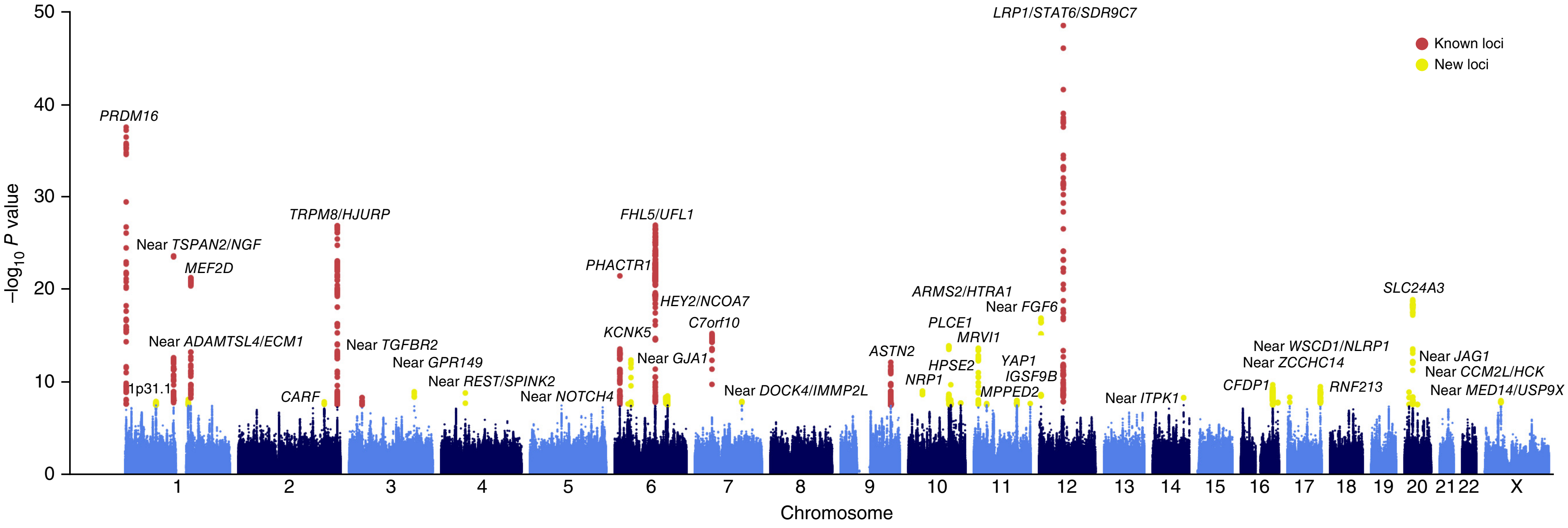
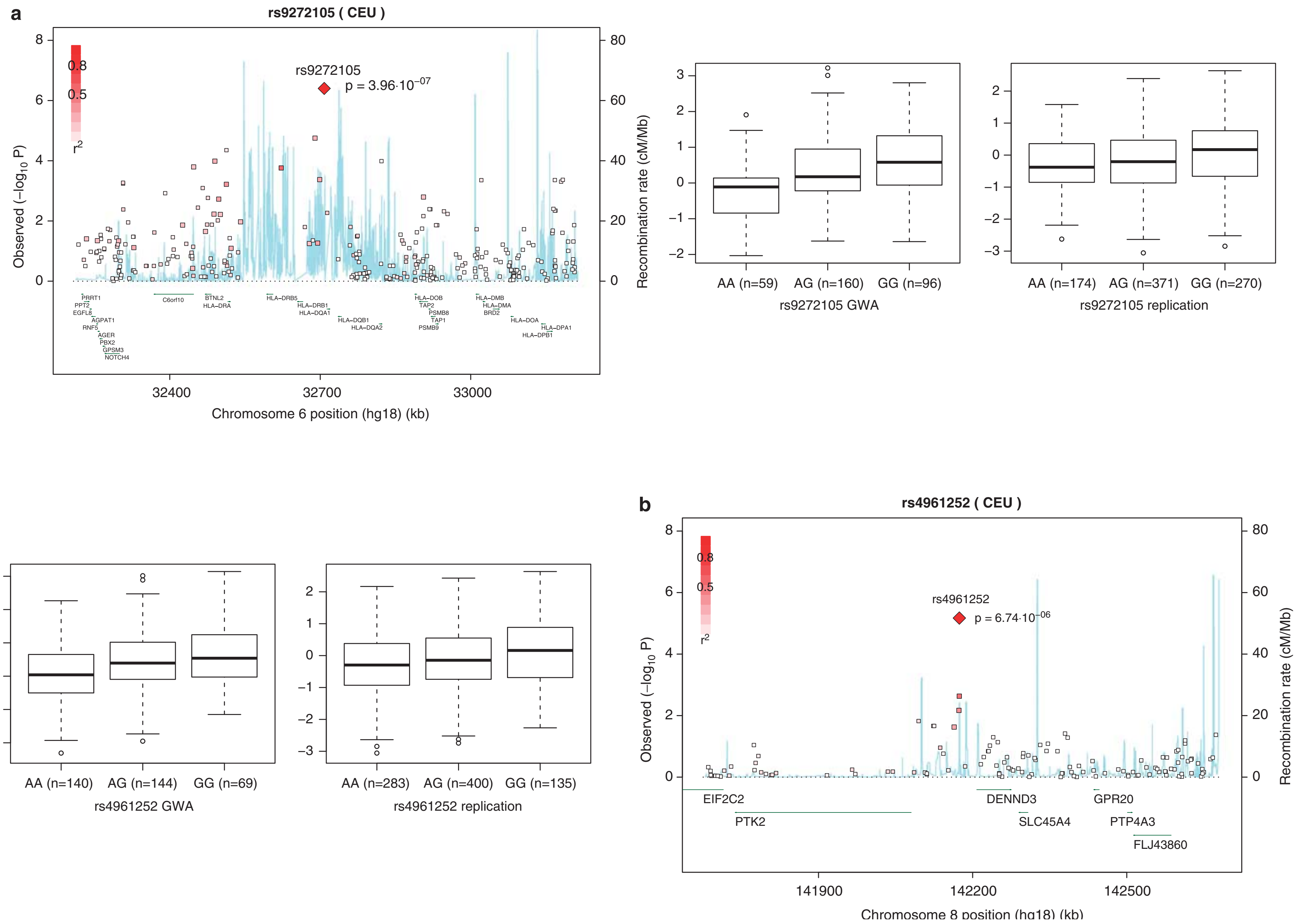
Max Planck Institute
of Psychiatry
German Research Institute of Psychiatry

The Case for Non-Linearity in Statistical Genetics and Beyond

The age of linearity - GWAS

Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine




a
rs9272105 (CEU)
rs9272105
p = 3.96·10^{-07}
0.8
0.5
r^2
Observed (−log$_{10}$ P)
Recombination rate (cM/Mb)
Chromosome 6 position (hg18) (kb)
32400
32700
33000
PRRT1
PPT2
EGFL8
AGPAT1
RNF5
AGER
PBX2
GPSM3
NOTCH4
C6orf10
BTNL2
HLA–DRA
HLA–DRB5
HLA–DRB1
HLA–DQA1
HLA–DQB1
HLA–DQA2
HLA–DOB
TAP2
PSMB8
TAP1
PSMB9
HLA–DMB
HLA–DMA
BRD2
HLA–DOA
HLA–DPA1
HLA–DPB1
AA (n=59)
AG (n=160)
GG (n=96)
rs9272105 GWA
AA (n=174)
AG (n=371)
GG (n=270)
rs9272105 replication
b
rs4961252 (CEU)
rs4961252
p = 6.74·10^{-06}
Chromosome 8 position (hg18) (kb)
141900
142200
142500
EIF2C2
PTK2
DENND3
SLC45A4
GPR20
PTP4A3
FLJ43860
AA (n=140)
AG (n=144)
GG (n=69)
rs4961252 GWA
AA (n=283)
AG (n=400)
GG (n=135)
rs4961252 replication

XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. By **R. A. Fisher**, B.A. *Communicated by* Professor J. ARTHUR THOMSON. (With Four Figures in Text.)

Then a population in which this factor is the only cause of variability has its mean at

$$m = Pa + 2Qd - Ra,$$

so that

$$P(a-m) + 2Q(d-m) - R(a+m) = 0.$$

Let now

$$P(a-m)^2 + 2Q(d-m)^2 + R(a+m)^2 = \alpha^2 \quad . \quad . \quad . \quad . \quad . \qquad (I)$$

α^2 then is the variance due to this factor, for it is easily seen that when two such factors are combined at random, the mean square deviation from the new mean is equal to the sum of the values of α^2 for the two factors separately. In general the mean square deviation due to a number of such factors associated at random will be written

$$\sigma^2 = \Sigma\alpha^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (II)$$

To justify our statement that α^2 is the contribution which a single factor makes to the total variance, it is only necessary to show that when the number of such factors is large the distributions will take the normal form.

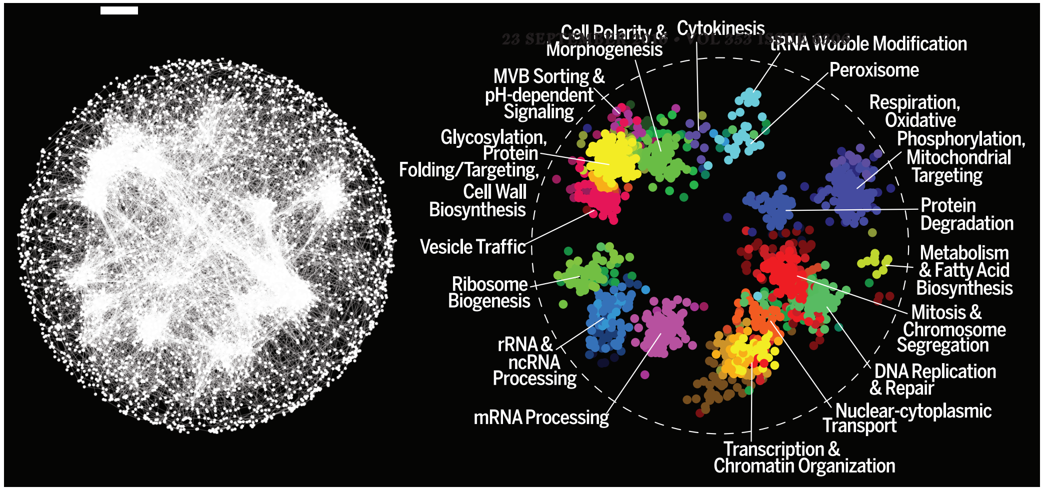
Transactions of the Royal Society of Edinburgh, 52: 399-433, (1918).

Is there no more?

YEAST GENETICS

A global genetic interaction network maps a wiring diagram of cellular function

Michael Costanzo,[1*] Benjamin VanderSluis,[2,3*] Elizabeth N. Koch,[2*]
Anastasia Baryshnikova,[4*] Carles Pons,[2*†] Guihong Tan,[1*] Wen Wang,[2] Matej Usaj,[1]
Julia Hanchard,[1,5] Susan D. Lee,[6] Vicent Pelechano,[7‡] Erin B. Styles,[1,5]
Maximilian Billmann,[8] Jolanda van Leeuwen,[1] Nydia van Dyk,[1] Zhen-Yuan Lin,[9]
Elena Kuzmin,[1,5] Justin Nelson,[2,10] Jeff S. Piotrowski,[1,11§] Tharan Srikumar,[12||]
Sondra Bahr,[1] Yiqun Chen,[1] Raamesh Deshpande,[2] Christoph F. Kurat,[1¶]
Sheena C. Li,[1,11] Zhijian Li,[1] Mojca Mattiazzi Usaj,[1] Hiroki Okada,[13] Natasha Pascoe,[1,5]
Bryan-Joseph San Luis,[1] Sara Sharifpoor,[1] Emira Shuteriqi,[1] Scott W. Simpkins,[2,10]
Jamie Snider,[1] Harsha Garadi Suresh,[1] Yizhao Tan,[1] Hongwei Zhu,[1]
Noel Malod-Dognin,[14] Vuk Janjic,[15] Natasa Przulj,[14,16] Olga G. Troyanskaya,[3,4]
Igor Stagljar,[1,5,17] Tian Xia,[2,18] Yoshikazu Ohya,[13] Anne-Claude Gingras,[5,9]
Brian Raught,[12] Michael Boutros,[8] Lars M. Steinmetz,[7,19] Claire L. Moore,[6]
Adam P. Rosebrock,[1,5] Amy A. Caudy,[1,5] Chad L. Myers,[2,10#]
Brenda Andrews,[1,5#] Charles Boone[1,5,11#]

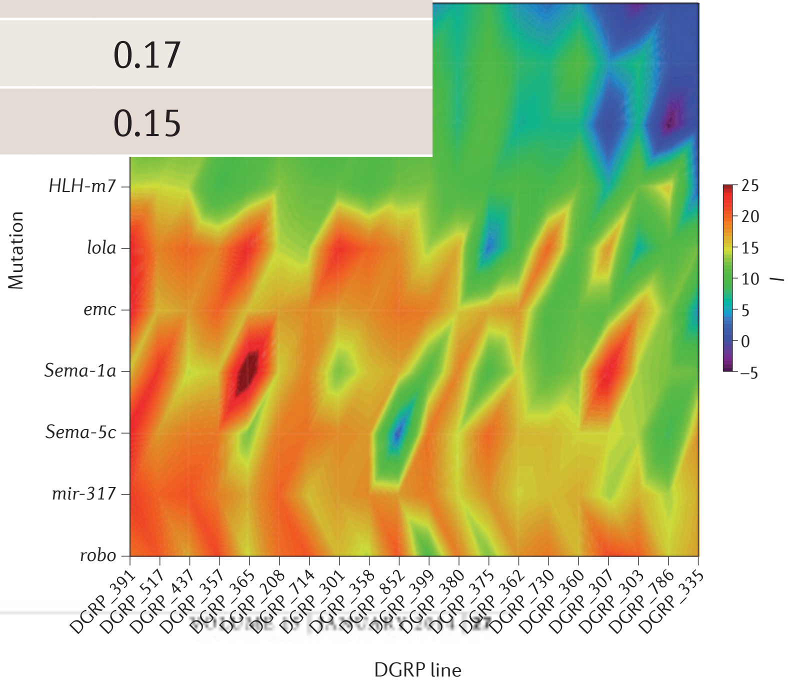


A global network of genetic interaction profile similarities. (**Left**) Genes with similar genetic interaction profiles are connected in a global network, such that genes exhibiting more similar profiles are located closer to each other, whereas genes with less similar profiles are positioned farther apart. (**Right**) Spatial analysis of functional enrichment was used to identify and color network regions enriched for similar Gene Ontology bioprocess terms.

Epistasis and quantitative traits: using model organisms to study gene–gene interactions

Trudy F. C. Mackay

Trait	**Observed**		**Expected**
	h^2	H^2	$H^2 = 2h^2/(1+h^2)$
Copulation latency	0.07	0.25	0.13
Startle response	0.16	0.58	0.28
Aggressive behaviour	0.09	0.78	0.17
Ethanol knock-down time	0.08	0.24	0.15



Even simple studies have more than linear effects in them … Example I

American Journal of Medical Genetics (Neuropsychiatric Genetics) 88:567–587 (1999)

Genome-Wide Search for Linkage of Bipolar Affective Disorders in a Very Large Pedigree Derived From a Homogeneous Population in Quebec Points to a Locus of Major Effect on Chromosome 12q23-q24

J. Morissette,[1] A. Villeneuve,[2] L. Bordeleau,[1] D. Rochette,[3] C. Laberge,[1] B. Gagné,[1] C. Laprise,[1] G. Bouchard,[4] M. Plante,[1] L. Gobeil,[1] E. Shink,[1] J. Weissenbach,[5] and N. Barden[1]*

[1] *Neuroscience, CHUL Research Center and Laval University, Québec, Canada*
[2] *Clinique Roy-Rousseau, Québec, Canada*
[3] *Complexe Hospitalier de la Sagamie, Chicoutimi, Québec, Canada*
[4] *Université du Québec à Chicoutimi, Chicoutimi, Québec, Canada*
[5] *Genoscope, Evry, France*

Molecular Psychiatry (2005) 10, 545–552
© 2005 Nature Publishing Group All rights reserved 1359-4184/05 $30.00
www.nature.com/mp

ORIGINAL RESEARCH ARTICLE

A genome-wide scan points to a susceptibility locus for bipolar disorder on chromosome 12

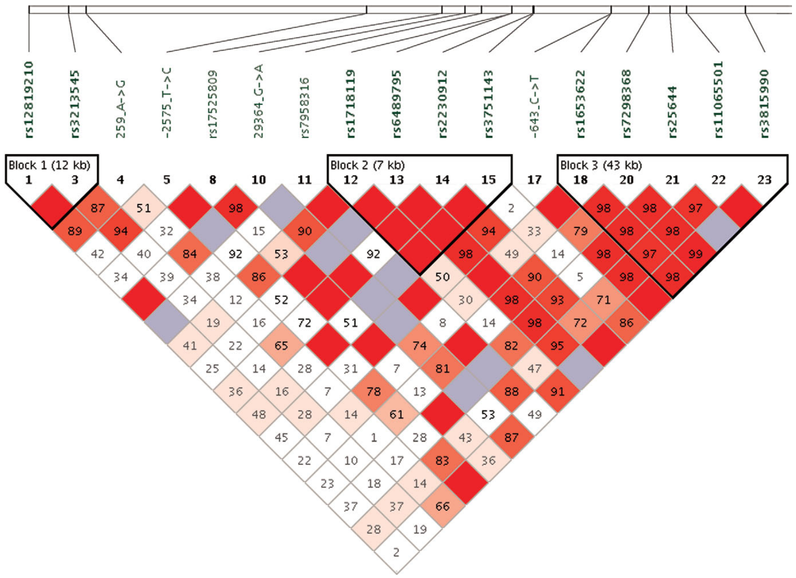
E Shink[1], J Morissette[2], R Sherrington[3] and N Barden[1]

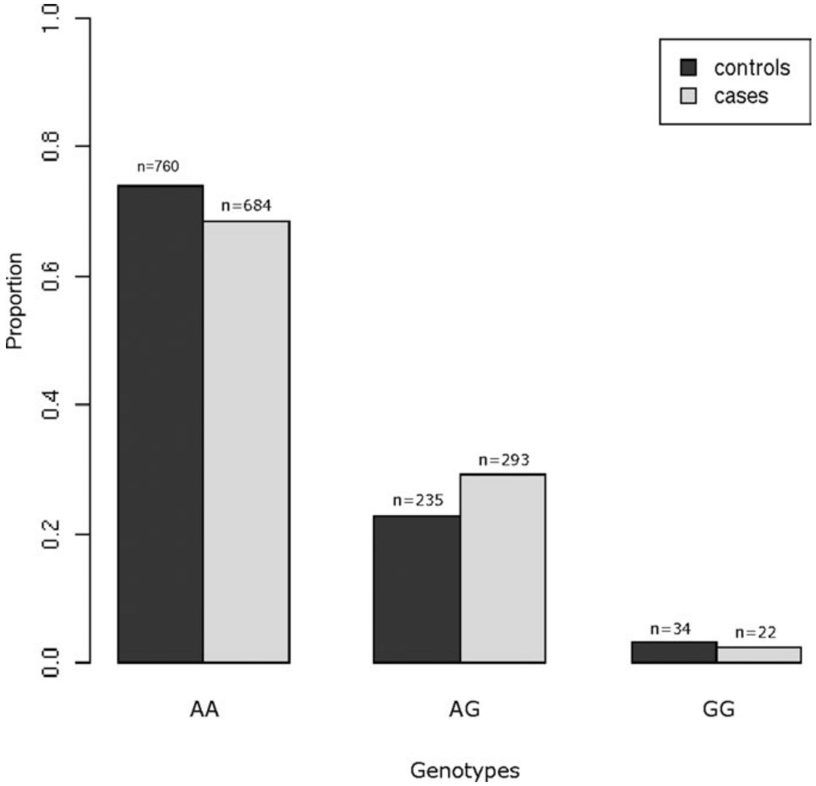
[1] *Neuroscience, CHUL Research Centre and Laval University, CHUQ Pavillon CHUL, Ste-Foy, Québec, Canada;*
[2] *Bioinformatics, CHUL Research Centre and Laval University, CHUQ Pavillon CHUL, Ste-Foy, Québec, Canada;*
[3] *Axys Pharmaceuticals Inc., South San Francisco, CA, USA*

Human Molecular Genetics, 2006, Vol. 15, No. 16 **2438–2445**
doi:10.1093/hmg/ddl166
Advance Access published on July 5, 2006

P2RX7, a gene coding for a purinergic ligand-gated ion channel, is associated with major depressive disorder

Susanne Lucae[1,†], Daria Salyakina[1,†], Nicholas Barden[2], Mario Harvey[2], Bernard Gagné[2], Michel Labbé[2], Elisabeth B. Binder[1,‡], Manfred Uhr[1], Marcelo Paez-Pereda[3], Inge Sillaber[3], Marcus Ising[1], Tanja Brückl[1], Roselind Lieb[1], Florian Holsboer[1] and Bertram Müller-Myhsok[1,*]

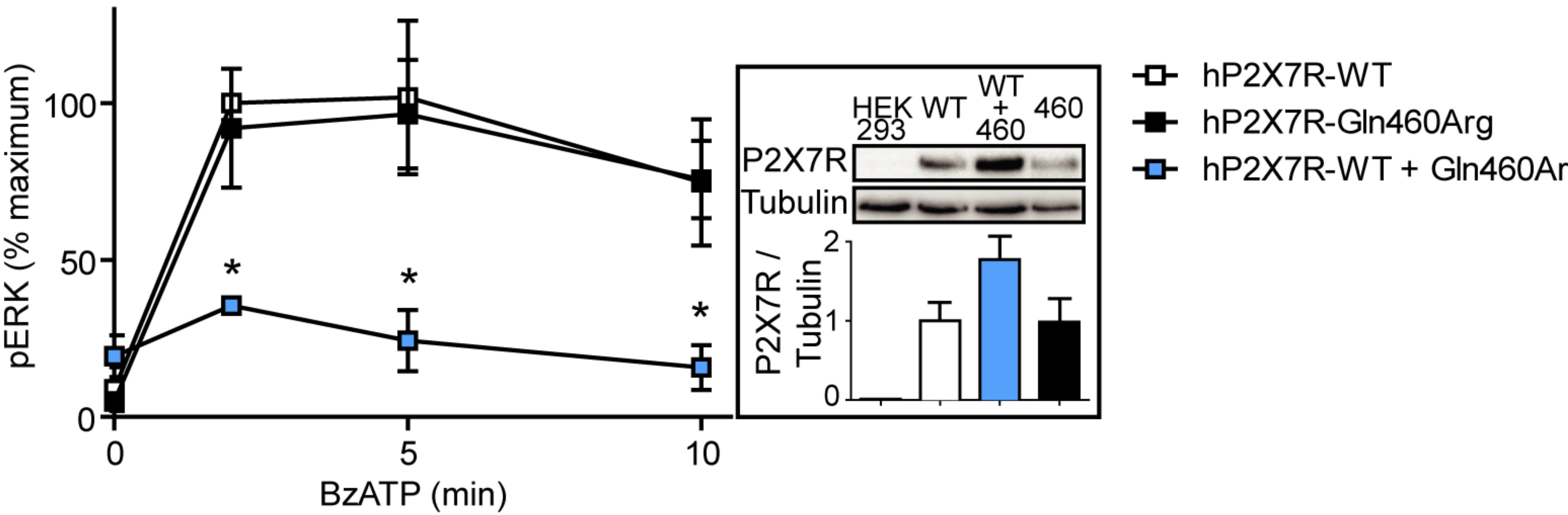




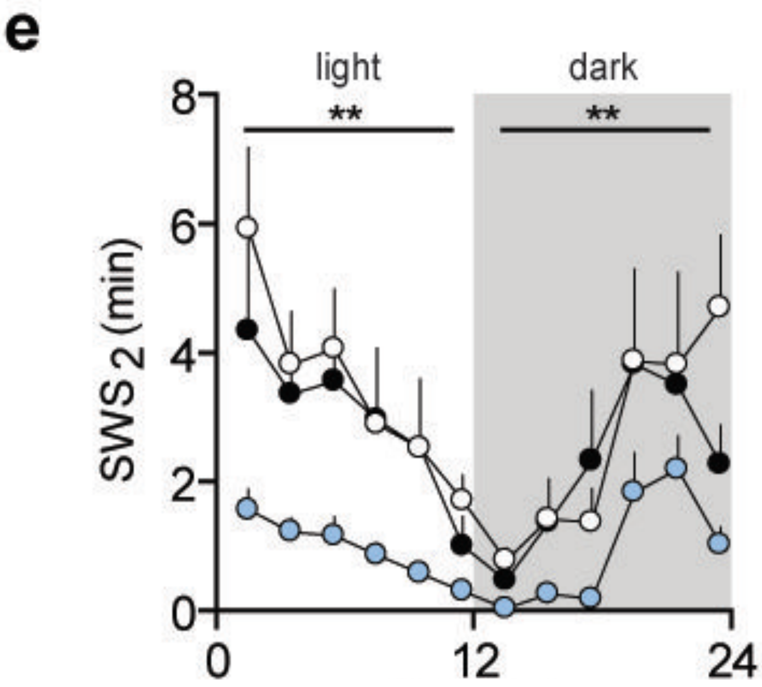
RESEARCH ARTICLE

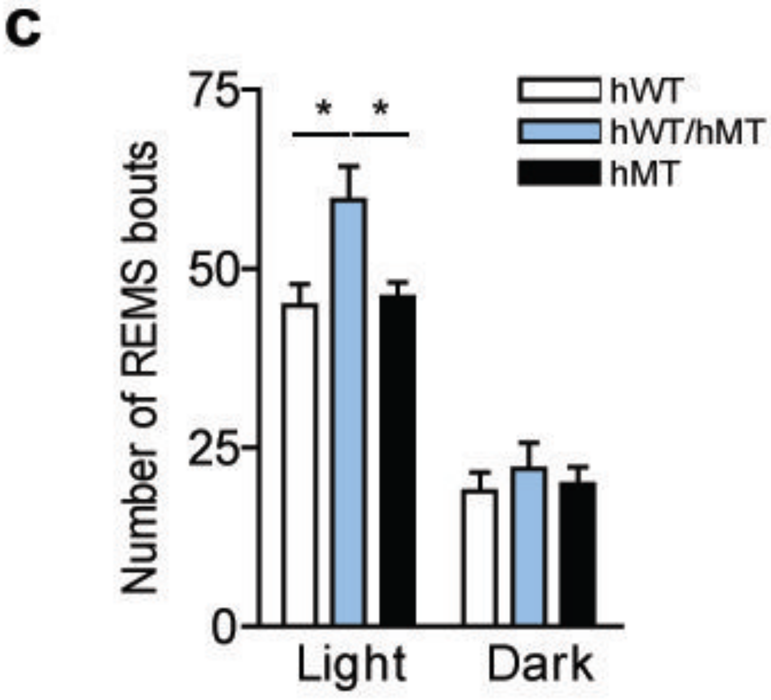
Co-Expression of Wild-Type P2X7R with Gln460Arg Variant Alters Receptor Function

Fernando Aprile-Garcia[1,2], Michael W. Metzger[3], Marcelo Paez-Pereda[3], Herbert Stadler[4], Matías Acuña[1], Ana C. Liberman[1], Sergio A. Senin[1], Juan Gerez[1], Esteban Hoijman[5], Damian Refojo[3], Mišo Mitkovski[6], Markus Panhuysen[4], Walter Stühmer[6], Florian Holsboer[3,7], Jan M. Deussing[3], Eduardo Arzt[1,2,3]*



P2X7R, sleep in mice

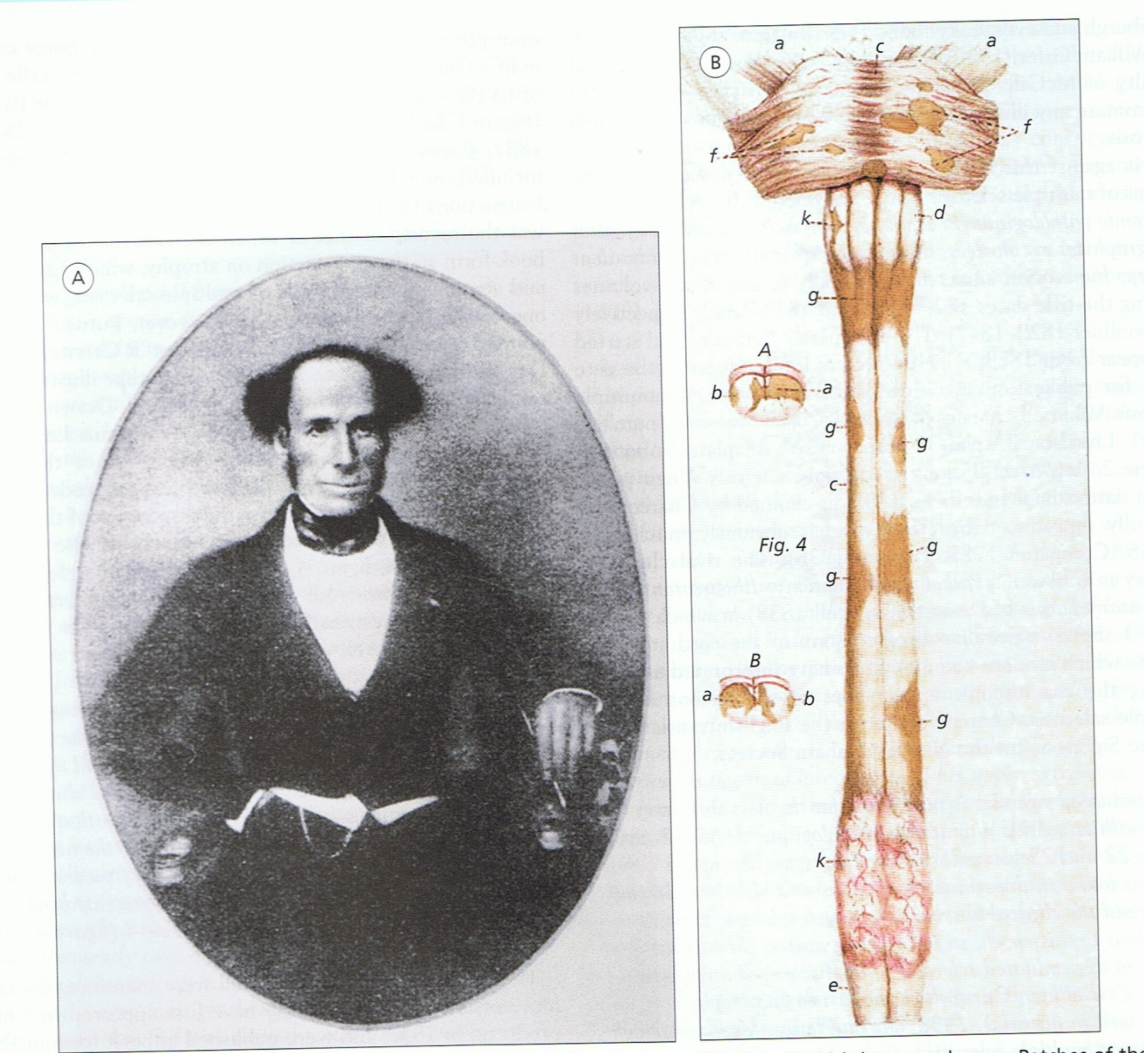




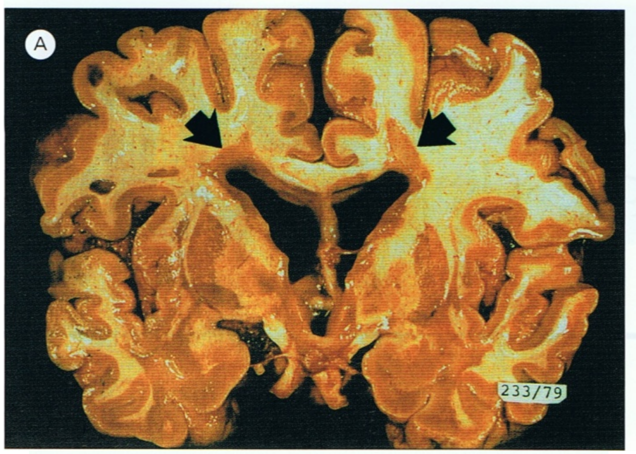
P2X7R, Aprile-Garcia, submitted

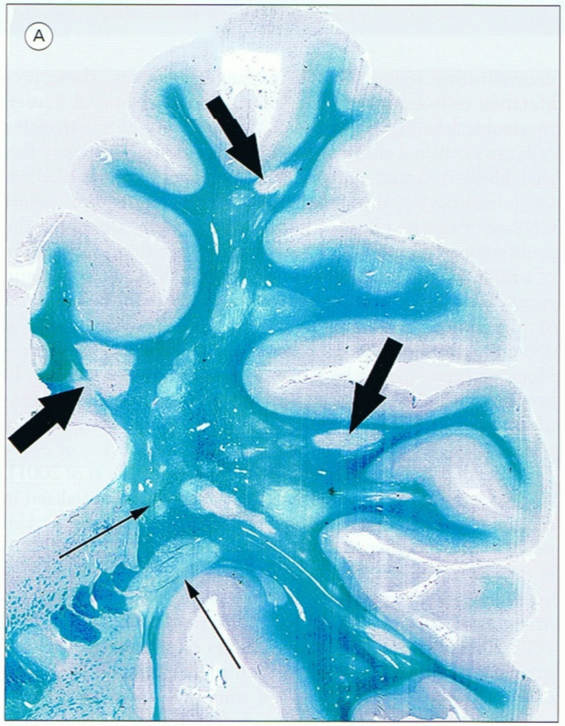
Even simple studies have more than linear effects in them … Example II

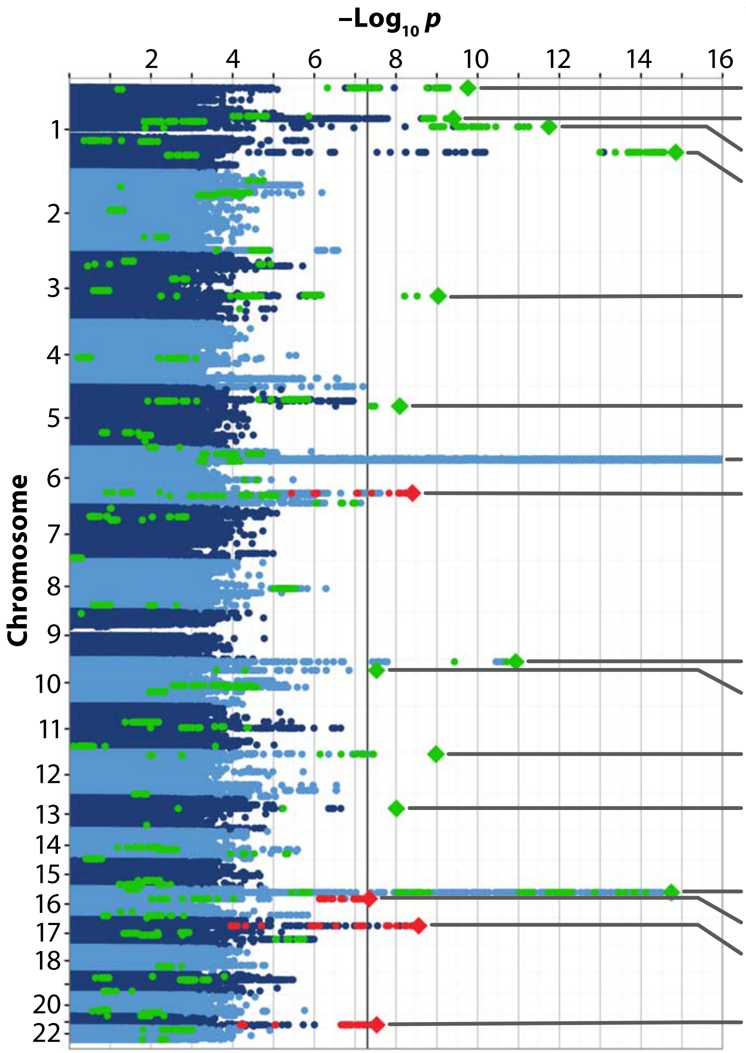
MS: an inflammatory, demyelinating disease



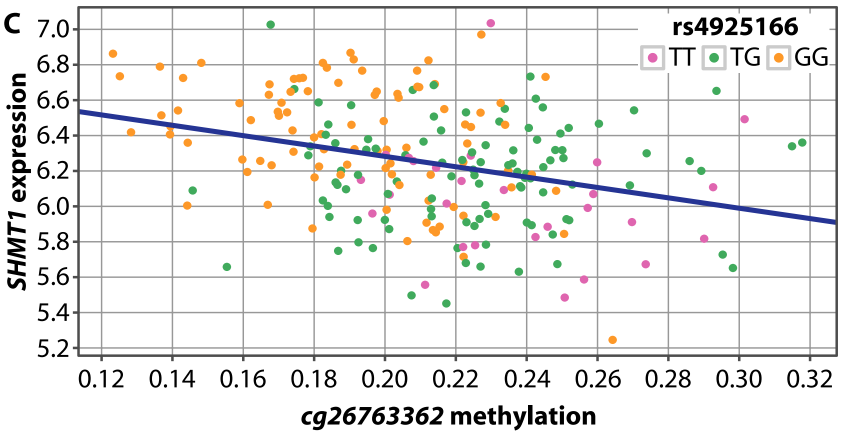
Carswell R 1838 aus Mc Alpine, Multiple Sclerosis

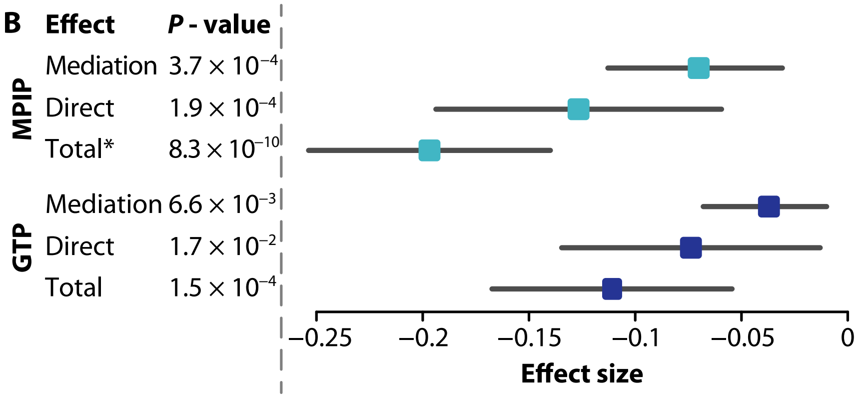


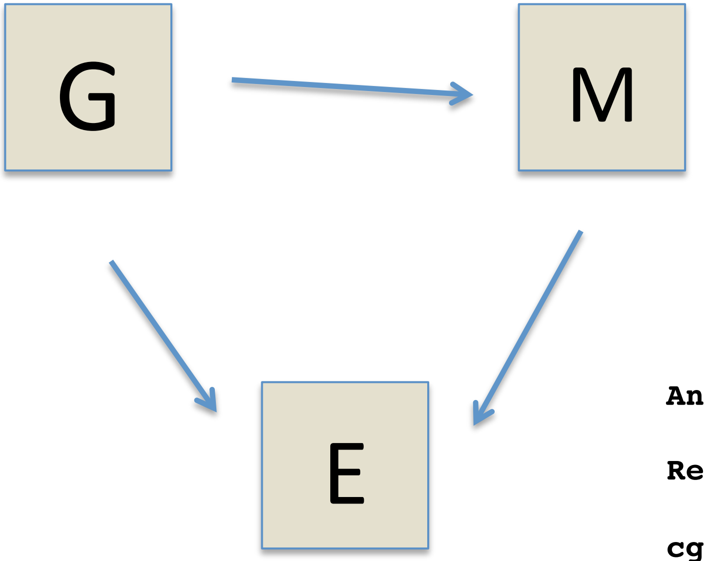




Variant	MA	OR	Gene
rs10797431	T	0.84	*MMEL1*
rs6689470	A	1.24	*EVI5*
rs2300747	G	0.75	*CD58*
rs7535818	G	0.76	*RGS1*
rs2681424	C	0.86	*CD86*
rs6859219	A	0.84	*ANKRD55*
rs3104373	T	2.90	*HLA-DRB1**
rs4364506	**A**	**0.84**	***L3MBTL3***
rs2182410	T	0.84	*IL2RA*
rs1891621	G	0.87	Intergenic
rs1800693	C	1.17	*TNFRSF1A*
rs2812197	T	0.86	*DLEU1*
rs6498168	T	1.23	*CLEC16A*
rs34286592	**T**	**1.21**	***MAZ***
rs4925166	**T**	**0.85**	***SHMT1***
rs2836425	**T**	**1.22**	***ERG***







Type	MPIP		GTP	
	Effect	*p*-value	Effect	*p*-value
Total effect	0.197	0	0.111	1.5e-04
Direct effect	0.127	1.9e-04	0.074	1.7e-02
Causal mediation effect	0.070	3.7e-04	0.037	6.6e-03
Proportion mediated	0.355	3.7e-04	0.332	6.6e-03

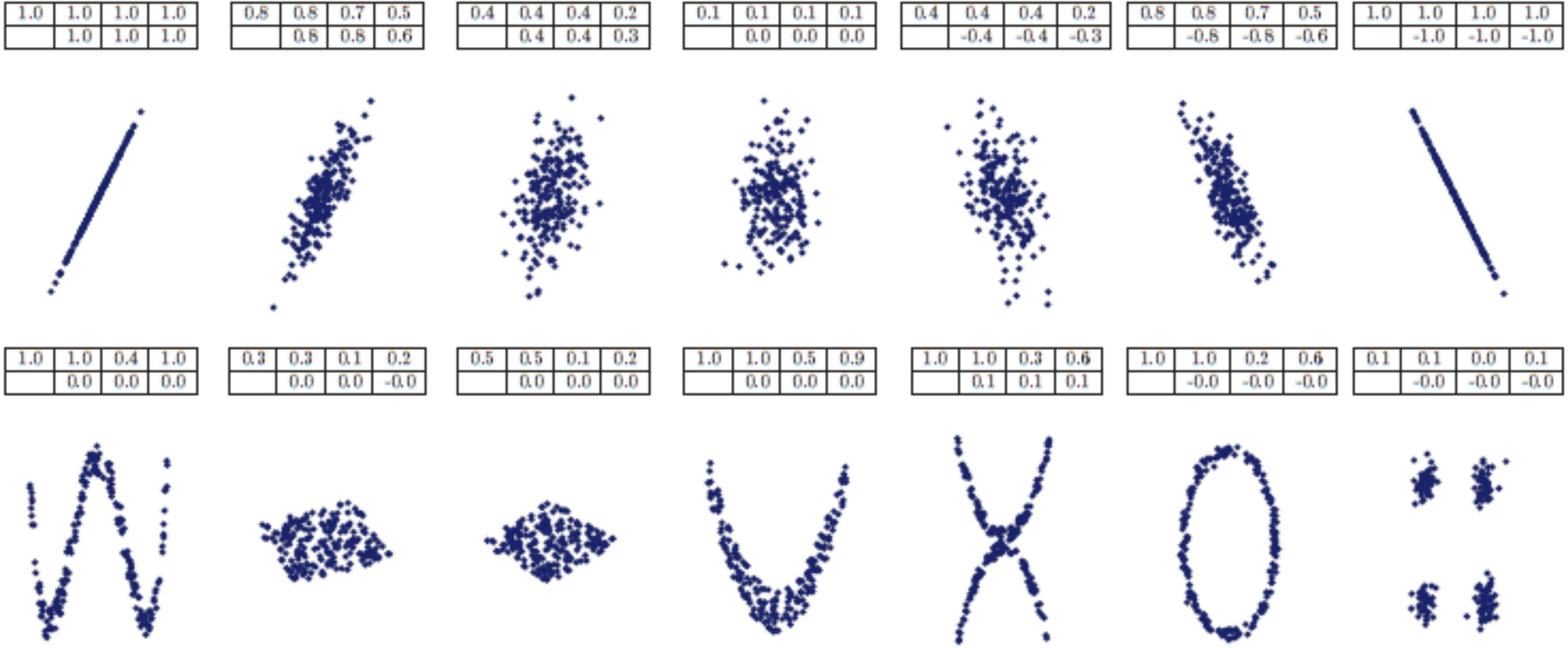
table S8. Causal mediation analysis.

```
Analysis of Variance Table

Response: ILMN_1811933_ori
                      Df  Sum Sq Mean Sq F value    Pr(>F)
cg26763362             1  2.5999 2.59995 28.1281 2.761e-07
rs4925166              1  1.7481 1.74806 18.9118 2.094e-05
cg26763362:rs4925166   1  0.4713 0.47129  5.0988   0.02492
Residuals            220 20.3352 0.09243
```

First tries at analysing non-linearities generally

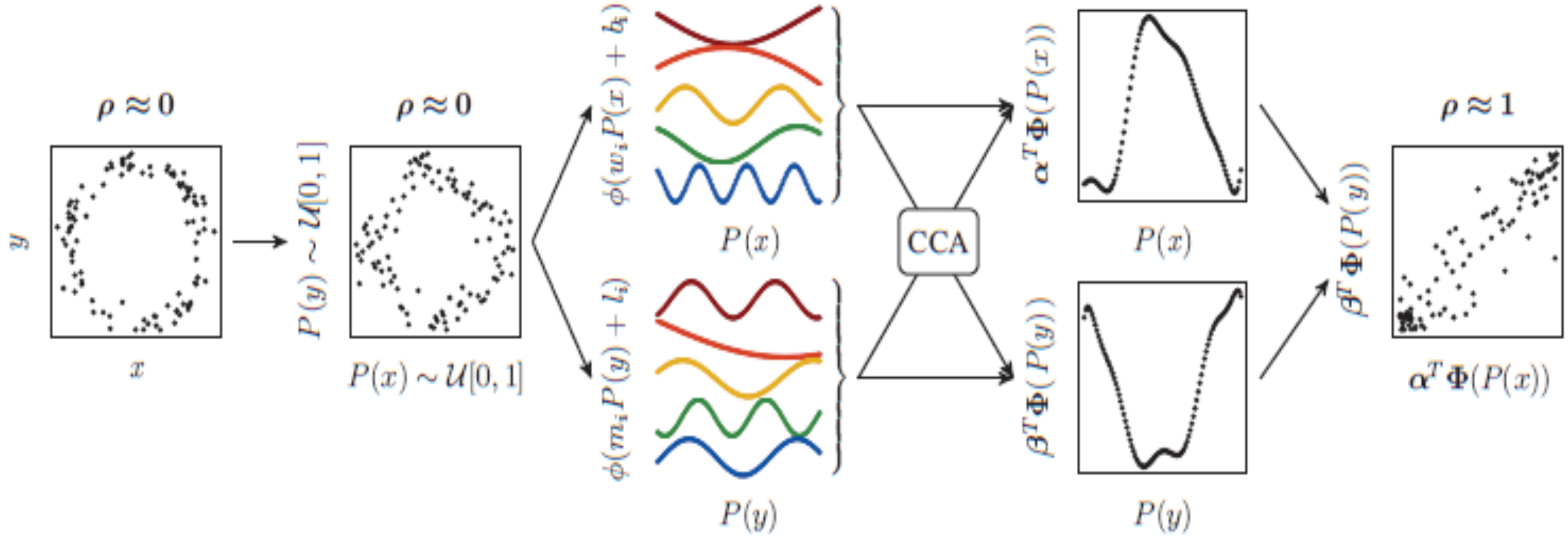
The Randomized Dependence Coefficient

David Lopez-Paz, Philipp Hennig, Bernhard Schölkopf
Max Planck Institute for Intelligent Systems
Spemannstraße 38, Tübingen, Germany
`{dlopez,phennig,bs}@tue.mpg.de`



The Randomized Dependence Coefficient

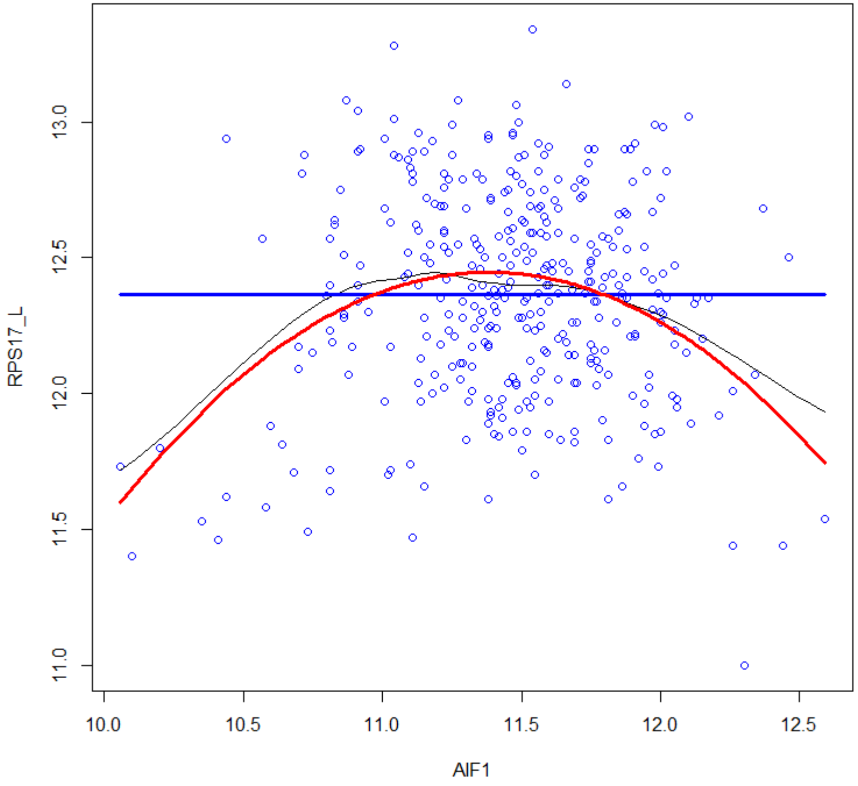
David Lopez-Paz, Philipp Hennig, Bernhard Schölkopf
Max Planck Institute for Intelligent Systems
Spemannstraße 38, Tübingen, Germany
`{dlopez,phennig,bs}@tue.mpg.de`



Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder

Divya Mehta[a,1], Torsten Klengel[a], Karen N. Conneely[b], Alicia K. Smith[c], André Altmann[a], Thaddeus W. Pace[c,d], Monika Rex-Haffner[a], Anne Loeschner[a], Mariya Gonik[a], Kristina B. Mercer[e], Bekh Bradley[c,f], Bertram Müller-Myhsok[a], Kerry J. Ressler[c,e,g], and Elisabeth B. Binder[a,c]

8302–8307 | PNAS | May 14, 2013 | vol. 110 | no. 20



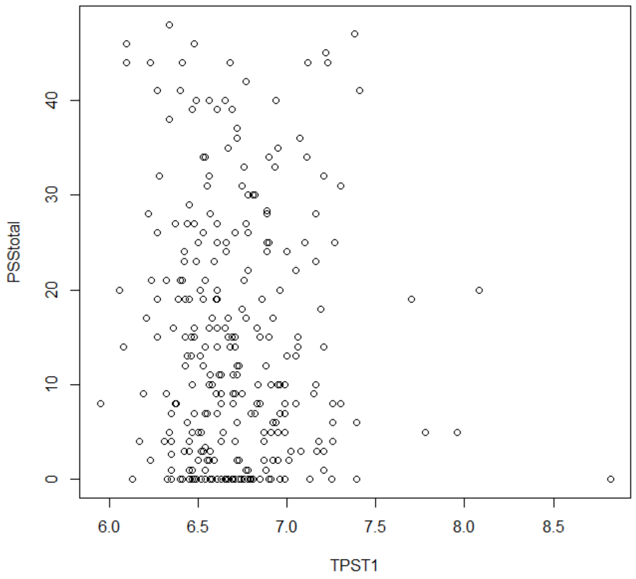
Stats again ....

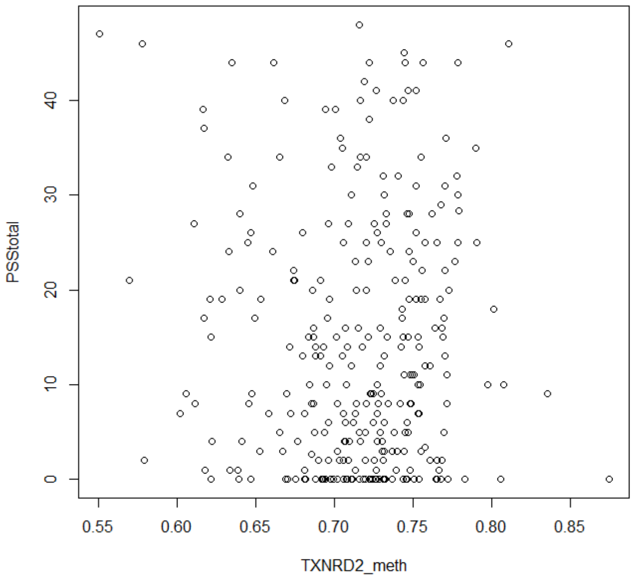
```
Pearson correlation:
Rho = -0.0015, P = 0.976
```

```
Randomized determination coefficient
RDC = 0.4445; P ~ 2.4e-11
```

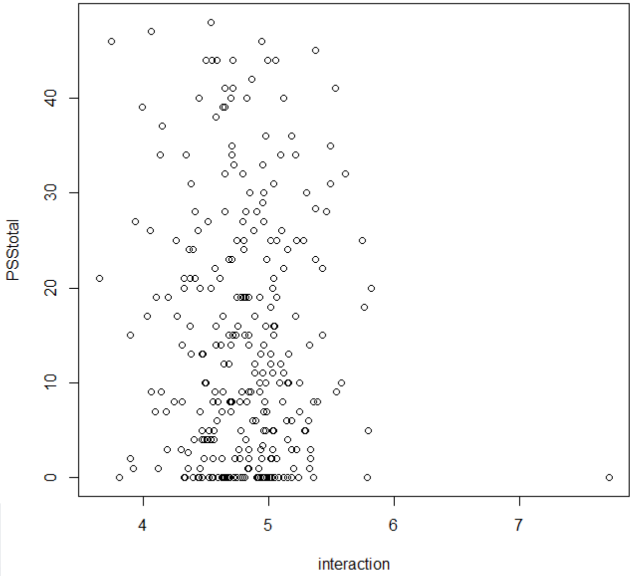
```
Analysis of Variance Table

Response: RPS17_L
                     Df  Sum Sq Mean Sq F value    Pr(>F)
AIF1                  1   0.000   0.0001   0.0010    0.9747
AIF1²                 1   5.442   5.4423 41.1463 4.433e-10 ***
AIF1³                 1   0.045   0.0451   0.3407    0.5598
Residuals           361 47.748   0.1323
```

Classic advice .. look at the data first

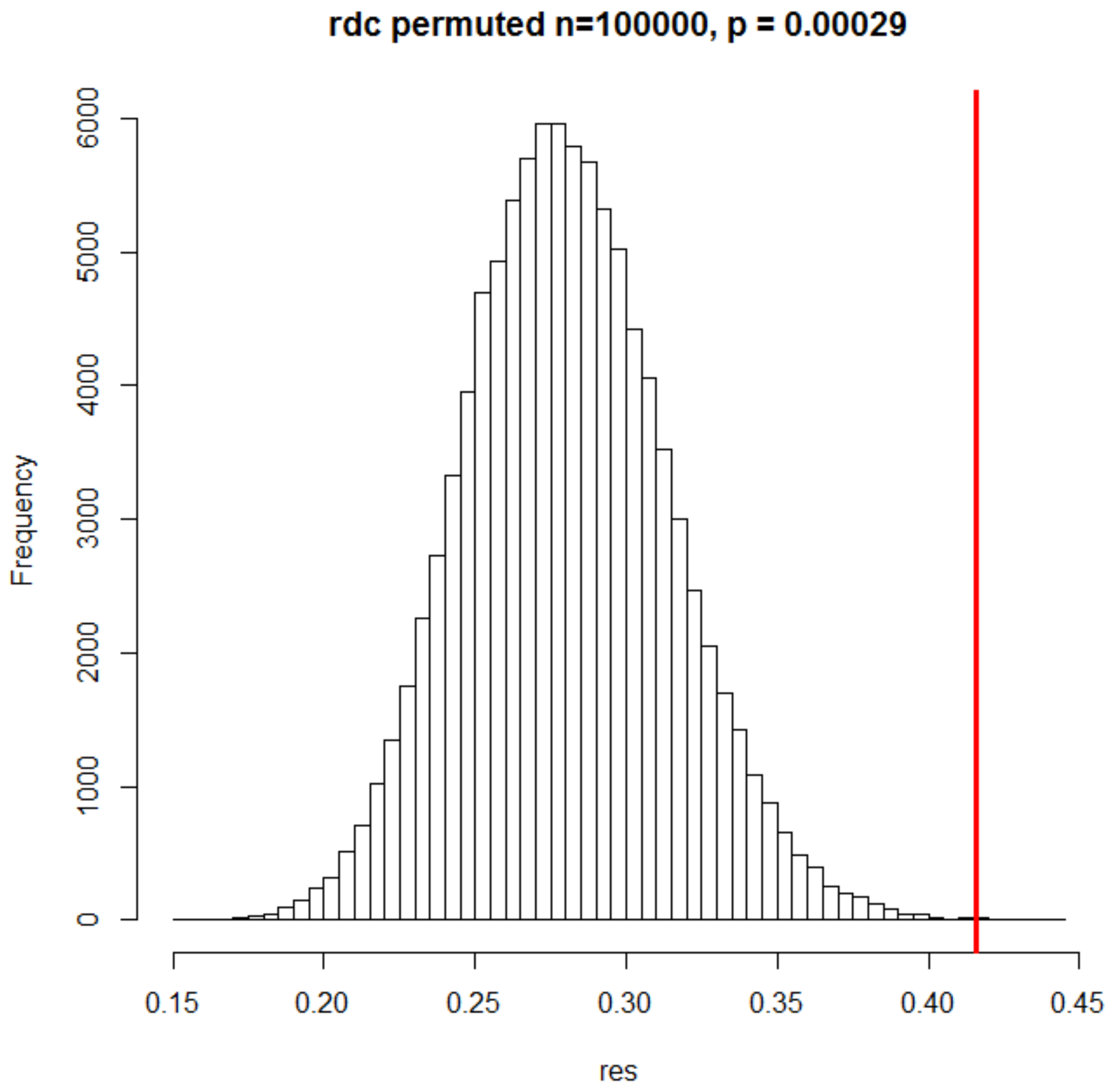




Bi-variate plots of PSStotal
vs TPST1 (expression),
TXNRD2 (methylation),
and their interaction term



RDC is 0.41 and significant …



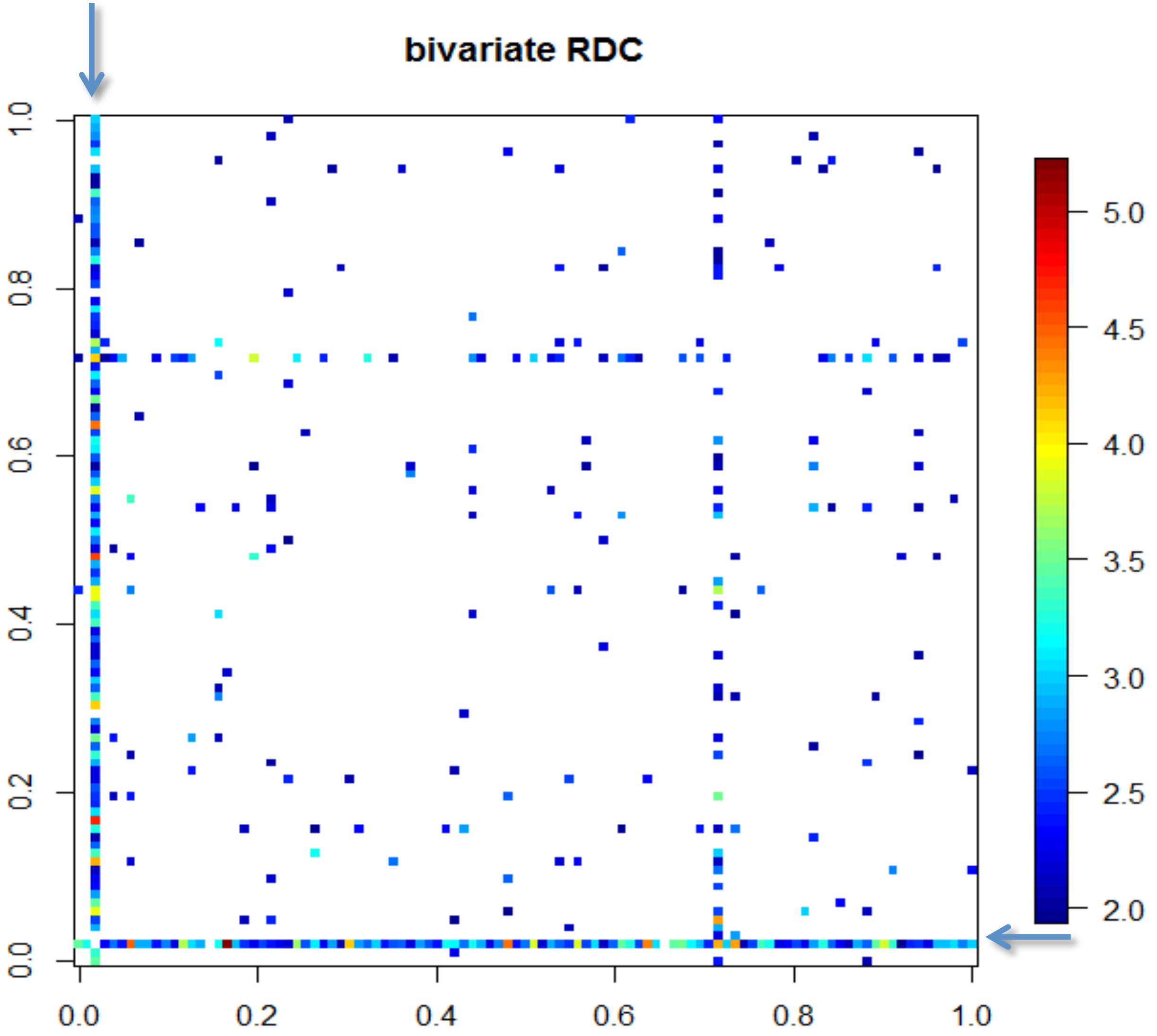
Evaluation of removable statistical interaction for binary traits

Jaya M. Satagopan[1] and **Robert C. Elston**[2]
[1]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

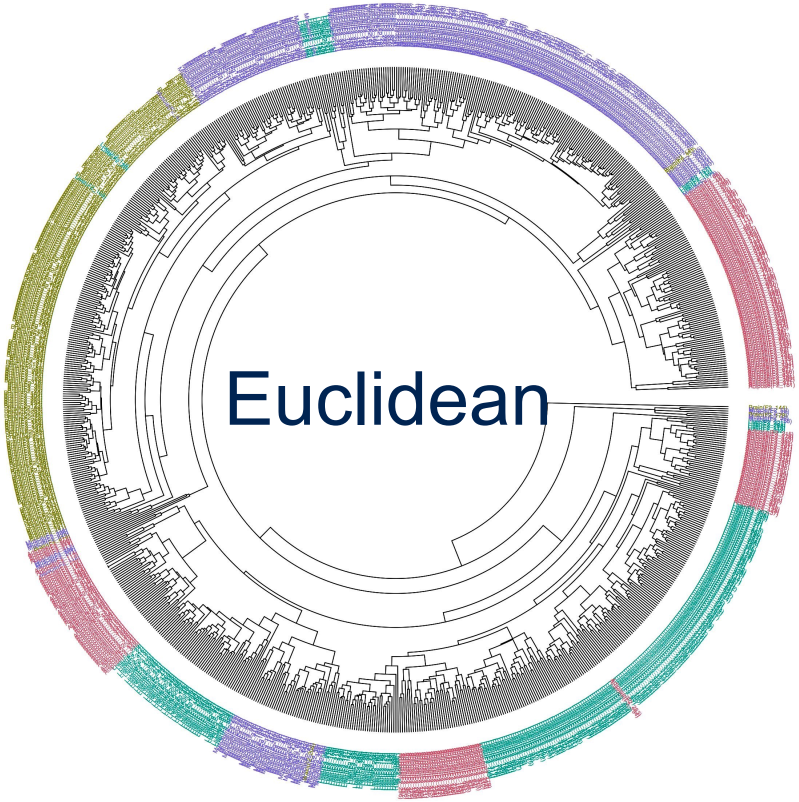
[2]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA

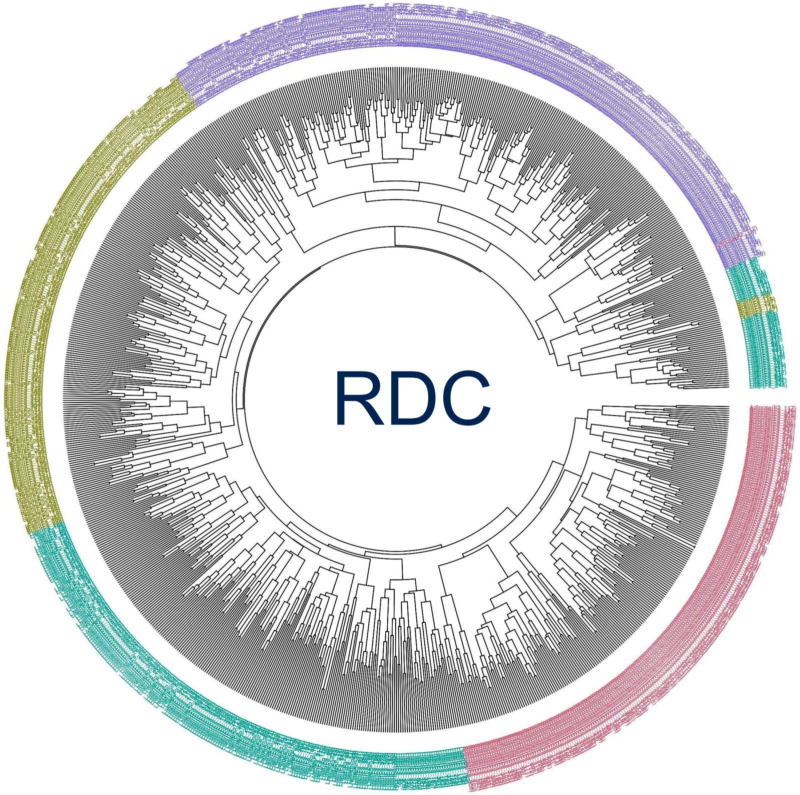
Give me an interaction and I'll transform it away ...

Predictive of higher-order interactions in lower-order analysis?



Can be used as a distance matrix in clustering





Kernel CCA

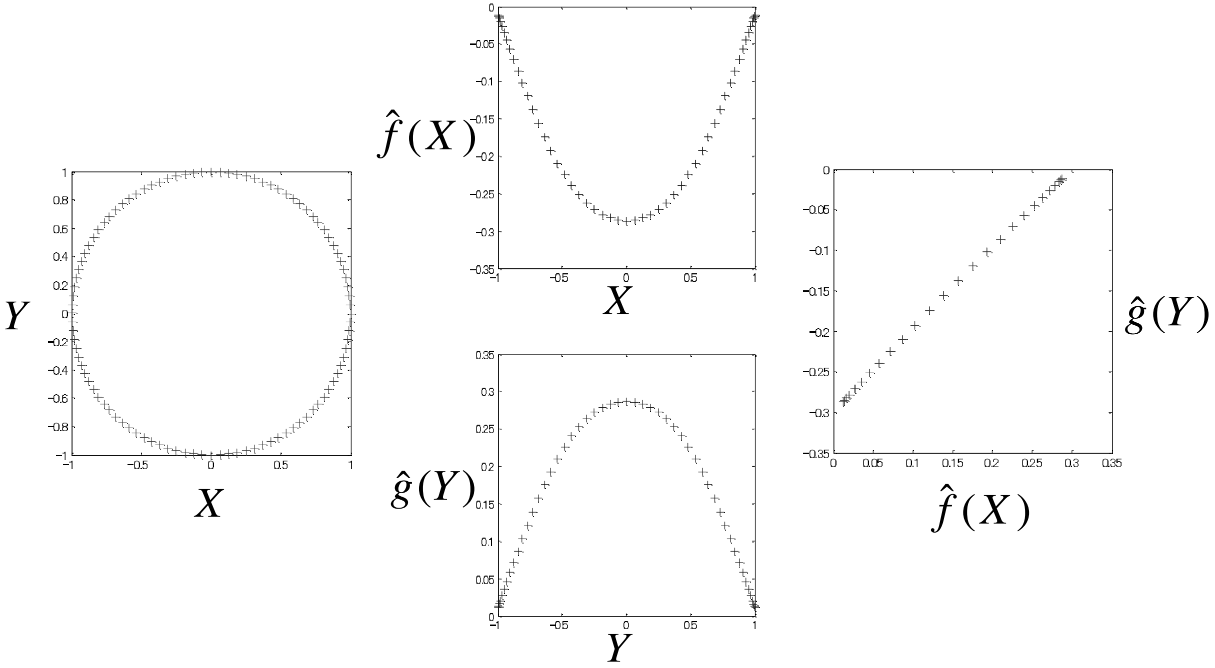
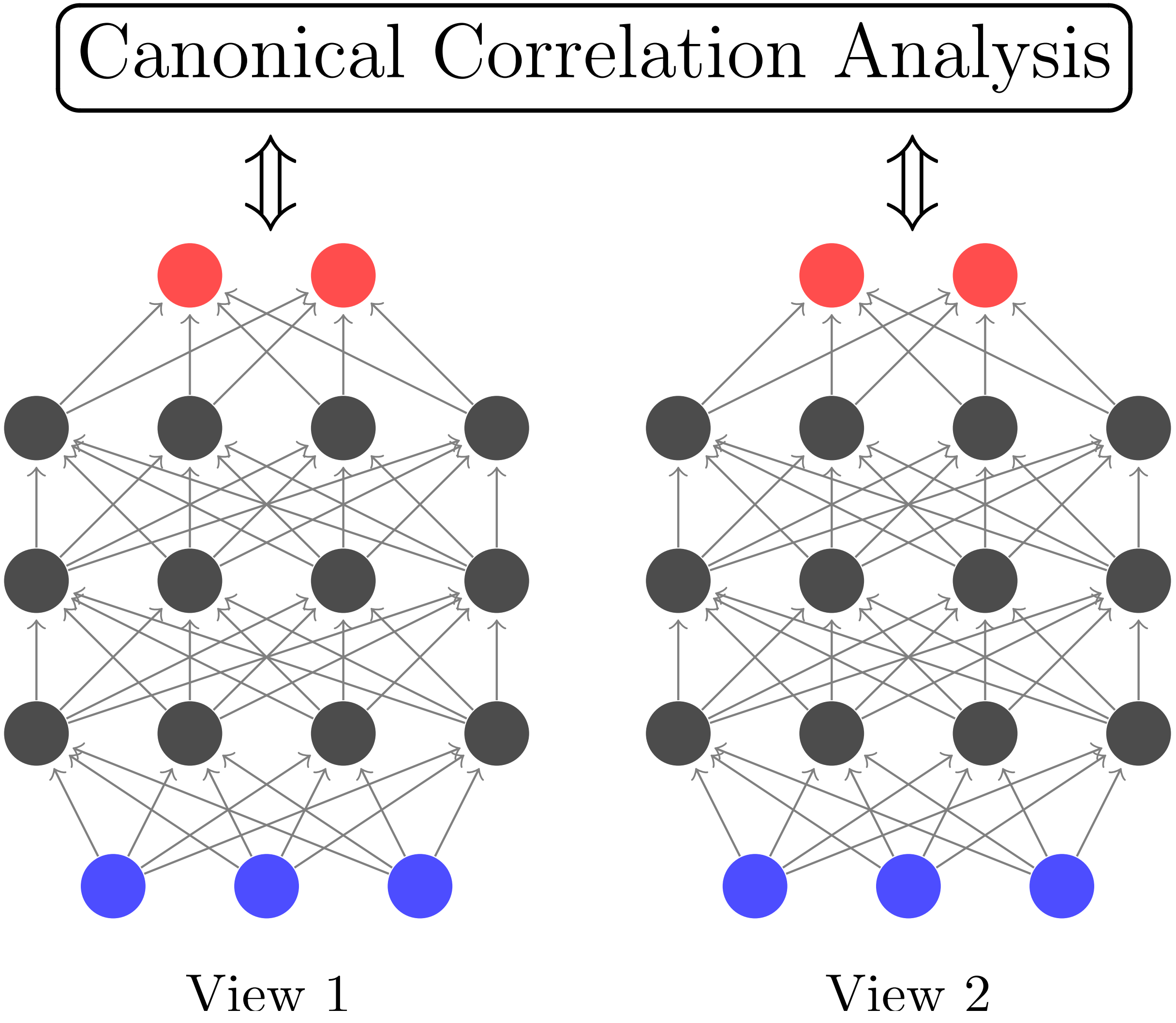


Figure 1: An example of kernel CCA. A Gaussian RBF kernel $k(x,y) = \exp\left(-\frac{1}{2\sigma^2}(x-y)^2\right)$ is used for both X and Y. Left: the original data. Center: derived functions $\widehat{f}(X_i)$ and $\widehat{g}(Y_i)$. Right: transformed data.

Journal of Machine Learning Research 8 (2007) 361-383

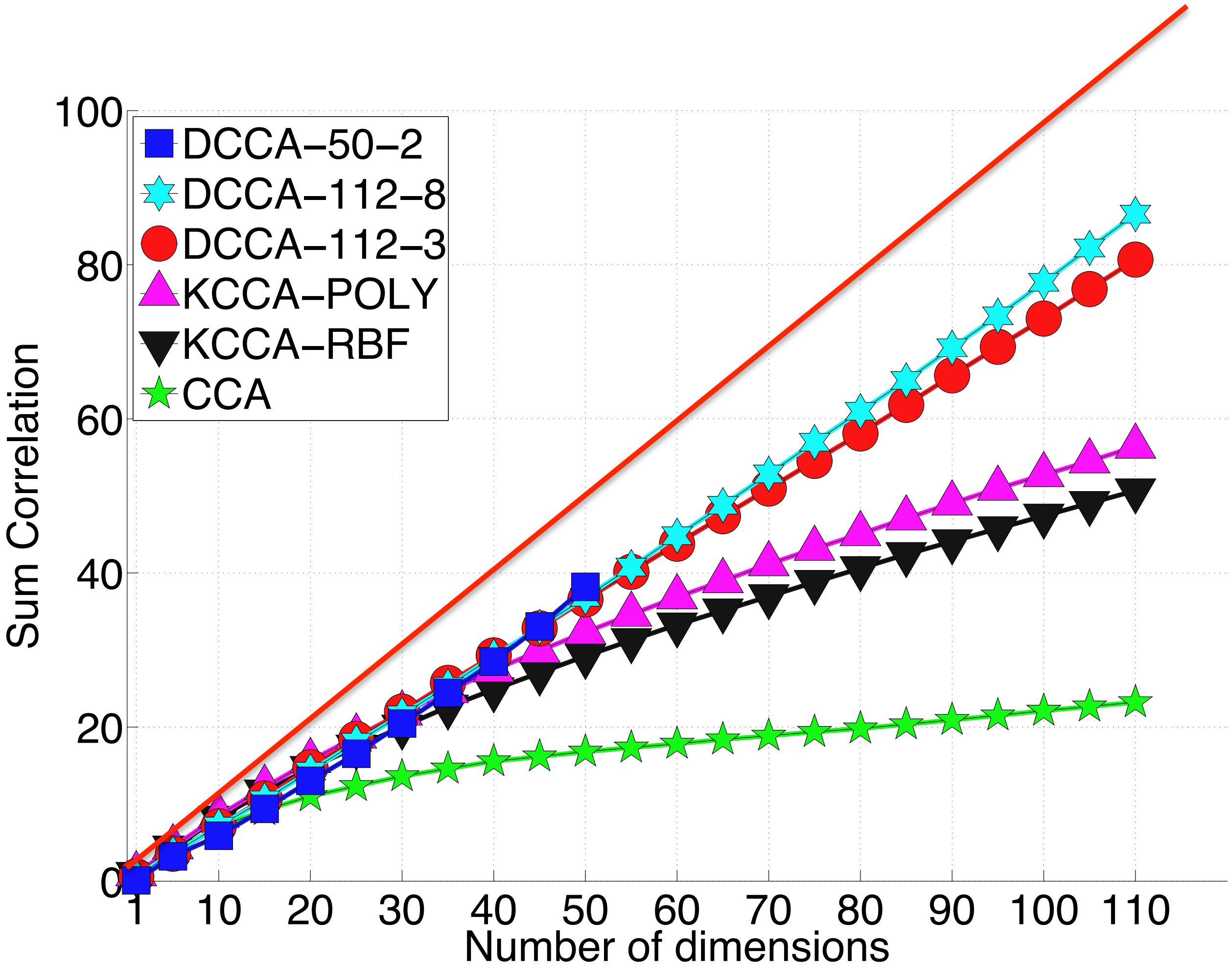


Canonical Correlation Analysis

View 1

View 2

Galen et al, ICML, 2013



Galen et al, ICML, 2013

A wishlist and a joint effort re statstics and machine elarning

Derive statistics from these tests

Define their distributions

Derive testing procedures based on these data

Implement feature extraction methods

Summing up ....

Bernard Prum († 2015)

You can do whatever you want ...

(as long as you pay the price ....)