Memorial Sloan Kettering Cancer Center

The genomics of normal tissue radiation response: principles, results, and opportunities

15 July 2020

Joseph O. Deasy (presenting), Jung Hun Oh, Sang Kyu Lee

Chair, Department of Medical Physics, MSKCC

“deasyj@mskcc.org” (email me for a copy of the slides)

Disclosure

- Co-founder and share-holder in PAIGE.AI, a computational pathology company.
- Funding support from
  - NCI (R01 and cancer center P30 grant)
  - The Breast Cancer Research Foundation
  - Simons Foundation
  - Philips, Corp.
  - Varian Oncology.

Thanks to

- **Jung Hun Oh, PhD**
- **Sang Kyu Lee, PhD**
- Sarah Kerns, PhD
- Barry Rosenstein, PhD

Memorial Sloan Kettering Cancer Center

“Although it has been known for many decades that radiotherapy patients differ in their radiosensitivity based on their genetic makeup differences, the ability to understand this variability and to potentially include it in treatment planning is only now becoming feasible. This talk will cover basic concepts in genetic variability, the impact on radiosensitivity (called radiogenomics), and recent results in modeling this variability. The talk will finish with thoughts about the next ten years of opportunities and potential progress in radiogenomics.”

“There is a lot to be gained by a much better understanding of the responses of normal tissues (and tumors) to a whole range of dose-volume distributions.”

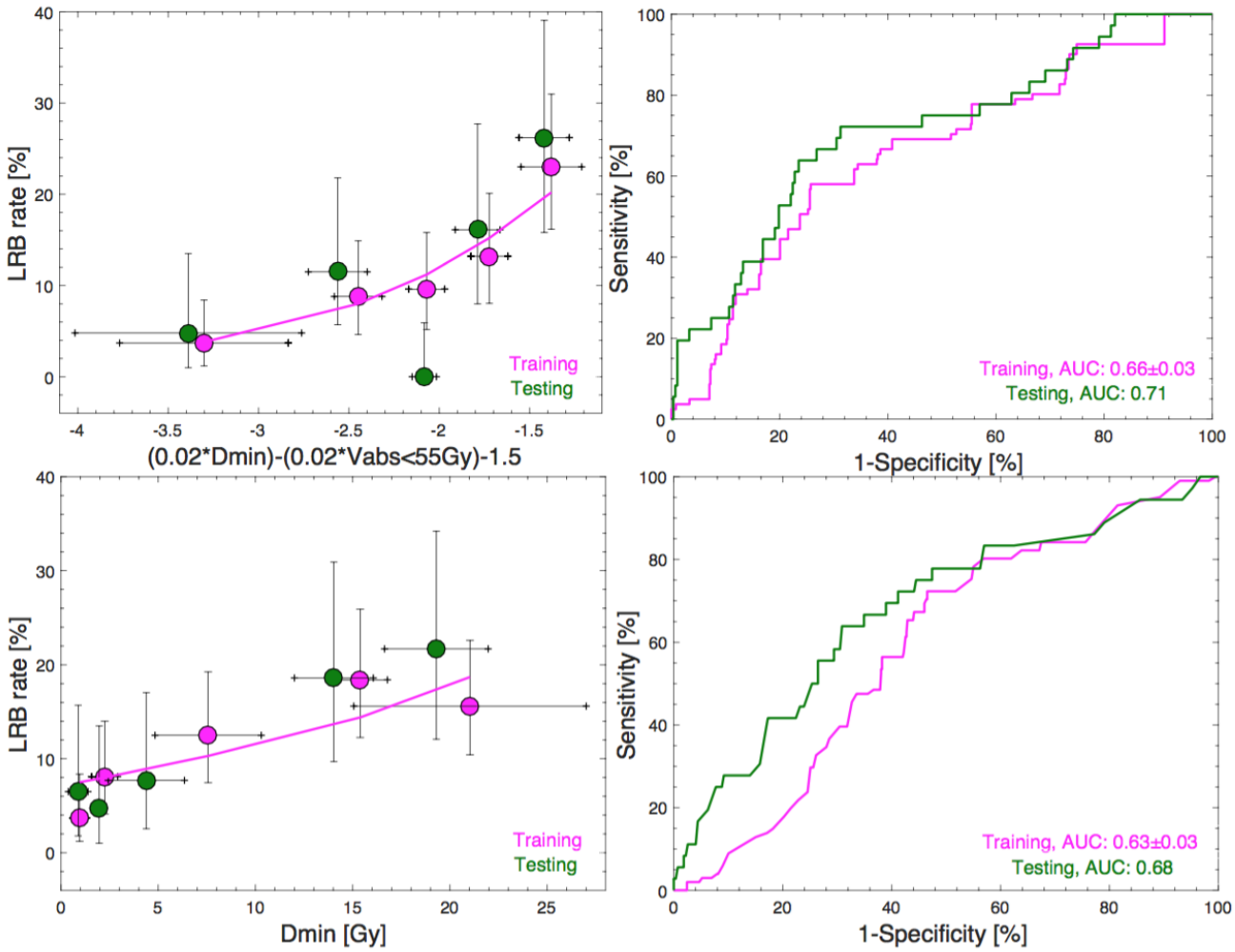
-- Michael Goitein (2007)

Pooled cohort analysis demonstrates the importance of rectal sparing in preventing late rectal bleeding

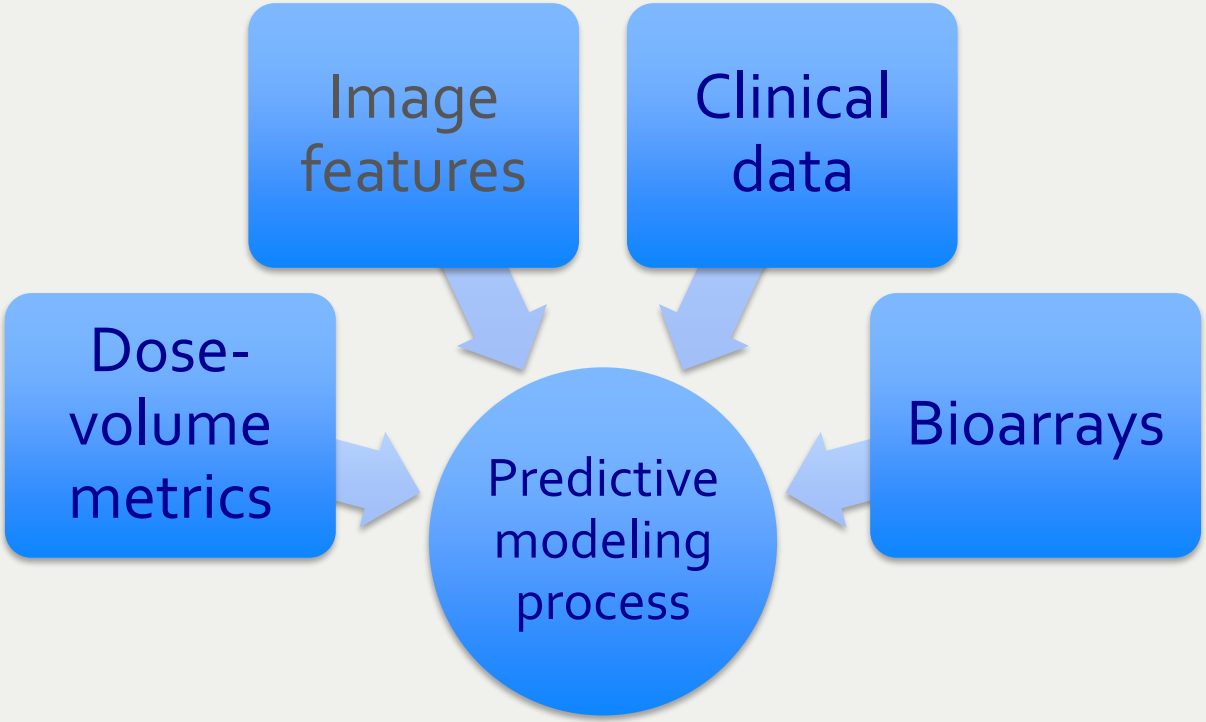
M Thor[1§], A Jackson[1], M J Zelefsky[2], G Steineck[3], Á Karlsdòttir[4], M Høyer[5], M Liu[6], N J Nasser[2], S E Petersen[5], V Moiseenko[7] and J O Deasy[1]

[1]Dept of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, US
[2]Dept of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, US
[3]Division of Clinical Cancer Epidemiology, Dept. of Oncology, Institute of Clinical Sciences, the Sahlgrenska Academy at the University of Gothenburg, Gothenburg, SE
[4]Dept of Oncology, Haukeland University Hospital, Bergen, NO
[5]Dept of Oncology, Aarhus University Hospital, Aarhus, DK
[6]British Columbia Cancer Agency, Vancouver Cancer Center, Vancouver, CA
[7]Dept of Radiation, Medicine and applied sciences, University of California San Diego, La Jolla, US

- 989 patients
- treated with 3DCRT or IMRT to
- 70-86.4Gy@1.8-2.0Gy/fraction

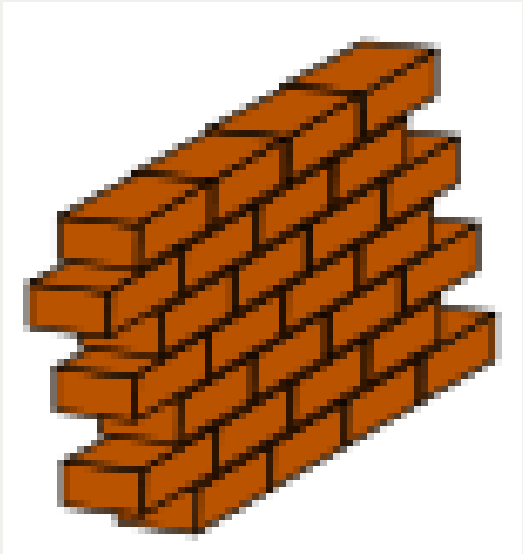

LRB rate [%]
(0.02*Dmin)-(0.02*Vabs<55Gy)-1.5
Training
Testing
Sensitivity [%]
1-Specificity [%]
Training, AUC: 0.66±0.03
Testing, AUC: 0.71
LRB rate [%]
Dmin [Gy]
Training
Testing
Sensitivity [%]
1-Specificity [%]
Training, AUC: 0.63±0.03
Testing, AUC: 0.68

Predictive modeling: inputs



The Past:
Outcomes = physics *or* biology

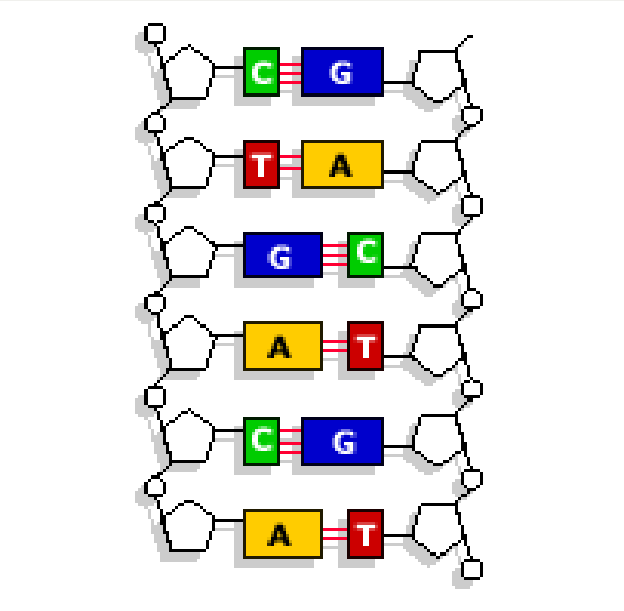
Outcomes determined by physics: prescription dose, normal tissue volume effects



Outcomes determined by clinical radiobiology factors: stage, grade, fractionation, hypoxia, radiosensitivity, proliferation

JOD

About DNA



- Double-stranded DNA molecule held together by chemical components called bases
- Adenine (A) bonds with thymine (T); cytosine(C) bonds with guanine (G)
- These letters form the "code of life". Estimated to be about 2.9 billion base-pairs in the human genome wound into 24 distinct bundles, or chromosomes
- Written in the DNA are about 30,000 genes which human cells use as starting templates to make proteins. These sophisticated molecules build and maintain our bodies

(taken from the BBC)

Genetic contributions to a complex phenotype

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

NATURE GENETICS VOLUME 42 | NUMBER 7 | JULY 2010

Key points

- Replicated single SNPs identified to date explain only ~5% of the phenotypic variance for height.
- Common SNPs in total explain another ~40% of phenotypic variance.
- Hence, most variation due to SNPs has been undetected in published GWASs because the effects of the SNPs are too small to be statistically significant.

‘**Radiogenomics**’ unfortunately used in two ways:

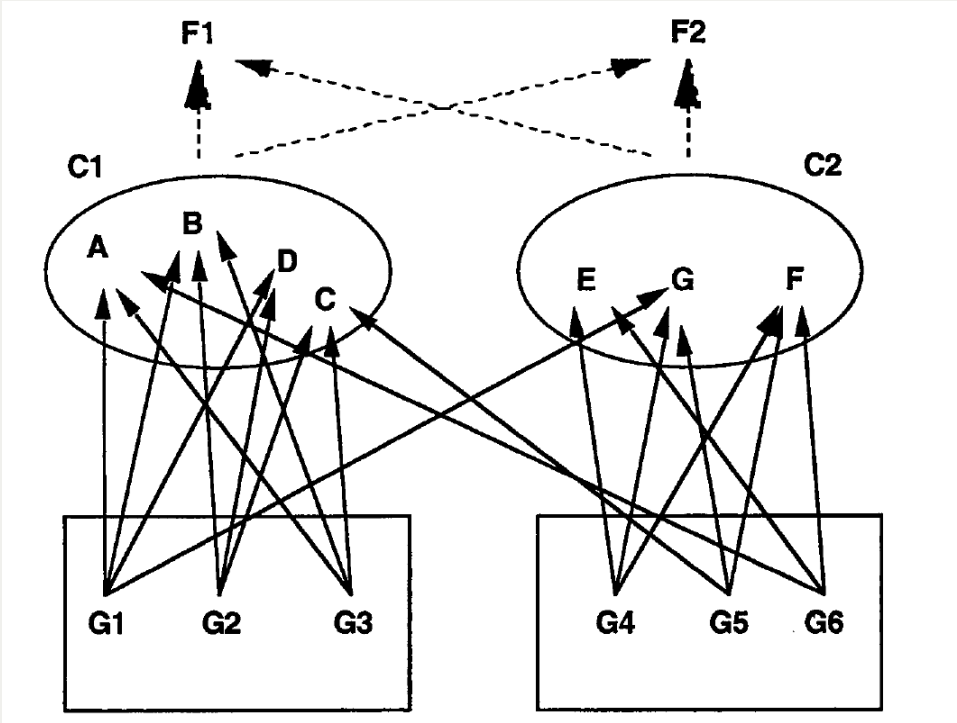
(1) imaging correlations with tumor genomics,
(2) correlations between ‘radiation response’ and genomics.

I prefer ‘radiation response genomics’ to be clear.

Evolution drives modularity, but imperfectly

COMPLEX ADAPTATIONS AND THE EVOLUTION OF EVOLVABILITY

GÜNTER P. WAGNER[1] AND LEE ALTENBERG[2,3]
[1]*Center for Computational Ecology, Department of Biology, Yale University, New Haven, Connecticut 06511*
E-mail: gpwag@peaplant.biology.yale.edu
[2]*Hawaii Institute of Geophysics and Planetology, University of Hawaii at Manoa, Manoa, Hawaii 96822*



Evolution (1996)

Key points

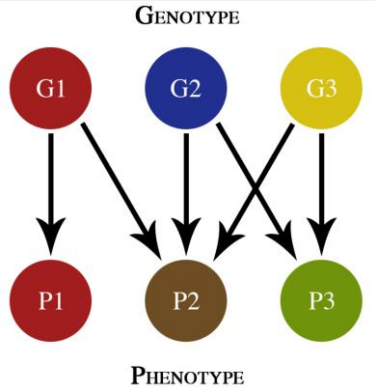
- Evolution works by introducing random differences in the machinery responsible for any defined phenotype.
- Genotypic modularity makes evolution feasible
- But inevitably, evolution randomly tries “crosstalk” links, to see if they are favorable

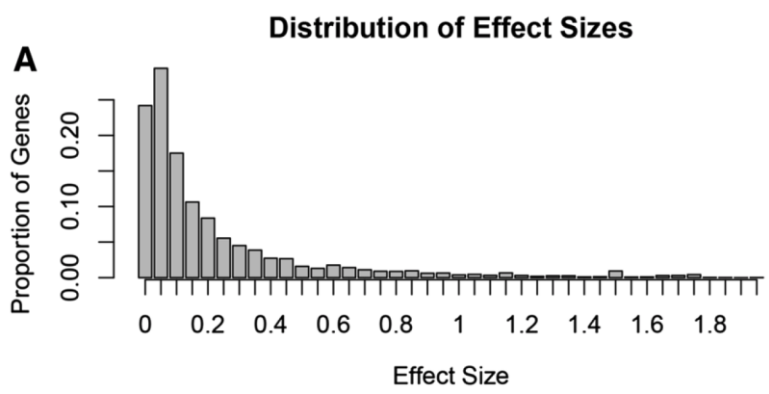
Small genetic effects on many processes is typical

The ubiquity of pleiotropy in human disease

Kevin Chesmore[1] · Jacquelaine Bartlett[2] · Scott M. Williams[2,3]

Human Genetics (2018) 137:39–44





Key points

- Single genes contribute to multiple biological processes
- Complex phenotypes are typically impacted by many gene,
- Genes impact phenotypes mostly through small effects
- Hence, many SNPs are likely to impact typical complex phenotypes

[Phenotype: a defined characteristic or endpoint, e.g., height, xerostomia, survival time, etc.]

Towards a complete resolution of the genetic architecture of disease

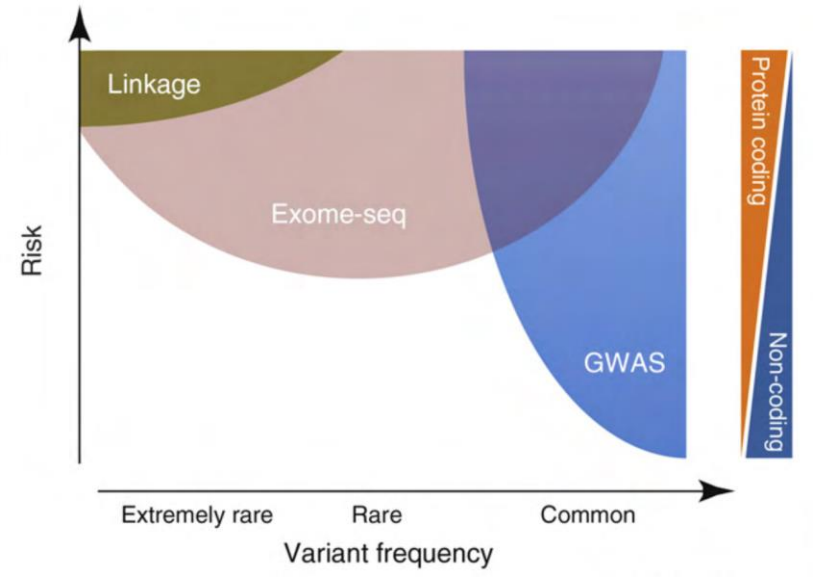
Andrew B. Singleton[1], John Hardy[2], Bryan J. Traynor[1,3] and Henry Houlden[2,4]

Trends in Genetics 26 (2010)

"...several hundred thousand SNPs throughout the genome are typed in a large series of disease cases and disease-free controls. Allele and genotype frequencies at each of these SNPs are then compared between the case group and the control group to detect alleles or genotypes that are over-represented in one group versus the other. A statistically significant association implies that there is a risk variant close to the associated SNP (or plausibly that the associated SNP is the risk variant)"

Polygenetic risk should be the expectation

Risk vs. prevalence



But…

- Large differences in radiosensitivity are **common, not rare.**
- Common alleles typically have small effect sizes
- Damage response and wound healing are complex responses

The polygenetic risk model hypothesis:

"Genetic differences in toxicity risk are likely to be highly polygenic, and unlikely to arise from a few rare alleles of high effect size."

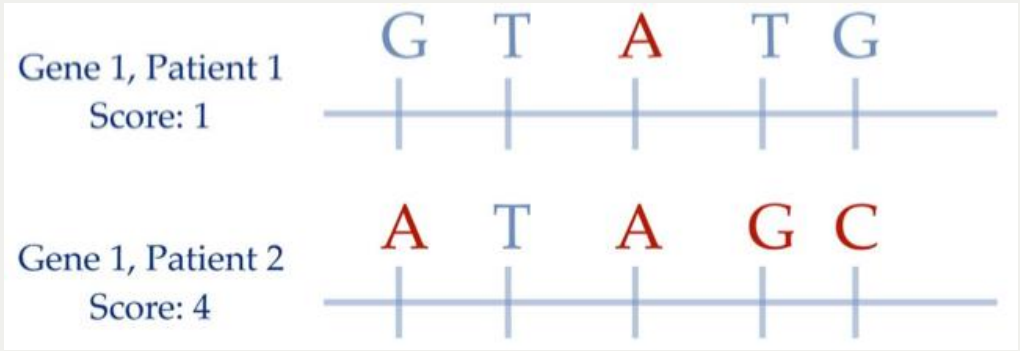
One approach: group SNPs into pathways

Pathway analysis of genome-wide data improves warfarin dose prediction

Roxana Daneshjou[1], Nicholas P Tatonetti[2], Konrad J Karczewski[1,3], Hersh Sagreiya[1], Stephane Bourgeois[4], Katarzyna Drozda[5], James K Burmester[6], Tatsuhiko Tsunoda[7], Yusuke Nakamura[7], Michiaki Kubo[7], Matthew Tector[8], Nita A Limdi[9], Larisa H Cavallari[5], Minoli Perera[10], Julie A Johnson[11], Teri E Klein[1], Russ B Altman[1,12*]

From SNP-SIG 2012: Identification and annotation of SNPs in the context of structure, function, and disease
Long Beach, CA, USA. 14 May 2012

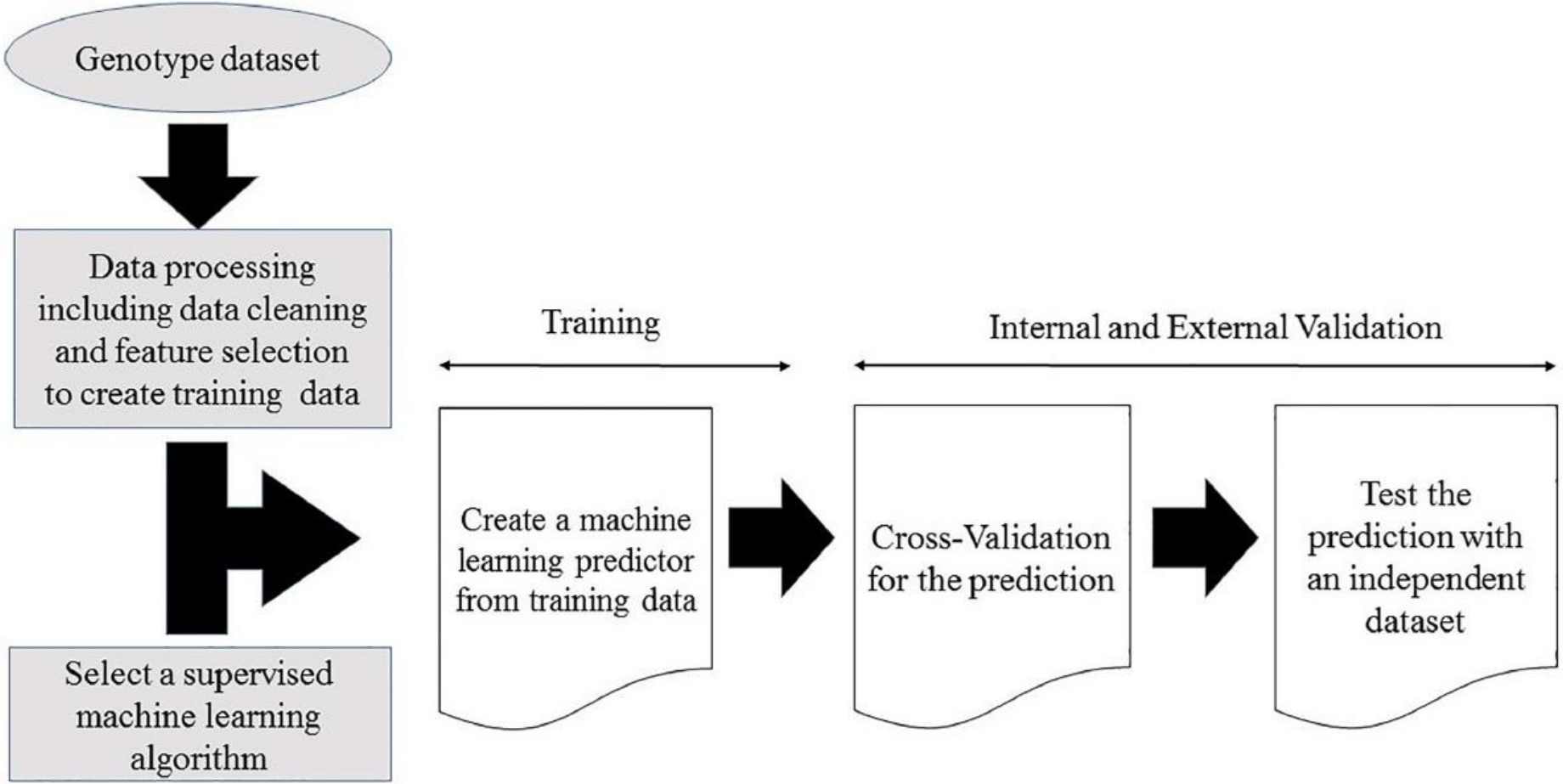
BMC Genomics 2013, **14**(Suppl 3):S11



Key points

- Method aggregates SNPs along a biologically important pathway known to affect warfarin dosing – the enzymes of its metabolic pathway.
- Focused on metabolic enzymes because of their similar direction of effect on warfarin – degradation.
- Still not highly predictive

[minor allele: nucleotide that is infrequent, yet seen in populations at that location]


Genotype dataset
Data processing including data cleaning and feature selection to create training data
Select a supervised machine learning algorithm
Training
Internal and External Validation
Create a machine learning predictor from training data
Cross-Validation for the prediction
Test the prediction with an independent dataset

SCIENTIFIC REPORTS

OPEN

Computational methods using genome-wide association studies to predict radiotherapy complications and to identify correlative molecular processes

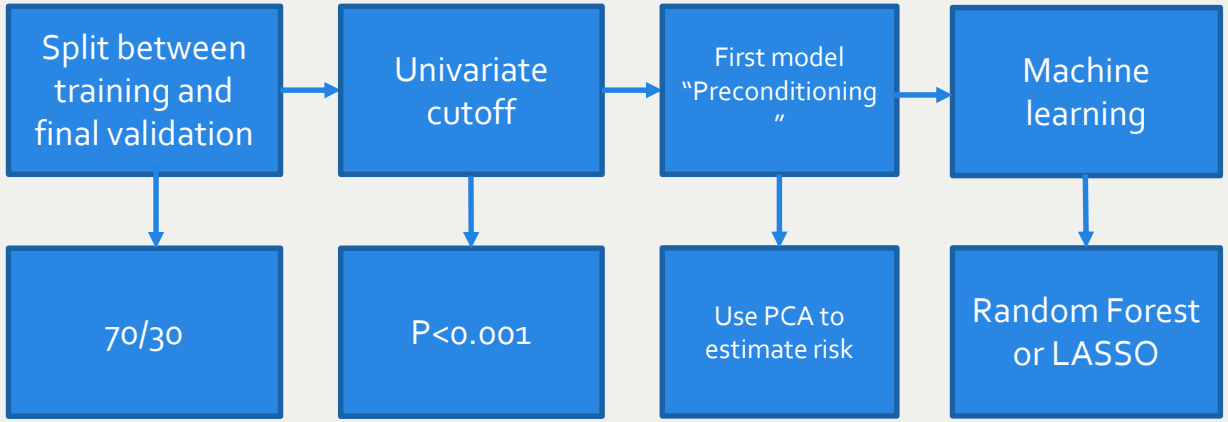
Received: 28 October 2016
Accepted: 23 January 2017
Published: 24 February 2017

Jung Hun Oh[1], Sarah Kerns[2], Harry Ostrer[3], Simon N. Powell[4], Barry Rosenstein[5] & Joseph O. Deasy[1]

Rationale

- Our goal is to predict how the risk of radiation toxicity varies between patients, based on germ line genome characteristics.
- Previous single-SNP models must overcome multiple-testing correction due to a large number of SNPs being evaluated
  - Important SNPs may fail to achieve genome-wide significance
  - Furthermore, clinical radiosensitivity is known to be a complex phenotype involving many genes.
- Therefore, we have taken a **many-SNP** approach to developing predictive models, using machine learning methods

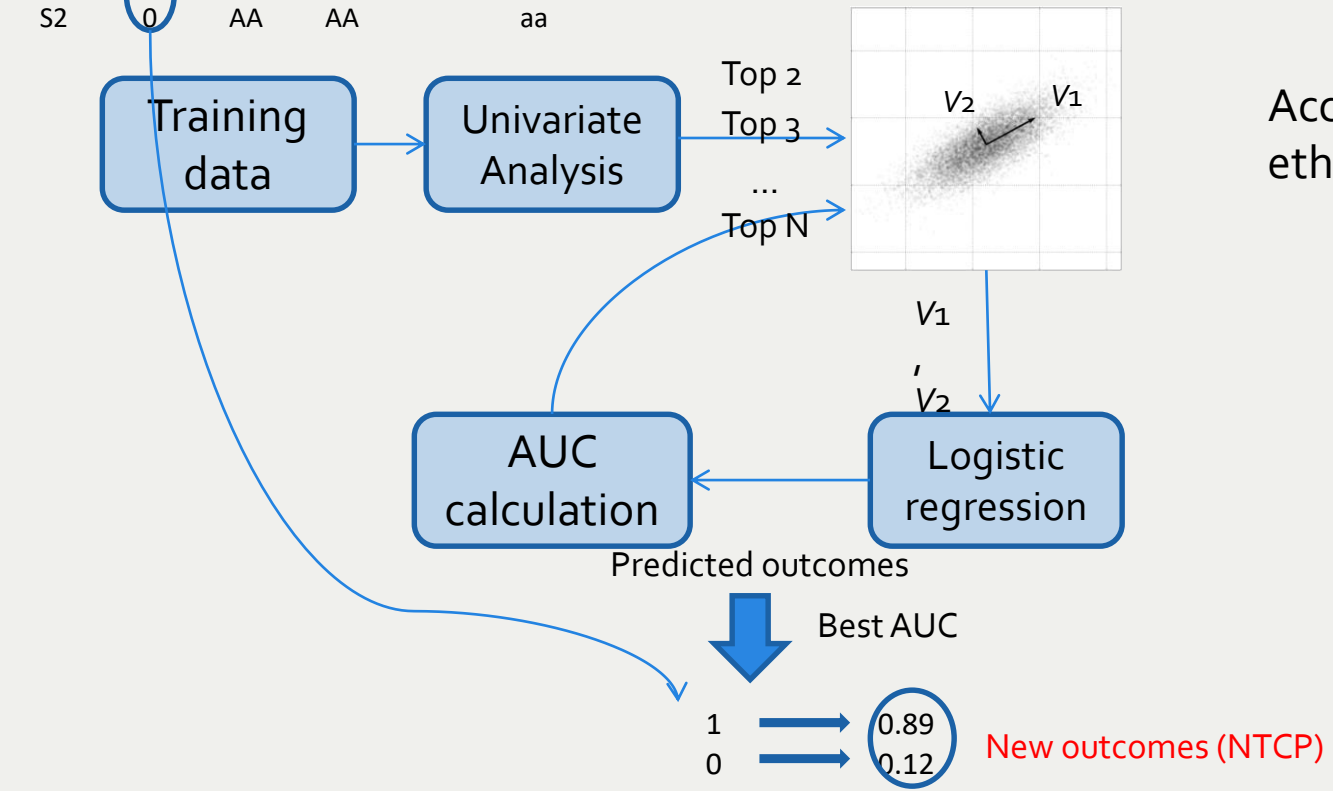
Building the model



*What is '**preconditioning**'?*
We attempt to replace the observed outcome with another outcome that takes account certain risk factors, introducing a more correct outcome ranking for the final genetic analysis. (Idea taken from Hastie and Tibshirani, The Elements of Statistical Learning)

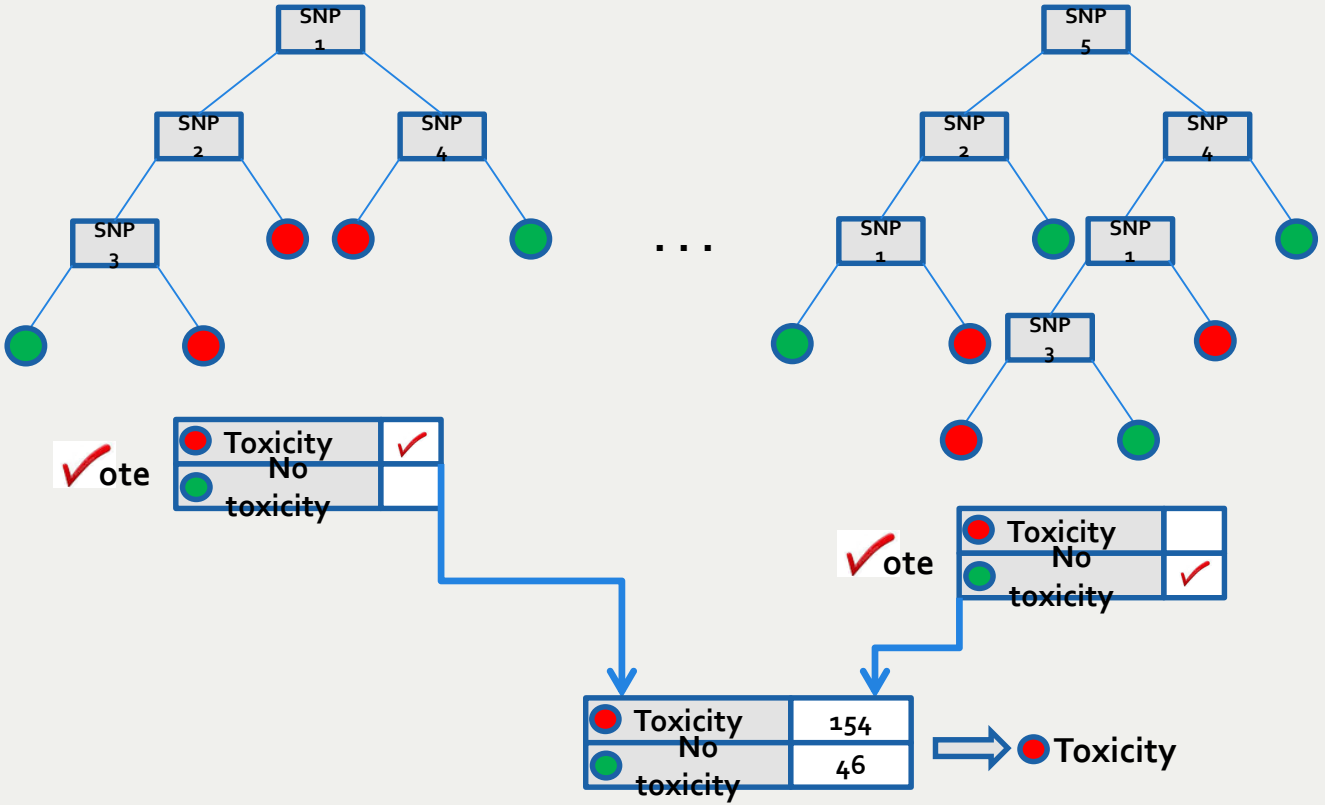
Preconditioning (1st phase) model

Sample	Toxicity	snp 1	snp 2	…	snp N
S1	1	AA	aa		Aa
S2	0	AA	AA		aa

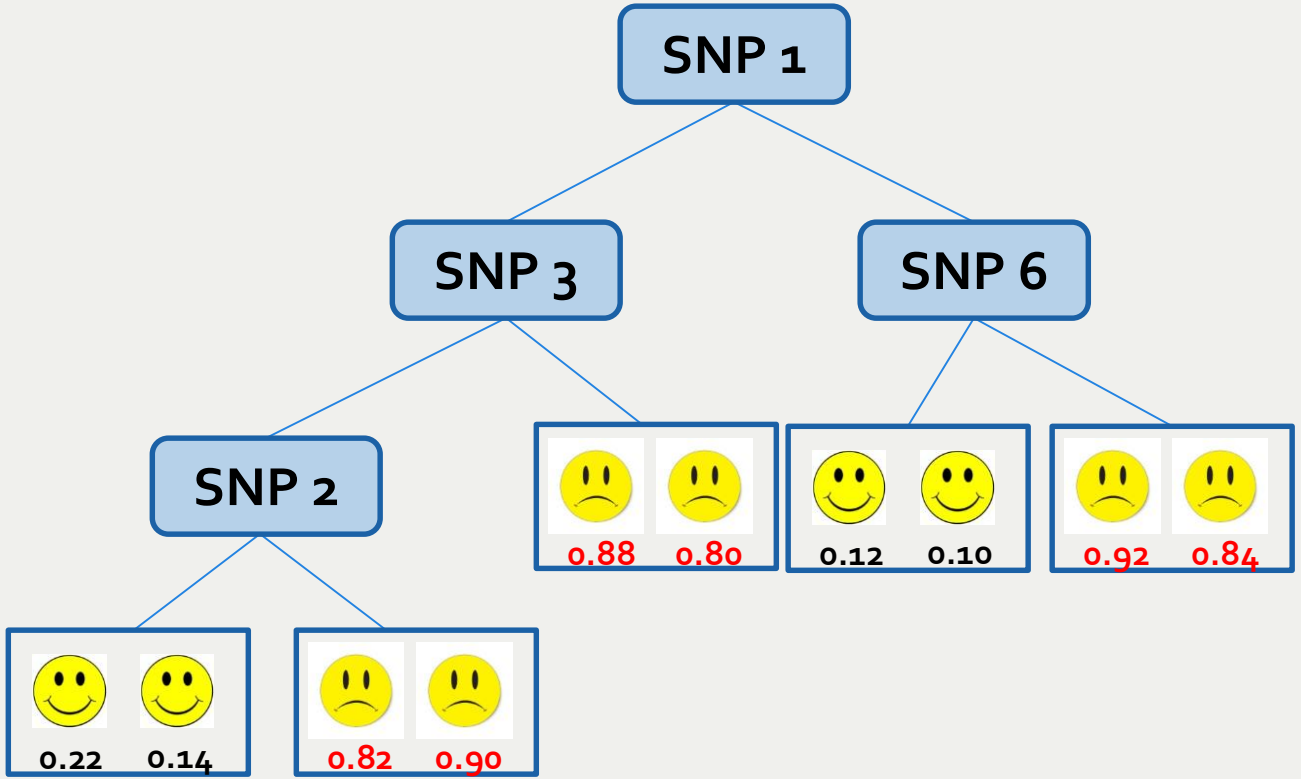


Accounts for ethnicity

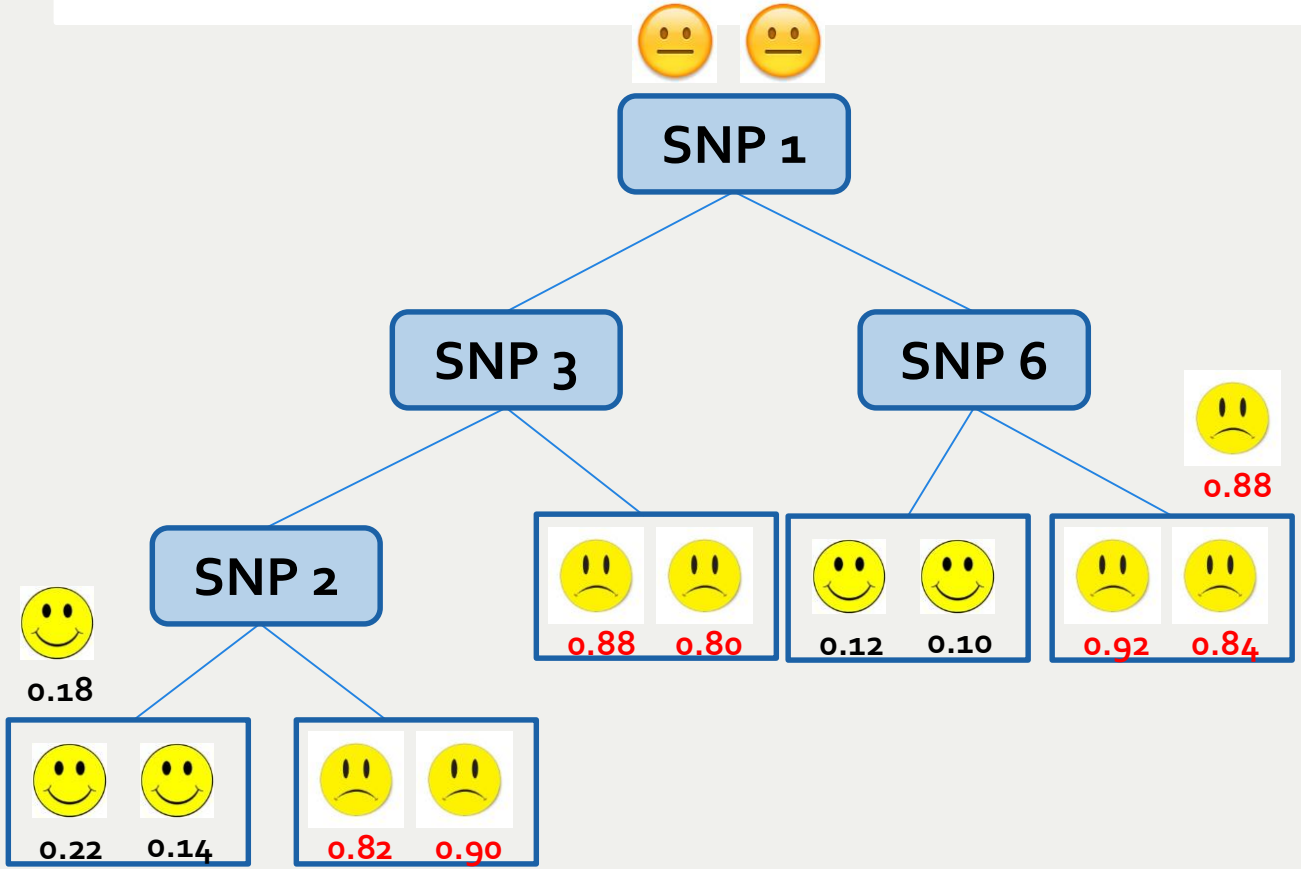
Random Forest



Random Forest - training



Random Forest - testing



Performance test

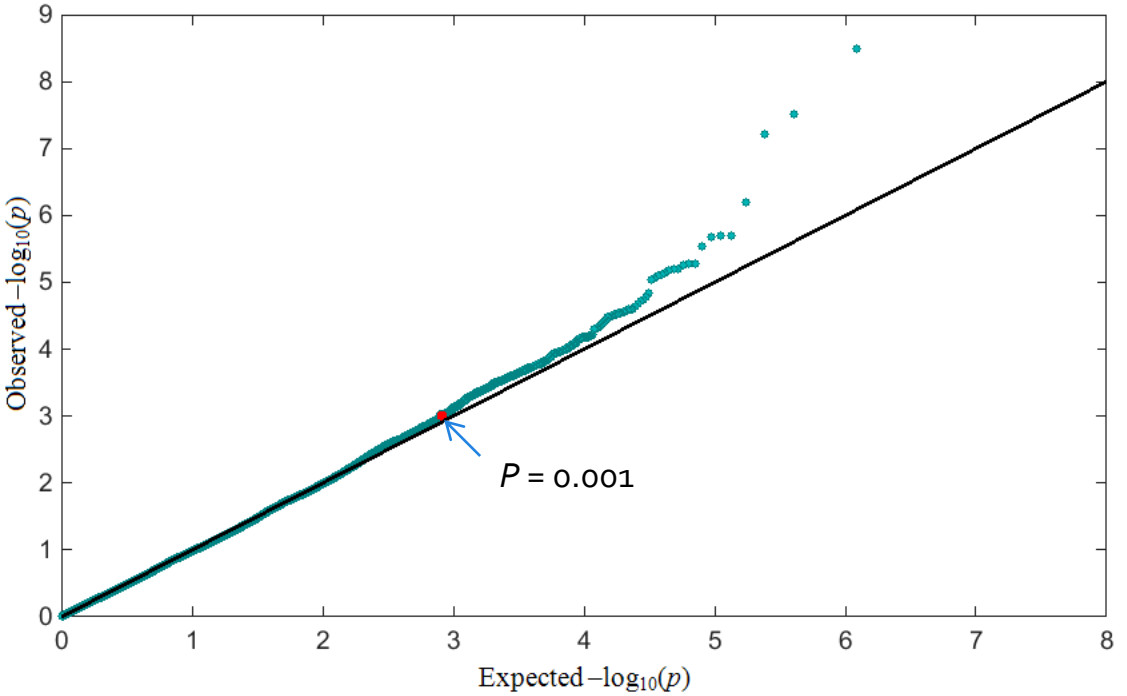
Dataset

- 368 patients with prostate cancer
  - DNA was genotyped using Affymetrix genome wide array (v6.0)
- Quality control
  - Missing rate > 5% of samples
  - MAF < 5%
  - Hardy-Weinberg equilibrium (p-value < 10^{-5})
  - 613,496 SNPs remained

Dataset for RB

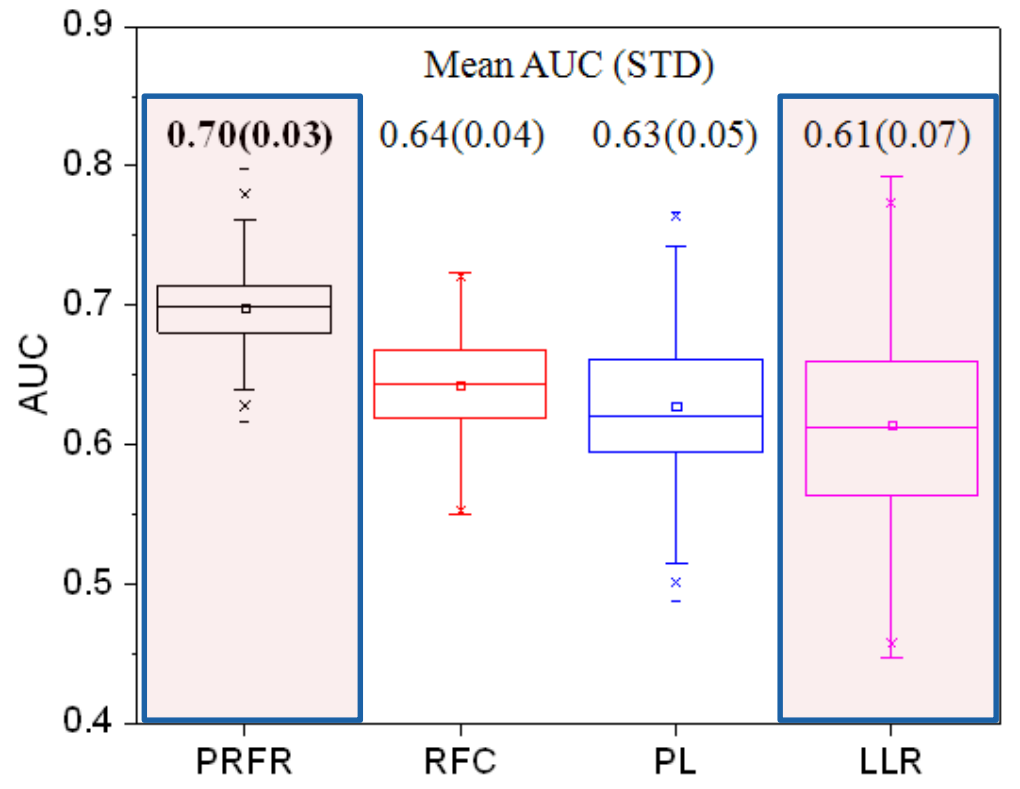
- Outcome: rectal bleeding
  - RTOG ≤ 1 (coded 0) vs RTOG ≥ 2 (coded 1)
- Data split: rectal bleeding
  - Training dataset
    - 243 samples
    - 49 events
    - 749 SNPs (p< 0.001; Chi-square test)
  - Validation dataset
    - 122 samples
    - 25 events
- 5-fold CV or bootstrapping with 100 iterations
- Additive model
  - Coded as the number of rare alleles

Q-Q plot for RB

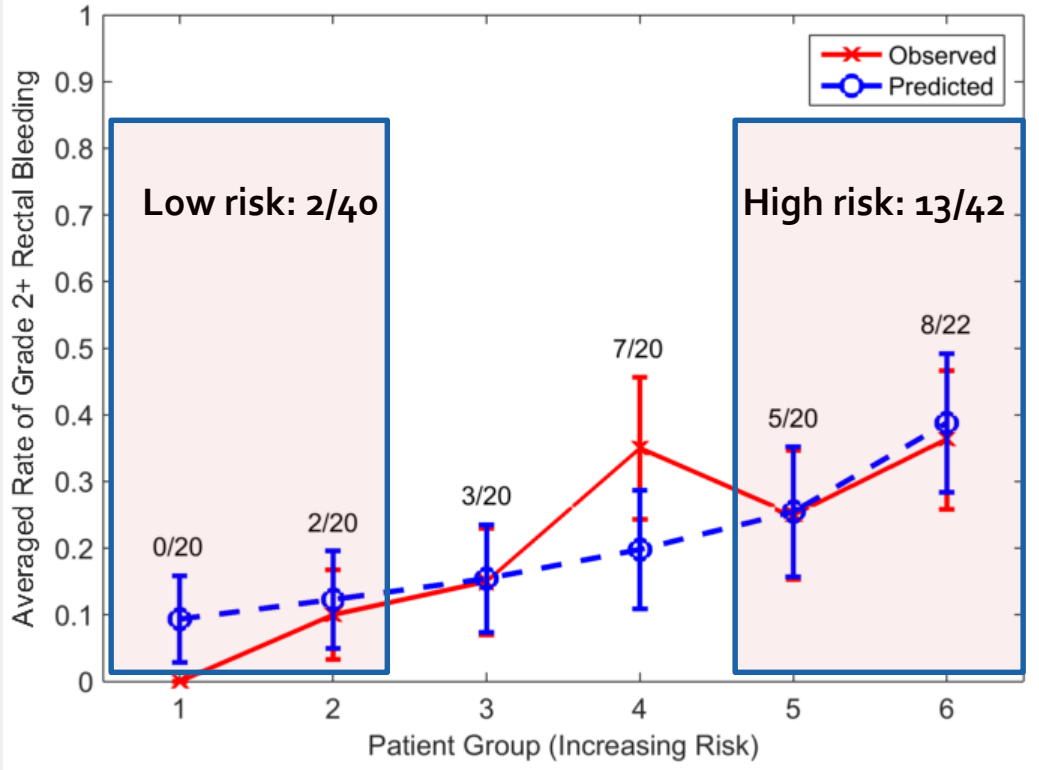


True randomness would result in correlations close to the straight line.

Model comparison for RB using validation data



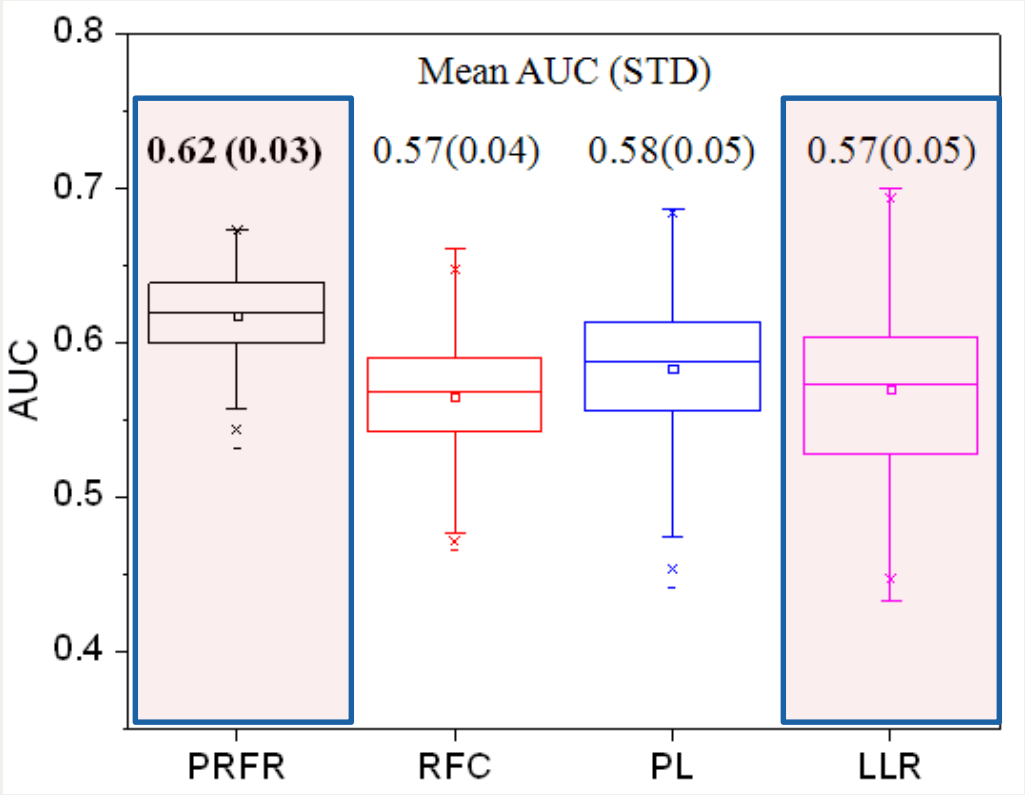
Results for RB using validation data



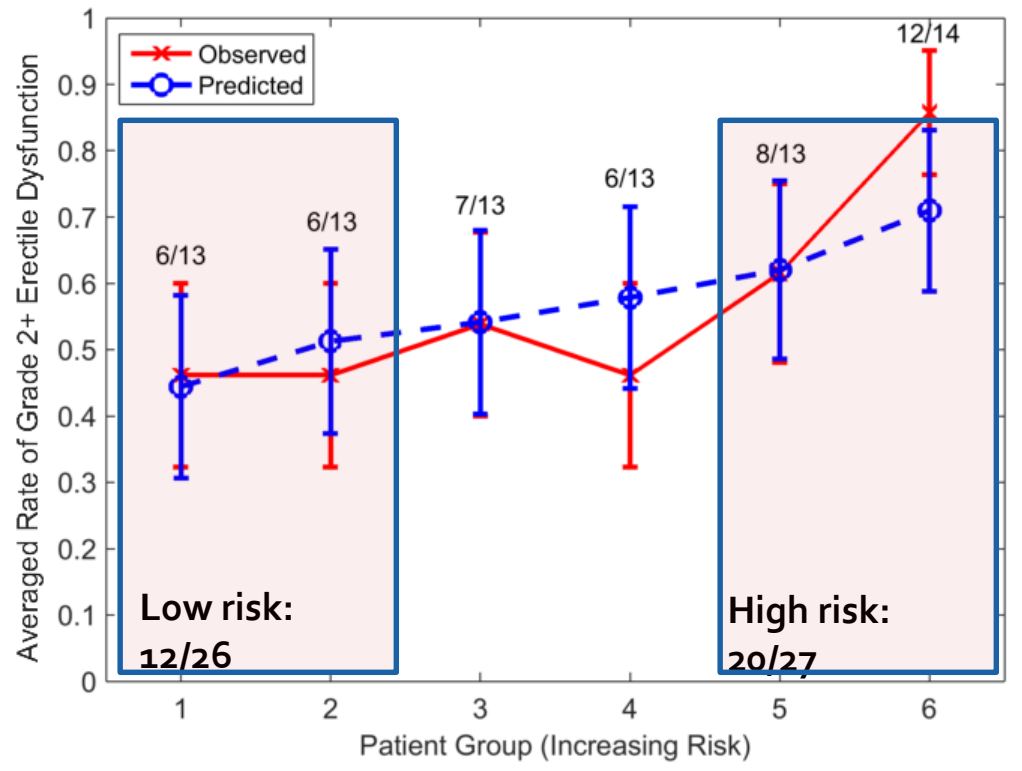
Dataset for ED

- Outcome: erectile dysfunction
  - SHIM ≤ 7 (coded 1) vs SHIM ≥ 16 (coded 0)
- Data split
  - Training dataset
    - 157samples
    - 88 events
    - 367 SNPs ($p < 0.001$; Chi-square test)
  - Validation dataset
    - 79 samples
    - 45 events

Results for ED using validation data



Results for ED using validation data

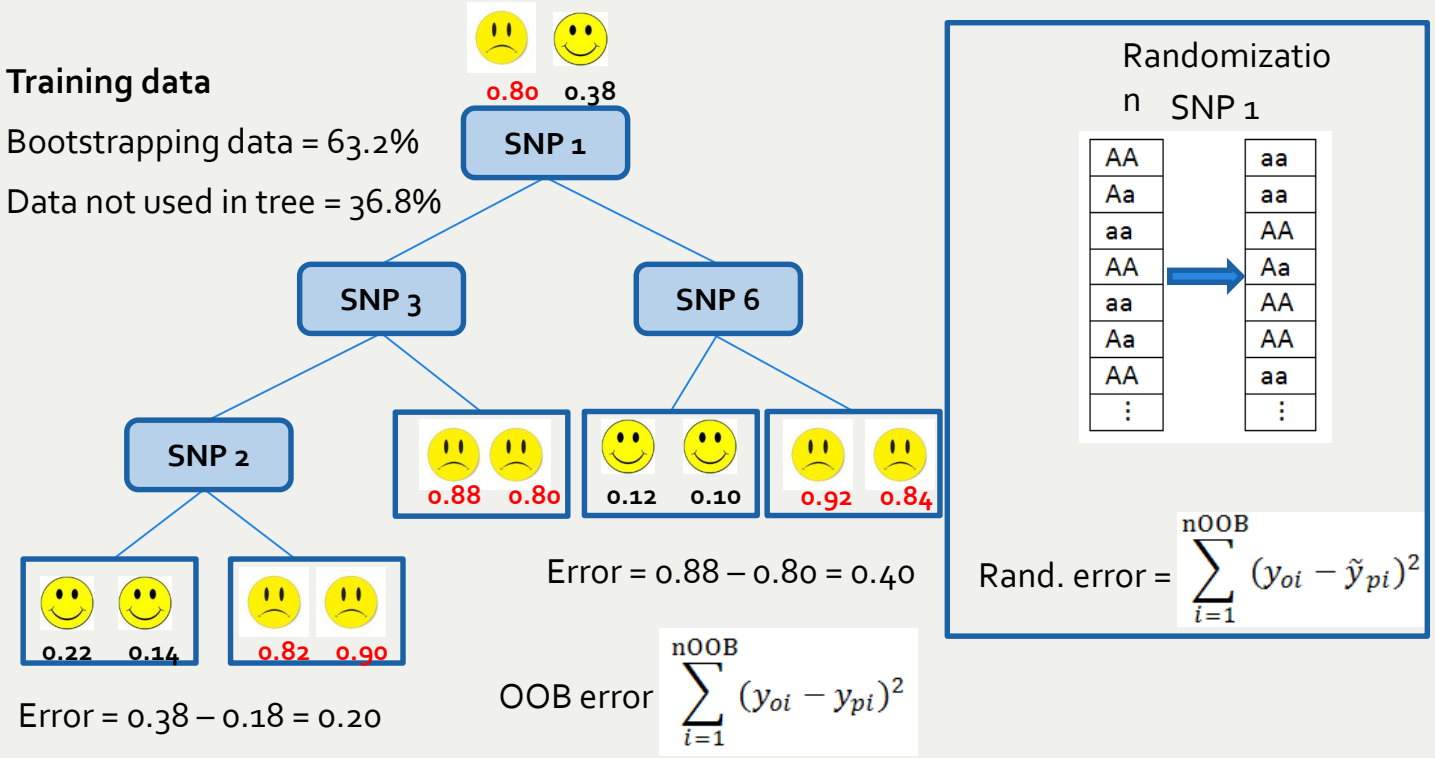


Identifying important biological processes in the resulting models

How important is any given SNP?

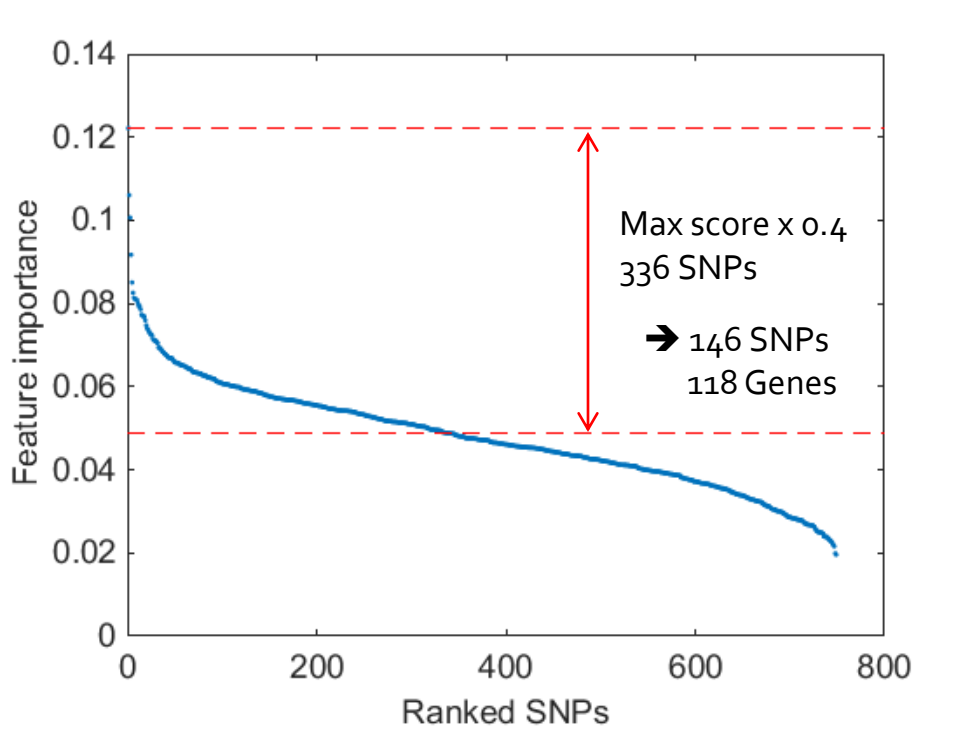
Test this for each SNP by shuffling allele values (0,1,2) between patients.

How important is any given SNP?

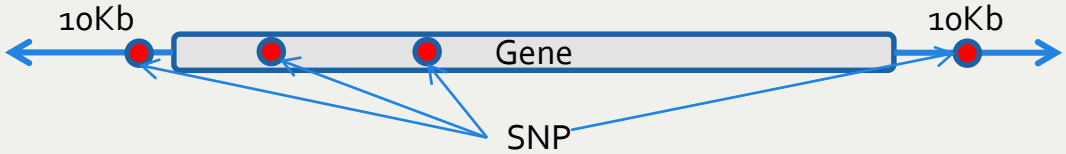


Importance of SNP 1 = (Rand. error - OOB error)/ nOOB

SNP importance for RB



Let's find any genes near important SNPs



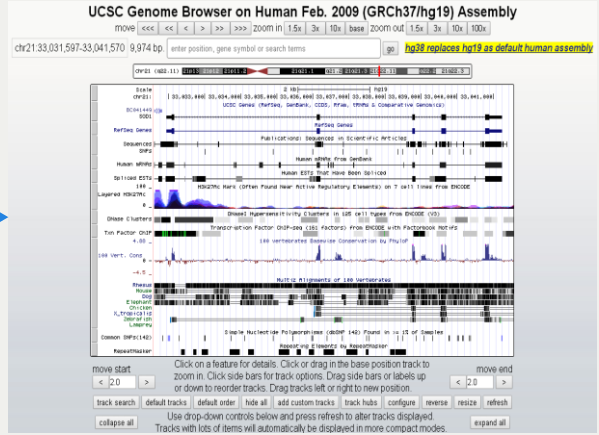
mysql client

```
C:\WINDOWS\system32\cmd.exe - mysql --user=genome --host=genome-mysql.cse.ucsc.edu -A
    -> "rs2065841",
    -> "rs11578253",
    -> "rs17471738",
    -> "rs10157706";
+------------+-----------+-------+------------+----------+---------+---------+
| rsId       | name      | chrom | chromStart | chromEnd | txStart | txEnd   |
+------------+-----------+-------+------------+----------+---------+---------+
| rs10157706 | NULL      | chr1  |   68109612 | 68109613 |    NULL |    NULL |
| rs1061615  | NM_006474 | chr1  |   13943723 | 13943724 | 13910251 | 13944452 |
| rs1061615  | NM_198389 | chr1  |   13943723 | 13943724 | 13910251 | 13944452 |
| rs1061615  | NM_001006625 | chr1  |   13943723 | 13943724 | 13911966 | 13944452 |
| rs1061615  | NM_001006624 | chr1  |   13943723 | 13943724 | 13911966 | 13944452 |
| rs10888857 | NM_147161 | chr1  |   55012358 | 55012359 | 55013806 | 55076085 |
| rs10888857 | NM_015547 | chr1  |   55012358 | 55012359 | 55013806 | 55100417 |
| rs10914991 | NULL      | chr1  |   35074859 | 35074860 |    NULL |    NULL |
| rs10916762 | NM_001039500 | chr1  |   20644751 | 20644752 | 20617411 | 20681387 |
| rs11578753 | NULL      | chr1  |   64783887 | 64783888 |    NULL |    NULL |
| rs12804736 | NM_003036 | chr1  |    2189579 |  2189580 |  2160133 |  2241652 |
| rs12934251 | NM_015557 | chr1  |    6205986 |  6205987 |  6161846 |  6240194 |
| rs16865763 | NULL      | chr1  |   31290146 | 31290147 |    NULL |    NULL |
| rs17034615 | NM_015074 | chr1  |   10311876 | 10311877 | 10270763 | 10441661 |
| rs17034615 | NM_183416 | chr1  |   10311876 | 10311877 | 10270763 | 10368655 |
| rs17471738 | NULL      | chr1  |   66956150 | 66956151 |    NULL |    NULL |
| rs2065841  | NM_001134673 | chr1  |   61728596 | 61728597 | 61547979 | 61928460 |
| rs2065841  | NM_005595 | chr1  |   61728596 | 61728597 | 61547979 | 61928460 |
| rs2065841  | NM_001145511 | chr1  |   61728596 | 61728597 | 61542945 | 61928460 |
| rs2065841  | NM_001145512 | chr1  |   61728596 | 61728597 | 61547533 | 61928460 |
| rs2065841  | NR_104180 | chr1  |   61728596 | 61728597 | 61714616 | 61791190 |
| rs2147384  | NULL      | chr1  |   35082967 | 35082968 |    NULL |    NULL |
| rs333167   | NM_001134673 | chr1  |   61717881 | 61717882 | 61547979 | 61928460 |
| rs333167   | NM_005595 | chr1  |   61717881 | 61717882 | 61547979 | 61928460 |
| rs333167   | NM_001145511 | chr1  |   61717881 | 61717882 | 61542945 | 61928460 |
| rs333167   | NM_001145512 | chr1  |   61717881 | 61717882 | 61547533 | 61928460 |
| rs333167   | NR_104180 | chr1  |   61717881 | 61717882 | 61714616 | 61791190 |
| rs4653061  | NULL      | chr1  |   35072297 | 35072298 |    NULL |    NULL |
| rs4653062  | NULL      | chr1  |   35072386 | 35072387 |    NULL |    NULL |
| rs6671294  | NULL      | chr1  |   35042497 | 35042498 |    NULL |    NULL |
| rs709188   | NULL      | chr1  |    5749082 |  5749083 |    NULL |    NULL |
| rs7521055  | NM_001038633 | chr1  |   38109657 | 38109660 | 38076950 | 38100595 |
| rs7521055  | NM_001242910 | chr1  |   38109659 | 38109660 | 38076950 | 38100595 |
| rs7521055  | NM_001242909 | chr1  |   38109659 | 38109660 | 38076950 | 38100595 |
| rs7521055  | NM_001242908 | chr1  |   38109659 | 38109660 | 38076950 | 38100595 |
+------------+-----------+-------+------------+----------+---------+---------+
35 rows in set (0.52 sec)

mysql>
```

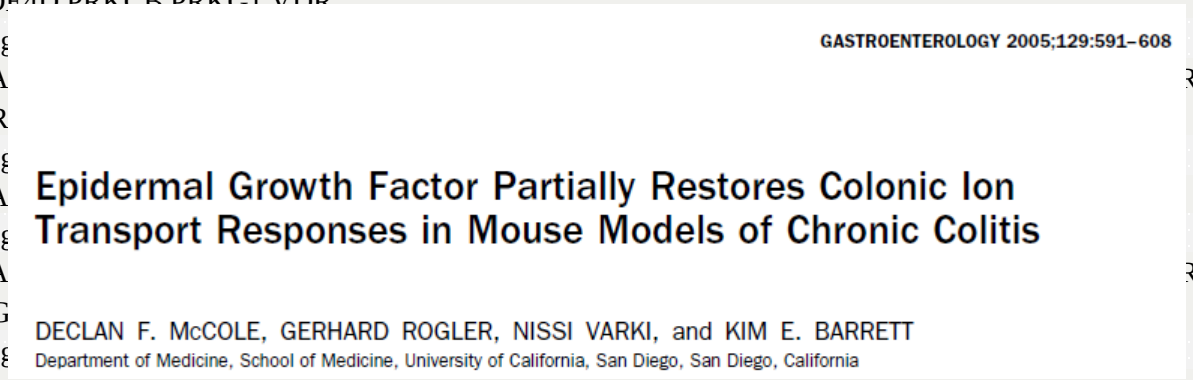
query

USCS Genome Browser



Biological processes for RB

#	GO Processes/Genes
1	Regulation of ion transport CACNA1D,CCL13,DPP6,GCK,GNB4,GPR63,HOMER1,IL1RAPL1,JDP2,KCNIP4,KCNJ6,NLGN1,NOS1AP, PDE4D,PRKCB,PRKG1,VDR
2	Reg CA DR
3	Reg CA
4	Reg CA KG
5	Reg CACNA1D,GNB4,HOMER1,NLGN1,NOS1AP,PDE4D,PRKCB,PRKG1
6	Regulation of transmembrane transport CACNA1D,CCL13,DPP6,GNB4,HOMER1,IL1RAPL1,KCNIP4,KCNJ6,NLGN1,NOS1AP,PDE4D,PRKCB,PR KG1
7	Regulation of transporter activity CACNA1D,GNB4,HOMER1,NLGN1,NOS1AP,PDE4D,PRKCB,PRKG1
8	Second-messenger-mediated signaling CACNA1D,GCK,GUCY1A2,HOMER1,JDP2,MCTP2,PDE4D,PRKG1
9	Regulation of system process CACNA1D,CTNNA2,FGG,FST,GPR63,GUCY1A2,NLGN1,NOS1AP,PDE4D,PRKCB,PRKG1,TENM4,TNR


GASTROENTEROLOGY 2005;129:591–608

Epidermal Growth Factor Partially Restores Colonic Ion Transport Responses in Mouse Models of Chronic Colitis

DECLAN F. McCOLE, GERHARD ROGLER, NISSI VARKI, and KIM E. BARRETT

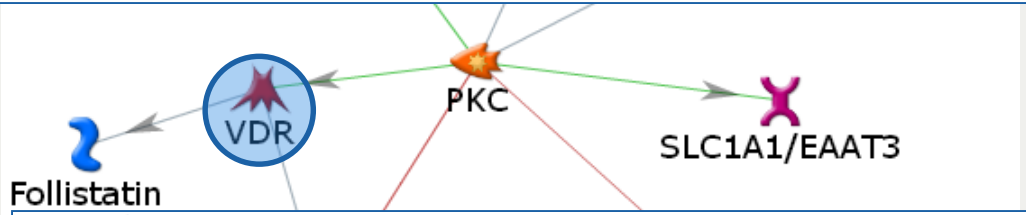
Department of Medicine, School of Medicine, University of California, San Diego, San Diego, California

Protein-protein interaction for RB

Am J Physiol Gastrointest Liver Physiol 294: G208–G216, 2008.
First published October 25, 2007; doi:10.1152/ajpgi.00398.2007.

Novel role of the vitamin D receptor in maintaining the integrity of the intestinal mucosal barrier

Juan Kong,[1] Zhongyi Zhang,[1] Mark W. Musch,[1] Gang Ning,[2] Jun Sun,[3] John Hart,[4] Marc Bissonnette,[1] and Yan Chun Li[1]



0888-8809/03/$15.00/0
Printed in U.S.A.

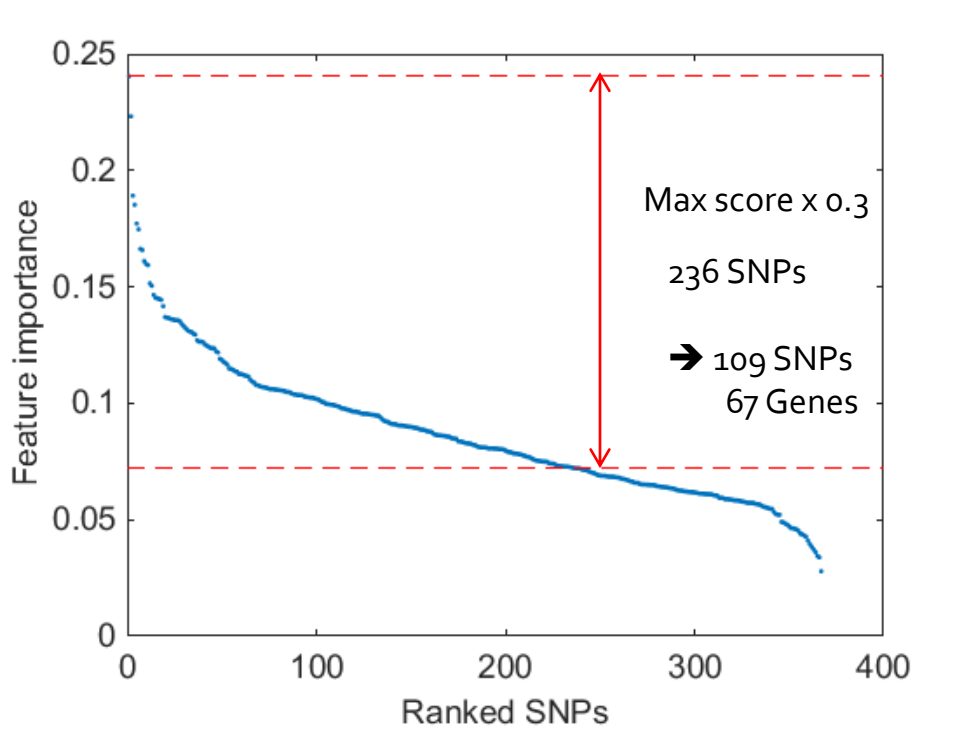
Molecular Endocrinology 17(12):2386–2392

doi: 10.1210/me.2003-0281

A Crucial Role for the Vitamin D Receptor in Experimental Inflammatory Bowel Diseases

MONICA FROICU, VERONIKA WEAVER, THOMAS A. WYNN, MARY ANN McDOWELL, JO ELLEN WELSH, AND MARGHERITA T. CANTORNA

SNP importance for ED



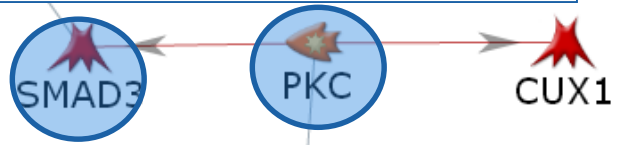
GO processes for ED

#	GO Processes/Genes
1	Negative regulation of heart contraction
	CXCR5,PDE4D,PRKCA,SPX
2	Negative regulation of blood circulation
	CXCR5,PDE4D,PRKCA,SPX
3	Neutrophil chemotaxis
	CXCR5,PDE4D,PRKCA
4	Neutrophil migration
	CXCR5,PDE4D,PRKCA
5	Granulocyte chemotaxis
	CXCR5,PDE4D,PRKCA
6	Granulocyte migration
	CXCR5,PDE4D,PRKCA
7	Regulation of blood circulation
	CXCR5,GLRX3,MAP2K1,PDE4D,PRKCA,SPX
8	Regulation of muscle system process
	CXCR5,GLRX3,MAP2K1,PDE4D,PRKCA
9	Regulation of muscle contraction
	CXCR5,MAP2K1,PDE4D,PRKCA
10	Positive regulation of cell migration
	CXCR5,DAB2IP,MAP2K1,PDE4D,PRKCA,SEMA5A,SMAD3

Protein-protein interaction for ED

Role of Increased Penile Expression of Transforming Growth Factor-β1 and Activation of the Smad Signaling Pathway in Erectile Dysfunction in Streptozotocin-Induced Diabetic Rats

Lu Wei Zhang, MD,* Shuguang Piao, MD, PhD,* Min Ji Choi, MS,* Hwa-Yean Shin, MS,* Hai-Rong Jin, MD,* Woo Jean Kim, PhD,* Sun U. Song, PhD,† Jee-Young Han, MD, PhD,‡ Seok Hee Park, PhD,‡ Mizuko Mamura, MD, PhD,§ Seong-Jin Kim, PhD,§ Ji-Kan Ryu, MD, PhD,* and Jun-Kyu Suh, MD, PhD*



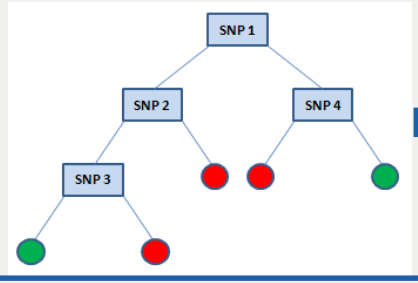
Altered Penile Vascular Reactivity and Erection in the Zucker Obese-Diabetic Rat

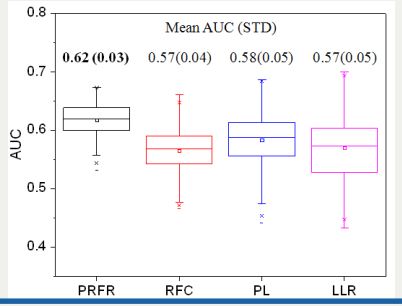
Christopher Wingard, PhD,* David Fulton, PhD,† and Shahid Husain, PhD‡

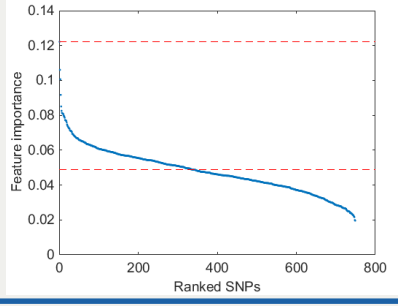
*Brody School of Medicine at East Carolina University—Physiology, Greenville, NC, USA; †Medical College of Georgia—Pharmacology, Augusta, GA, USA; ‡Medical College of South Carolina—Ophthalmology, Charleston, SC, USA

Summary

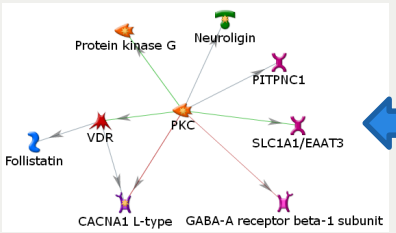
SNP level analysis







Gene level analysis



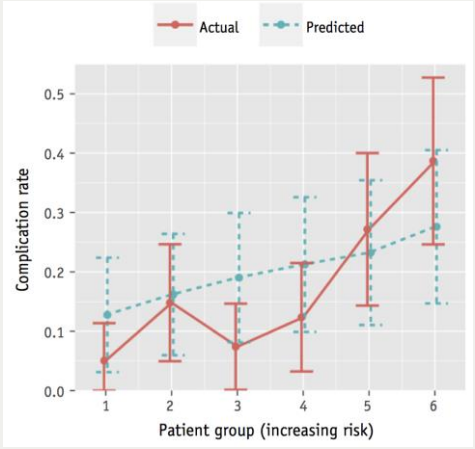
#	Processes
1	negative regulation of heart contraction
2	negative regulation of blood circulation
3	neutrophil chemotaxis
4	neutrophil migration
5	granulocyte chemotaxis

"Machine Learning on a Genome-wide Association Study to Predict Late Genitourinary Toxicity After Prostate Radiation Therapy"

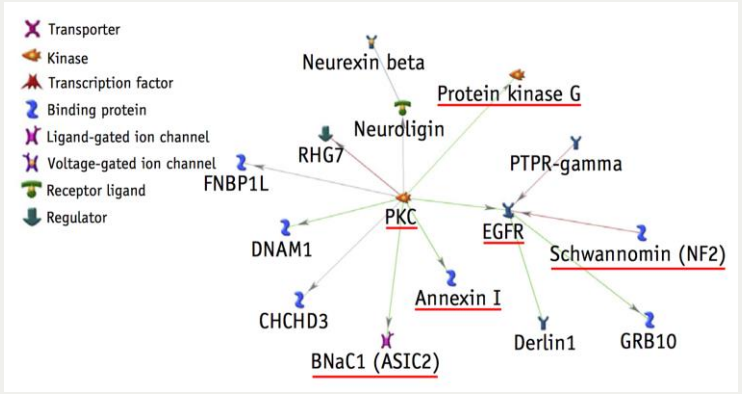
Approach:

Patient reported late (2 yr.) urinary toxicity following prostate radiotherapy (weak stream). 324 pts. (Mt. Sinai) 0.6M SNP arrays. Machine learning modeling building (PCA + Random Forests). Bioinformatics post model analysis.

Key results:



Validation of predictions in set-aside patient cohort not used for modeling.



Bioinformatics post model analysis of the most critical interacting sub-network. Red underline = previous publications related to urinary function.

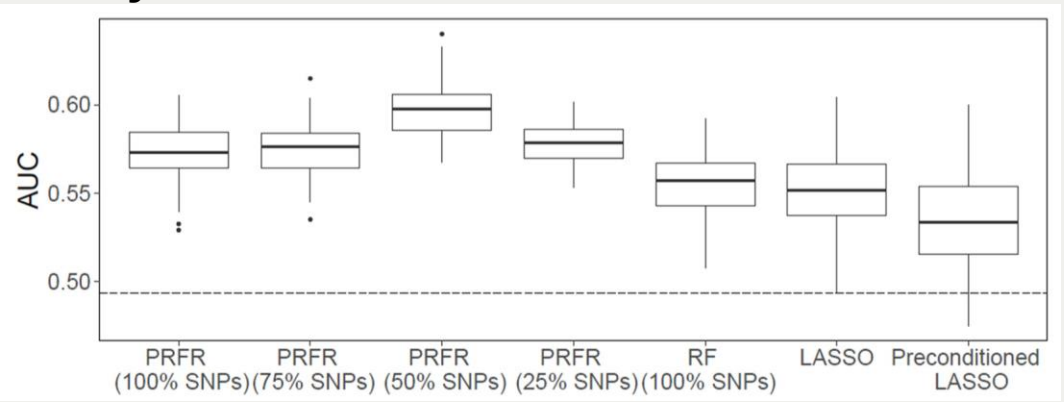
(S. Lee, S. Kerns, H. Ostrer, B. Rosenstein, J. Deasy, and J. H. Oh, Int J Radiation Oncology, Biology, Physics, early e-pub online.)

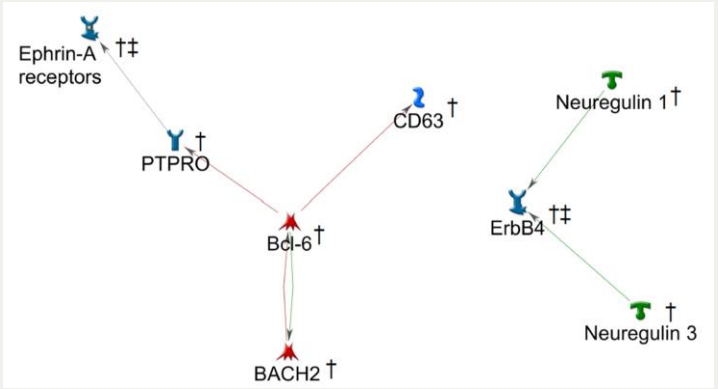
"Machine learning on genome-wide association studies to predict the risk of radiation-associated contralateral breast cancer in the WECARE Study"

Approach:

Fifty-two women with contralateral breast cancer and 153 women with unilateral breast cancer at increased risk of RCBC because they were < 40 years of age and received a scatter radiation dose > 1 Gy to the contralateral breast. Machine learning modeling building (PCA + Random Forests). Bioinformatics post model analysis.

Key results:





Bioinformatics post model analysis of connected sub-networks. Single cross- previously breast ca associated, dual cross – previously radiation induced carcinogen associated.

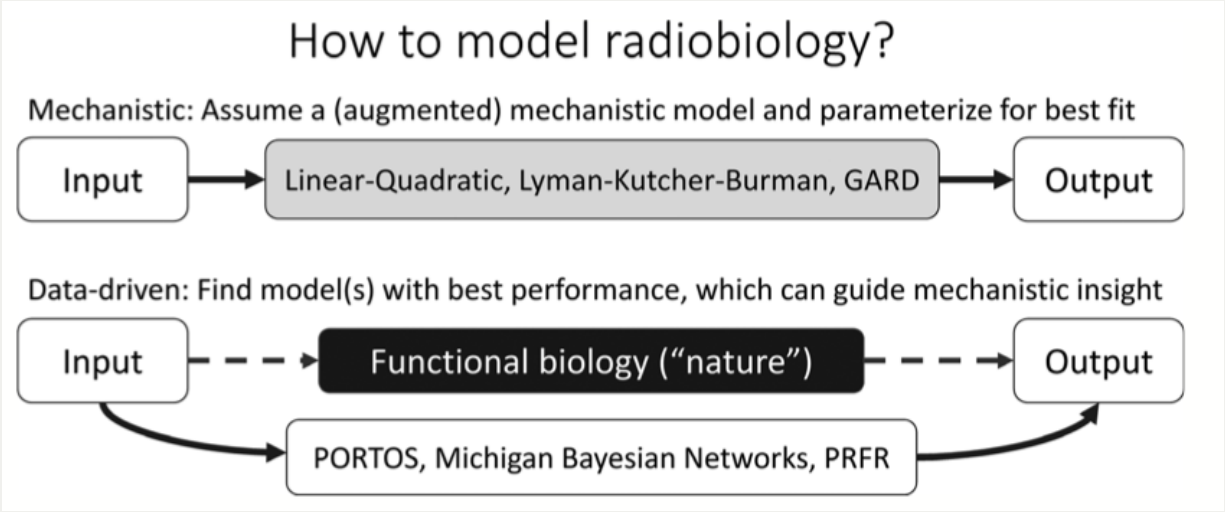
(Lee S, Liang X, Woods M, Reiner AS, Concannon P, Bernstein L, Lynch CF, Boice JD, Deasy JO, Bernstein JL, Oh JH. Plos One. 2020 Feb 27;15(2):e0226157)

Summary:

Appropriate machine learning on genomes can be a powerful tool for developing predictive models and understanding key biology…for a wide range of endpoints.

Genomics models in radiotherapy: From mechanistic to machine learning

Kang *et al.* Med. Phys. 47 (5), May 2020



$$NTCP = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{\frac{-u^2}{2}} du$$

$$t = \frac{D_{eff} - TD_{50} \cdot DMF_1 \cdot DMF_2 \cdot \ldots \cdot DMF_k}{m \cdot TD_{50} \cdot DMF_1 \cdot DMF_2 \cdot \ldots \cdot DMF_k}$$

Key points

- Groups trying to understand genetic modification of response
- Potential to put this into a Lyman-Kutcher-Berman type volume effect NTCP model as one or more Dose Modifying Factors (DMFs)

The future?

- Progress is slow due to cost and complexity of data (genomics + dose) collection.
- Data sharing is difficult – algorithm sharing more feasible.
- Many important endpoints probably do have ML-modelable genetic risk components
  - Brain radionecrosis
  - Xerostomia
  - Dysphagia
  - Etc.
- Not applicable to all endpoints – e.g. those dominated by inflammation such as pneumonitis
- Still a role for key large genetic effects (e.g. BRCA1,2, ATM, etc.)
- Definitely a role for tumor radiosensitivity variations but requires biopsy.