Identification of Candidate Regulatory SNPs by Integrative Analysis for Prostate Cancer Genome Data

Segun Jung

Sept 10, 2015

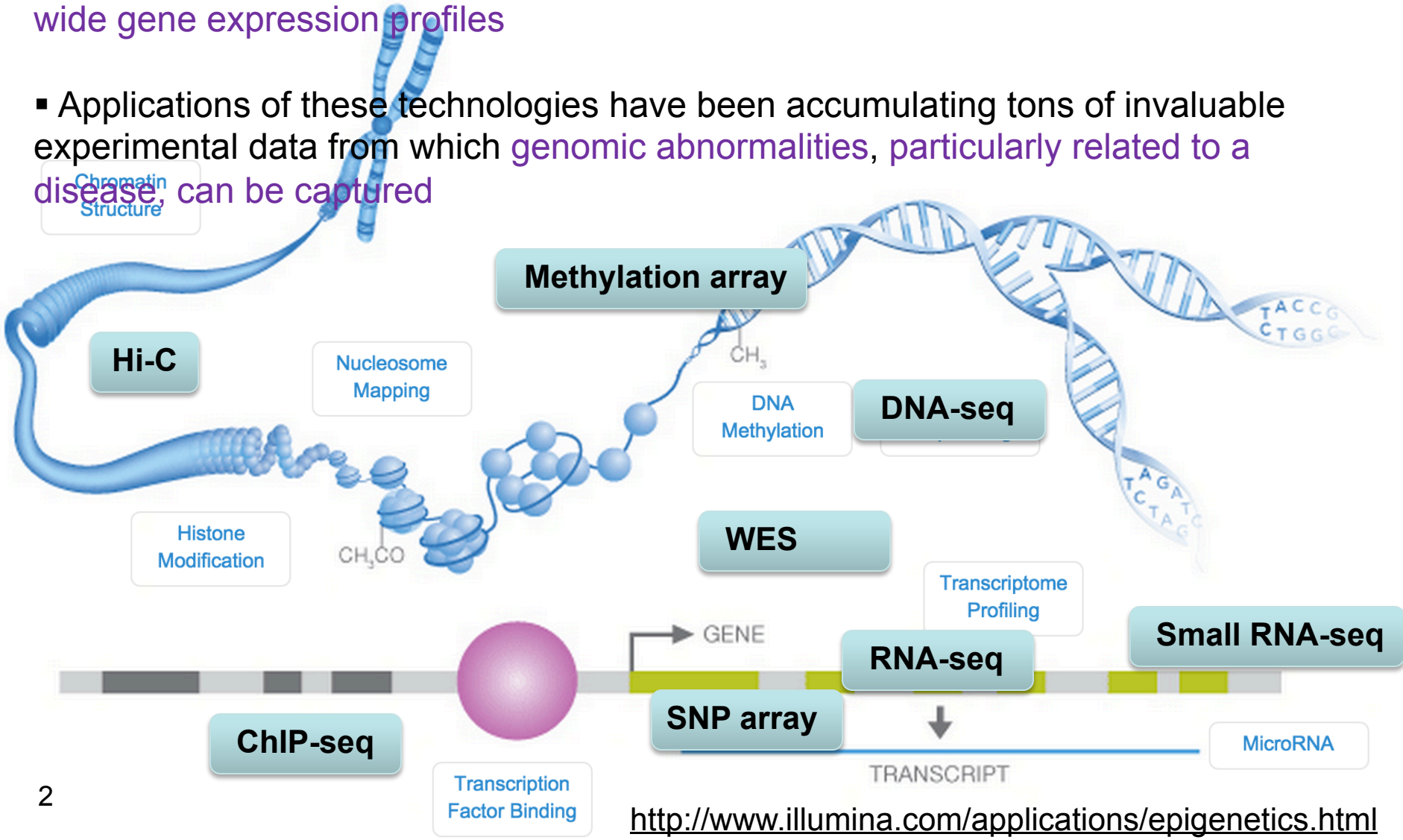
ACM-BCB 2015

Introduction

Cont.

- High-throughput technologies such as microarrays and next-generation sequencing have been extensively used to identify and characterize genome-wide gene expression profiles

- Applications of these technologies have been accumulating tons of invaluable experimental data from which genomic abnormalities, particularly related to a disease, can be captured

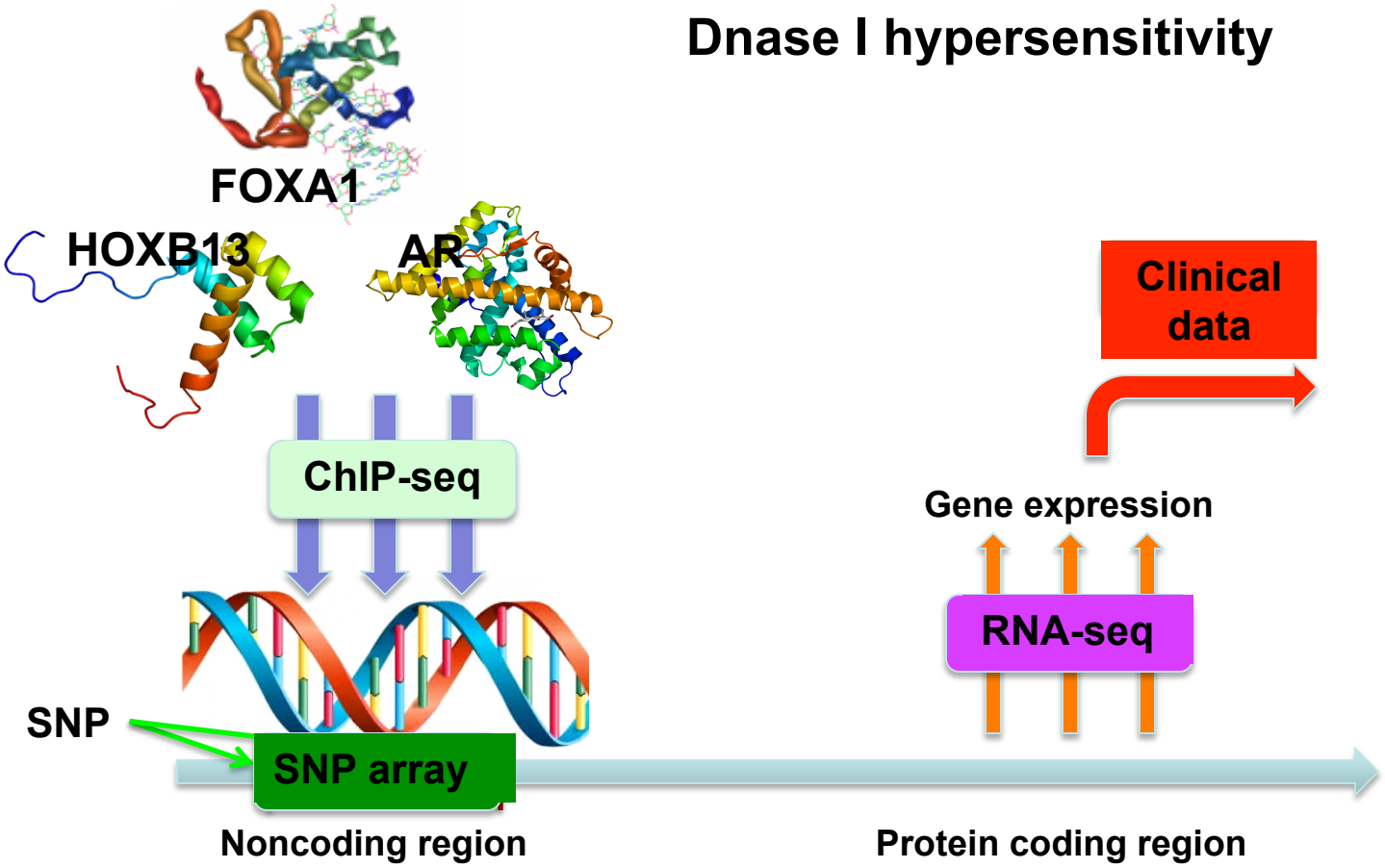


http://www.illumina.com/applications/epigenetics.html

Project design

Genes can be regulated by Transcription factors

Histon modification
Dnase I hypersensitivity

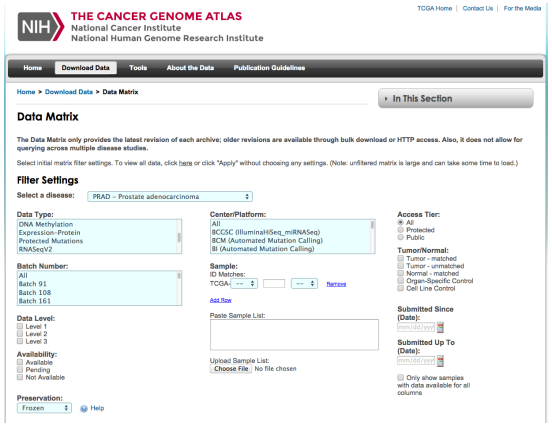


Materials and Methods

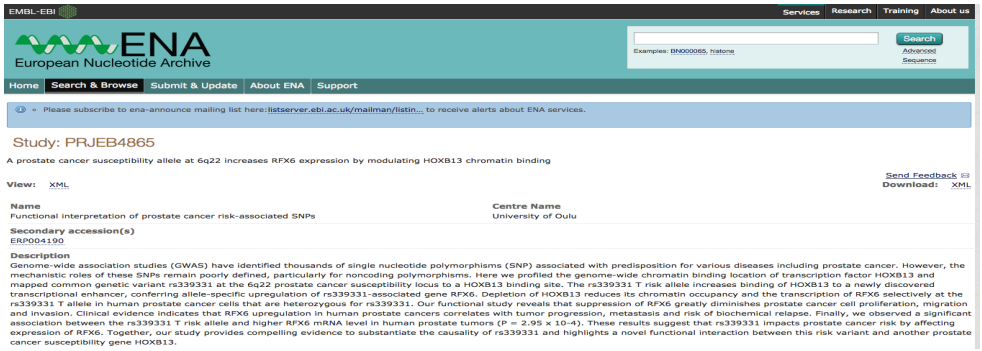
Cont.

- Regulatory SNP candidate identification
  - TCGA RNA-seq: 497 tumor and 52 matched normal samples
  - TCGA SNP array: 500 samples
  - TCGA clinical data: 369 patients
  - ChIP-seq data for the HOXB13 transcription factor

- Experimental validation
  - Microarray profiling of LNCaP control and HOX13 silencing cells



https://tcga-data.nci.nih.gov/tcga/



http://www.ebi.ac.uk/ena/data/view/PRJEB4865

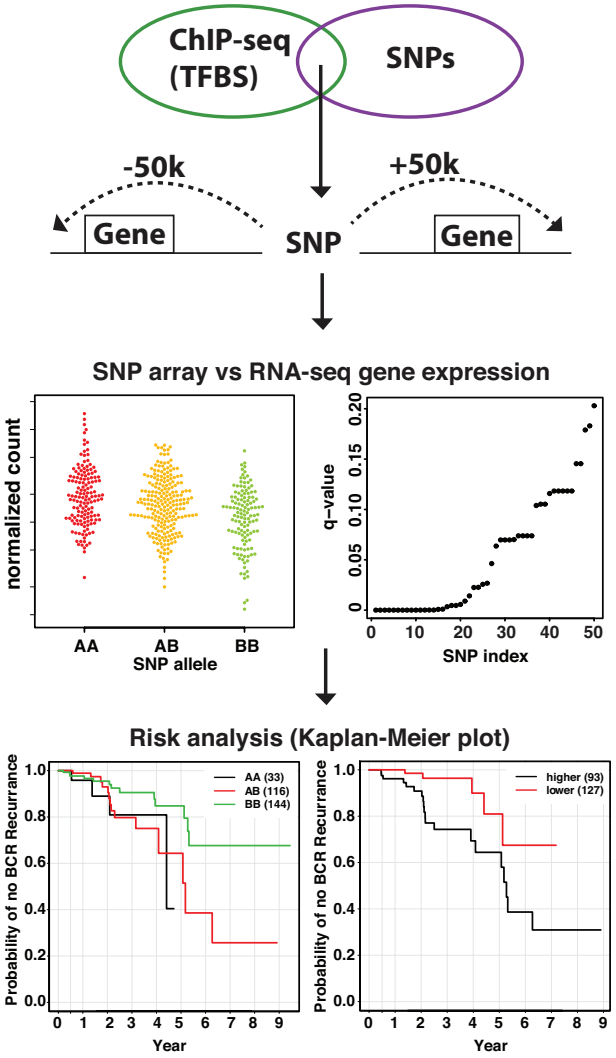
Materials and Methods

Step 1:
Find SNPs lying in a TF binding site (ChIP-seq peaks) based on genomic position

Step 2:
Search neighboring genes of each SNP

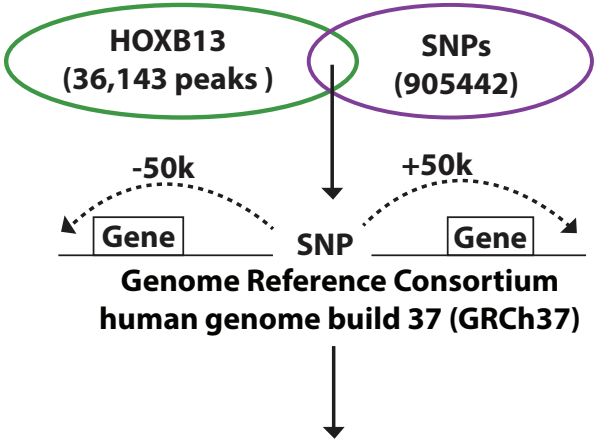
Step 3:
Identify significant allele-specific SNPs correlated with gene expression using SNP array and RNA-seq data

Step 4:
Risk analysis based on log-rank test for each SNP and gene in relation to clinical information



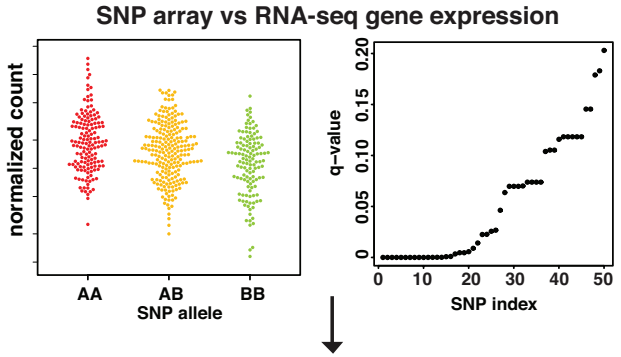
Results

Cont.

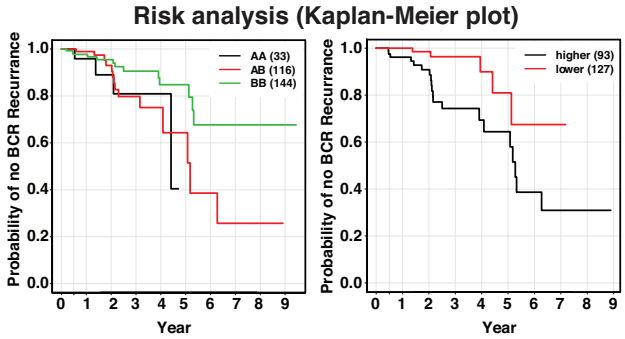


1946 SNPs

8168 SNP-gene pairs



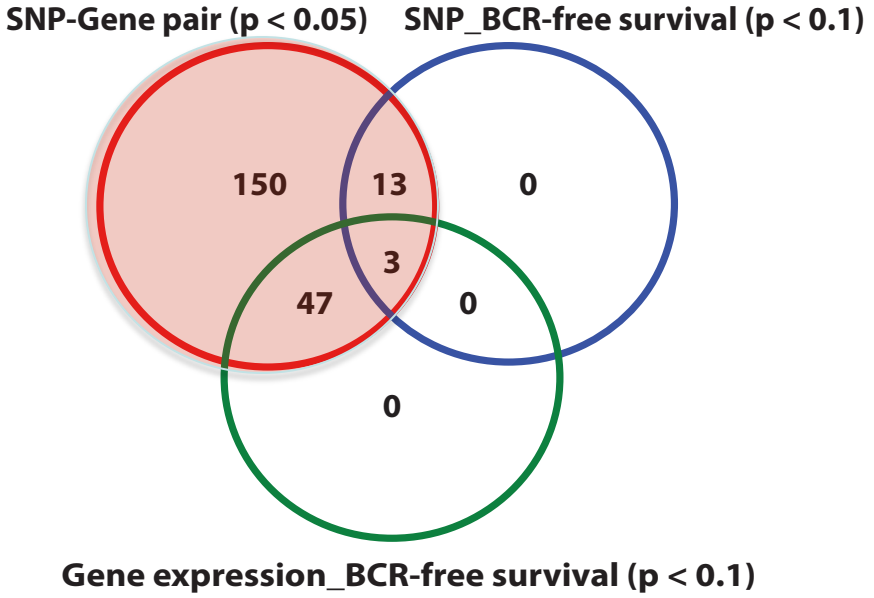
Significant SNP-gene pairs (213 by p-val < 0.05; 102 by q-val < 0.1)

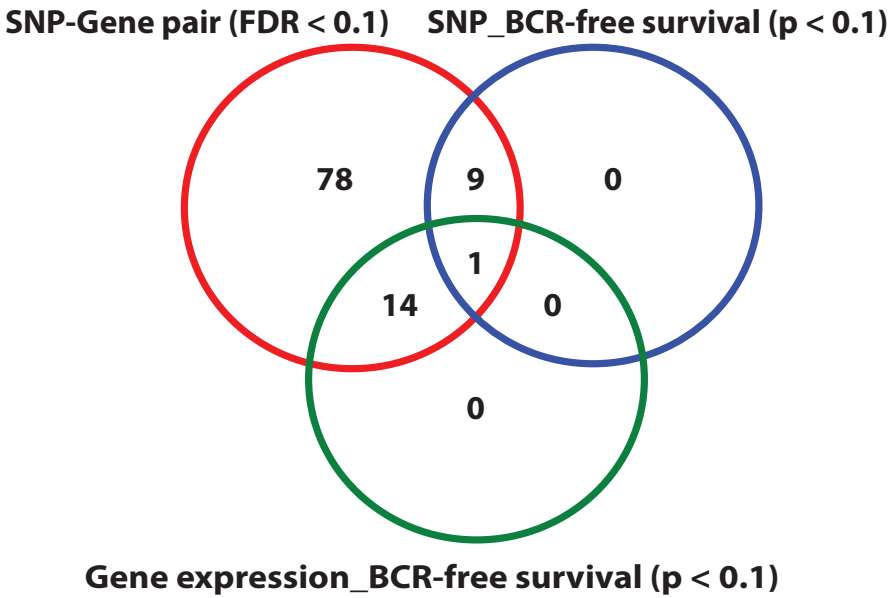


3 significant SNP-gene pairs associated with risk analysis from both SNP and gene expression

Results

Cont.





- Do our results include any eQTL?
- Any SNPs within a TF binding motif?
- Explore the top 3 SNP-gene candidates?

16 eQTLs

Cont.

SNP ID	Gene symbol	P-value	Allele A	Allele B	Allele frequency			Mean of normalized count		
					AA	AB	BB	AA	AB	BB
rs2742624	UPK3A	2.90E-46	A	G	63	202	229	2894.1	2220.5	786.8
rs2412106	CHURC1	7.95E-17	A	G	193	212	89	2170.1	2530.2	2768.8
rs1045270	WDYHV1	2.07E-13	A	G	210	218	66	722.6	579.8	514.8
rs3825393	KCTD10	2.51E-11	C	T	248	186	60	3321.6	3801.5	4391.2
rs6799720	PLOD2	1.21E-10	G	T	121	247	126	841.7	1257.3	1427.3
rs11689112	RALB	1.68E-10	A	C	244	202	48	4014.2	3536.1	2920.9
rs185397	GOT2	3.08E-10	A	G	65	182	247	7196.4	9322.1	7366
rs4325349	KRT86	4.42E-06	C	G	58	218	218	25.2	18.4	11.5
rs7894521	ECHDC3	2.61E-05	G	T	92	106	296	563.5	844.9	944.4
rs3746337	PYGB	3.45E-05	C	T	169	218	107	20172.3	18992.3	16455.2
rs10100297	MMP16	3.38E-04	C	T	97	211	186	50.2	45.2	35.8
rs3897474	GPR180	1.00E-03	A	G	200	204	90	582.1	554.1	508.8
rs11489585	RSBN1L	1.71E-03	A	G	271	187	36	698.7	778.8	836.3
rs2283119	ASAH1	8.46E-03	G	T	151	194	149	11760.6	12907.2	11621.8
rs3821747	RPL22L1	9.57E-03	A	G	315	150	29	2279.2	2896	2747.7
rs847377	AGR3	1.83E-02	C	T	202	231	61	362.3	429	487.6

- eQTL signatures from the Genotype-Tissue Expression (GTEx) portal (www.gtexportal.org/home/)

TF binding Motif Search

Cont.

rs447003, rs4796539, rs339331

Gene expression regulation

KRT6A
MED31
RFX6

SNP ID	Gene symbol	Gene name	Allele A	Allele B	Allele frequency			Mean of normalized count		
					AA	AB	BB	AA	AB	BB
rs447003	KRT6A	Keratin 6A	C	T	60	235	199	90.4	144	102
rs4796539	MED31	Mediator Complex Subunit 31	A	G	89	206	199	290	311	295
rs339331	RFX6	Regulatory Factor X, 6	T	C	263	186	45	117	69.6	22.6

ARTICLES

Nat Genet. 2014 Feb; 46(2):126-35

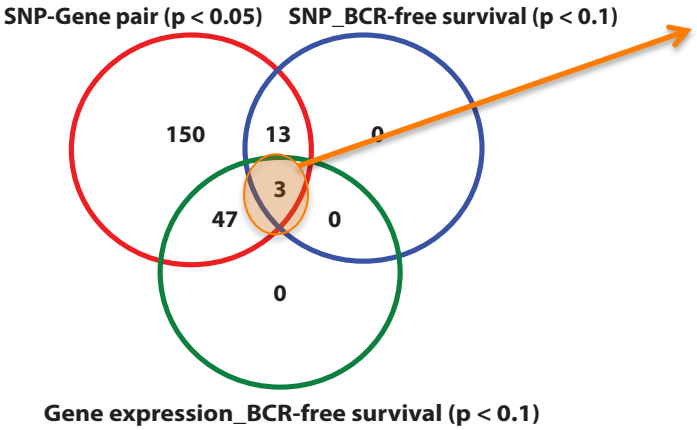
nature genetics

A prostate cancer susceptibility allele at 6q22 increases *RFX6* expression by modulating HOXB13 chromatin binding

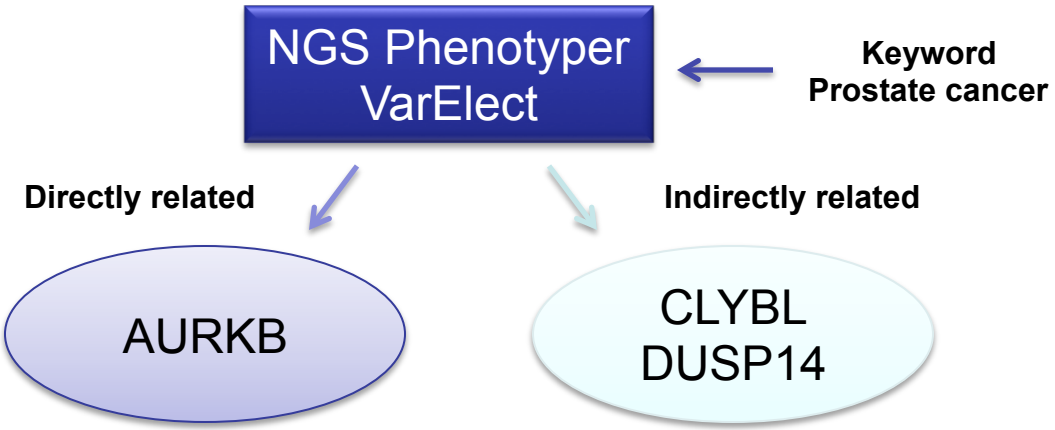
Qilai Huang[1,2,11], Thomas Whitington[3,4,11], Ping Gao[1,2], Johan F Lindberg[4], Yuehong Yang[1,2], Jielin Sun[5], Marja-Riitta Väisänen[6], Robert Szulkin[4], Matti Annala[7], Jian Yan[3], Lars A Egevad[8], Kai Zhang[1,2], Ruizhu Lin[1,2], Arttu Jolma[3,9], Matti Nykter[7], Aki Manninen[1,2], Fredrik Wiklund[4], Markku H Vaarala[6,10], Tapio Visakorpi[7], Jianfeng Xu[5], Jussi Taipale[3,9] & Gong-Hong Wei[1,2]

Genome-wide association studies have identified thousands of SNPs associated with predisposition to various diseases, including prostate cancer. However, the mechanistic roles of these SNPs remain poorly defined, particularly for noncoding polymorphisms. Here we find that the prostate cancer risk-associated SNP rs339331 at 6q22 lies within a functional HOXB13-binding site. The risk-associated T allele at rs339331 increases binding of HOXB13 to a transcriptional enhancer, conferring allele-specific upregulation of the rs339331-associated gene *RFX6*. Suppression of *RFX6* diminishes prostate cancer cell proliferation, migration and invasion. Clinical data indicate that *RFX6* upregulation in human prostate cancers correlates with tumor progression, metastasis and risk of biochemical relapse. Finally, we observe a significant association between the risk-

Top 3 Candidates Analysis Cont.



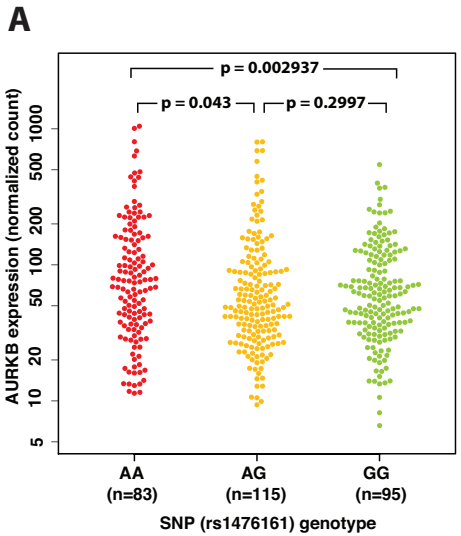
rs7992643	CLYBL	Citrate Lyase Beta Like
rs12938215	DUSP14	Dual Specificity Phosphatase 14
rs1476161	AURKB	Aurora Kinase B

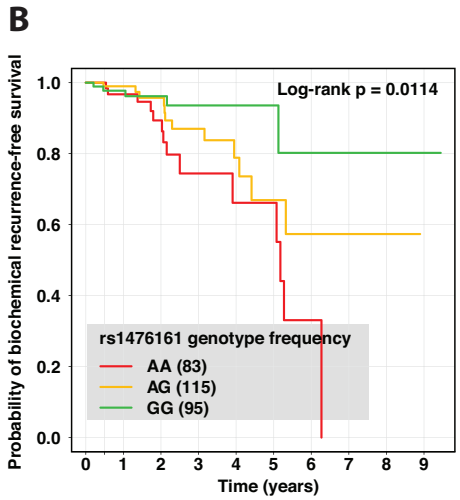


- **Aurora B is regulated by acetylation/deacetylation during mitosis in prostate cancer cells (FASEB J. 2012; 26(10):4057-67)**
- **Gene expression of Aurora kinases in prostate cancer and nodular hyperplasia tissues (Med Princ Pract. 2013; 22(2):138-43)**
- **Enhanced radiosensitivity of androgen-resistant prostate cancer: AZD1152-mediated Aurora kinase B inhibition (Radiat Res. 2011; 175(4):444-51)**

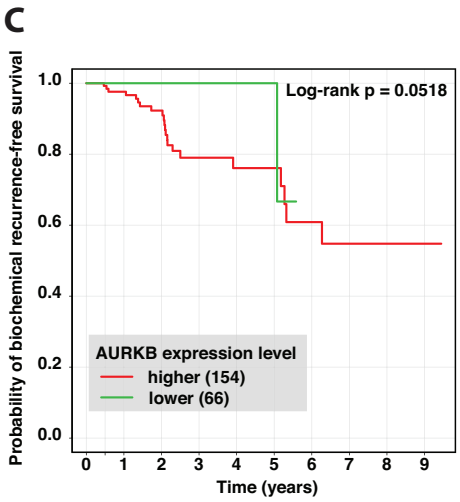
Top 3 Candidates Analysis

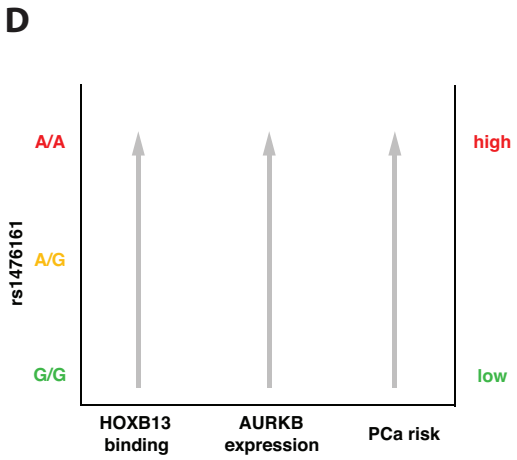
Cont.





- **Normalized mean read counts**
  AA (131), AG (95), GG (85)

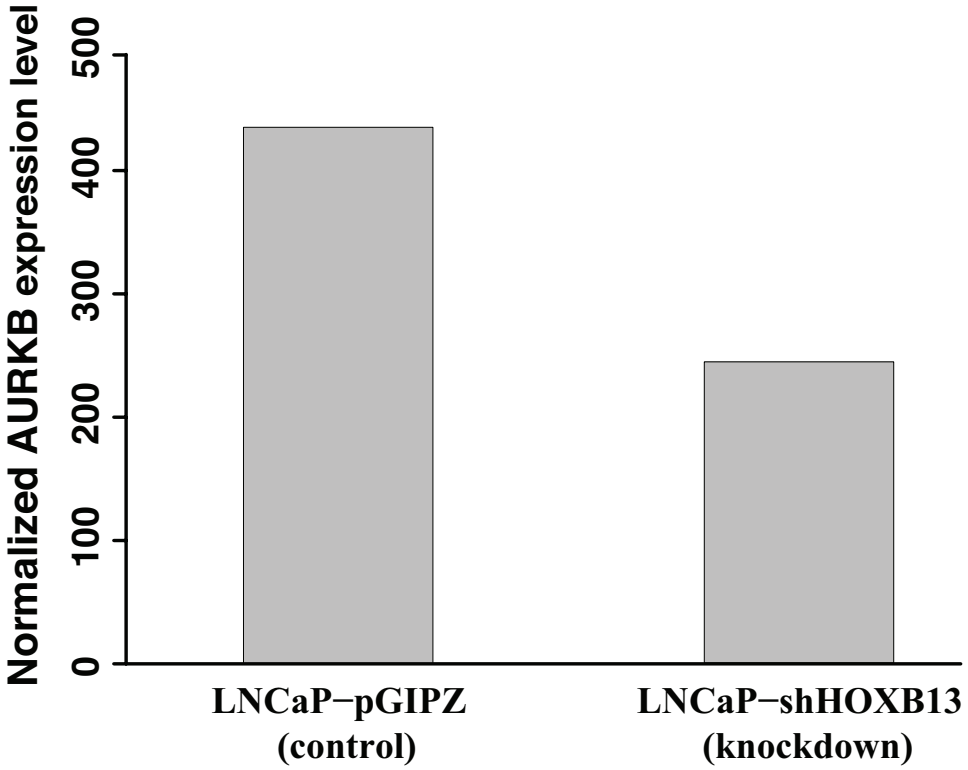




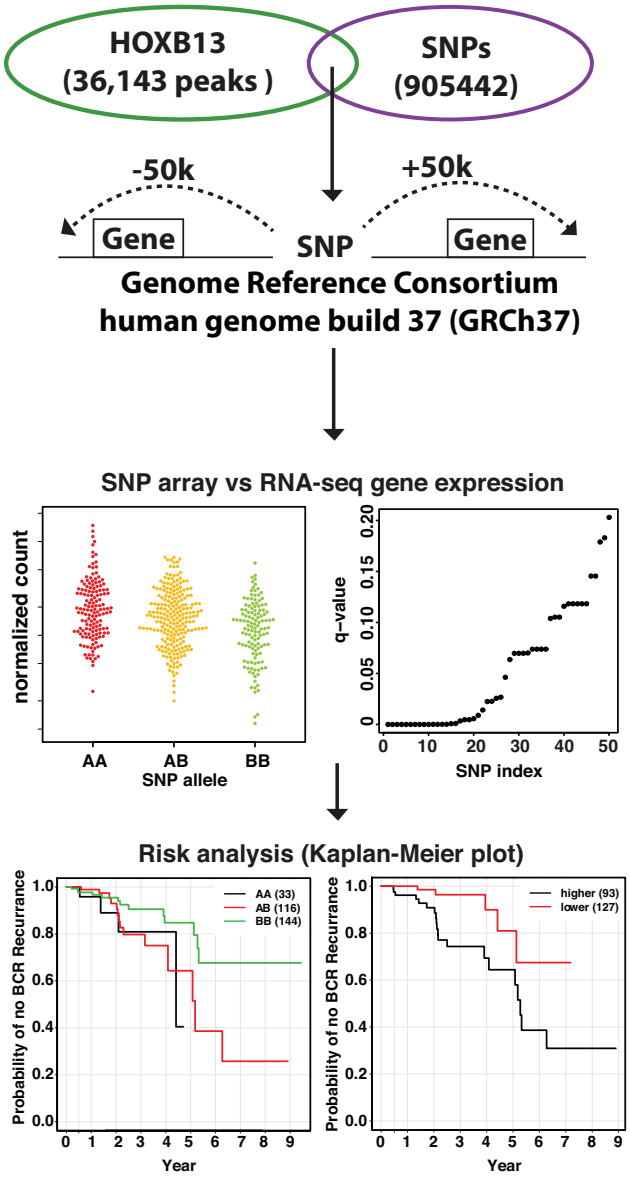
- **Allele frequency (dbSNP)**
  A: 56.290% (2819/5008)
  G: 43.710% (2189/ 5008)

Experimental Validation

- LNCaP-pGIPZ and LNCaP-shHOXB13 are the control and HOXB13 repressed cell, respectively.
- Knockout of HOXB13 diminishes AURKB gene expression level by about 2-fold.



Future work



- Different TFs for ChIP-seq
- Intersecting peak calls with other signals (e.g., H3K27ac, Dnase I hypersensitivity) to improve potential regulatory regions
- Incorporate external PCa dataset for validation
- Apply this scheme to other disease

Conclusions

- We presented an *in silico* methodology in conjunction with an experimental validation for identifying candidate regulatory SNPs located in the TF-bound noncoding regions

- We identified a novel rSNP and its target gene pair (rs1476161, AURKB) as a potential biomarker in PCa

- The method is not only suitable for prostate cancer, but for any other cancer types.

Acknowledgements

Dr. Ramana V. Davuluri (PI)

Dr. Hongjian Jin

Northwestern University
NUCATS
Clinical and Translational Sciences Institute

NATIONAL INSTITUTES OF HEALTH