Identification of Functional Lung Cancer Risk-Associated SNPs Through Examination of Allele-Specific Effects

C Yang[1], TR Stueve[1], E Tran[1], C Yan[1], CN Marconett[1], B Zhou[1], Z Borok[1], KD Siegmund[2], IA Laird-Offringa[*, 1]

[1]Norris Comprehensive Cancer Center, [2] Dept of Preventive Medicine, University of Southern California

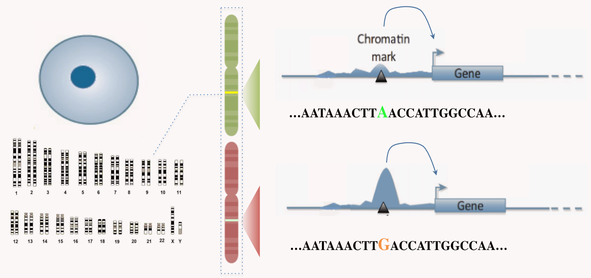
USC Norris Comprehensive Cancer Center
Part of the Keck School of Medicine of USC

Background

To date, a number of SNPs have been found to be associated with lung cancer risk through genome-wide association studies (GWASs). However, since most GWAS SNPs lie in non-coding regions and are co-inherited with hundreds of SNPs in linkage disequilibrium (LD), which SNP(s) play a causal role in the disease remains poorly understood.

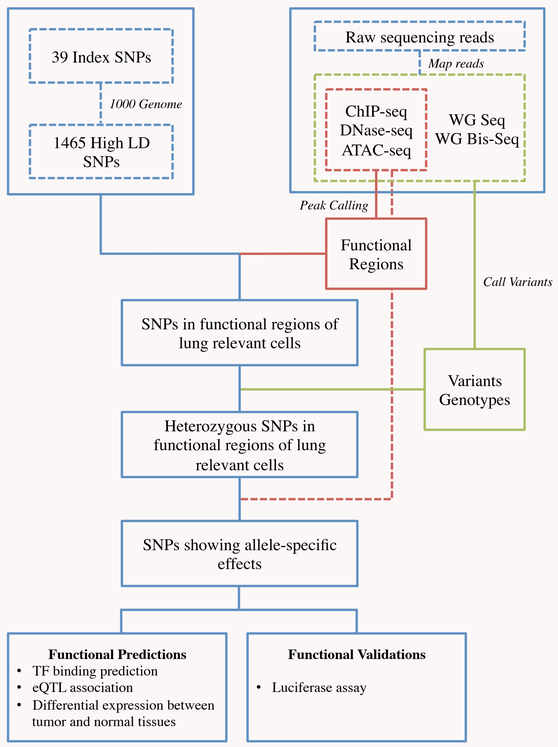
We aimed to identify causal SNPs associated with lung cancer risk by investigating allele-specific effects (ASE) in epigenetic marks.

Schematic plot of allele-specific effects (ASE)



Methods

Research Workflow

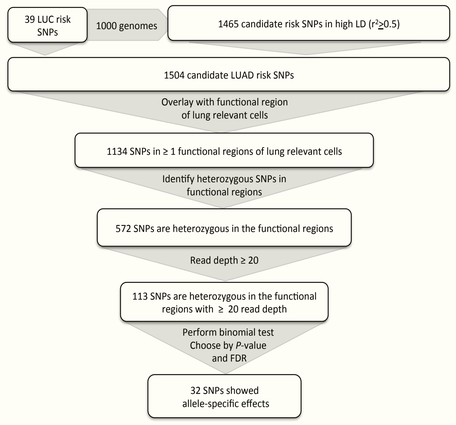


NGS data sources

ILO lab	ENCODE	Roadmap Epigenomics	DBTSS	GEO
• AEC • Basal Cells	• A549 • SAEC • NHLF • HPF • IMR90	• IMR90 • 24 Fetal Lung • 2 Human Lung	• 26 LUAD cell lines • SAEC • BEAS2B	• H1819 • H2087 • H3122

Results

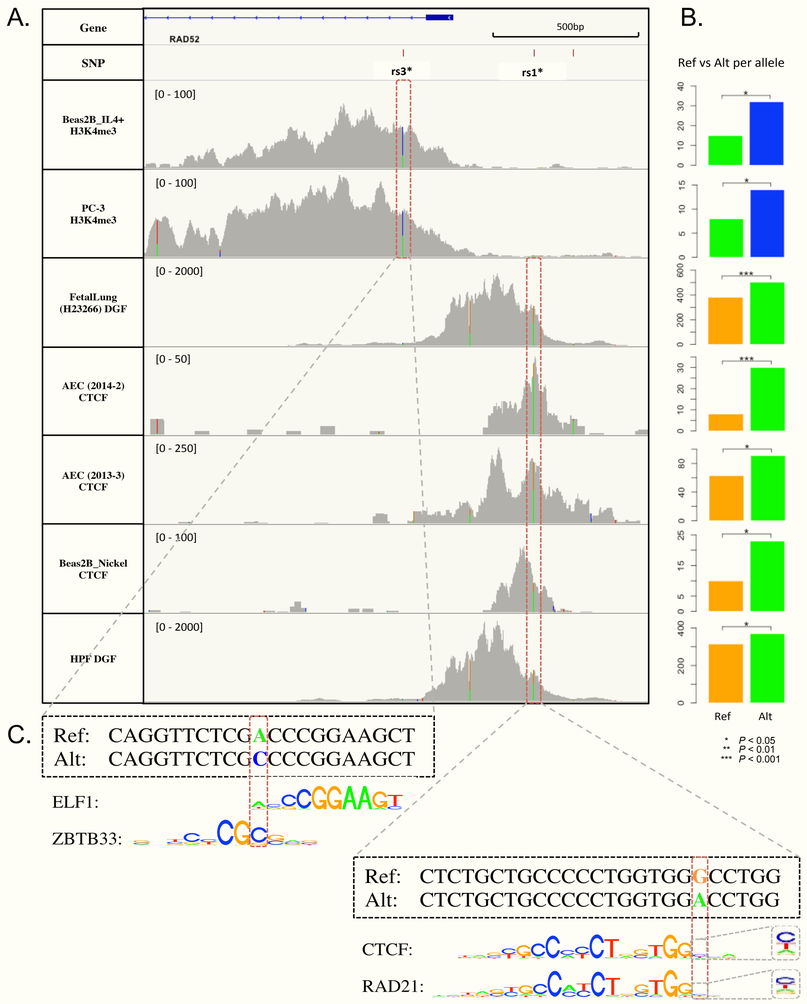
Number of SNPs at each step



Chosen SNPs that showed ASE in lung relevant cells

rsID	Region	Index SNPs	Genes	Potential function	r2	biofeature	eQTL	TF binding prediction	ASE in other cells types (ENCODE DGF)	Fine mapping
rs7*	5p15.33	rs2736100	TERT	Enhancer	0.81	Roadmap_FetalLung_DGF		GR	N	1st most sig SNP after imputation in 5p15.33[1]
rs7*	5p15.33	rs2736100	TERT	Enhancer	0.81	Roadmap_FetalLung_DGF Roadmap_FetalLung_DNase		MEIS1	N	2nd most sig SNP after imputation in 5p15.33[1]
rs3*	6p21.33	rs3117582	BAT3, APOM	Promoter	1	Nhlf_DGF	C4A, NOTCH4 …	E2F, NRF1	Y	7th in LUSQ[2]
rs3*	6p21.33	rs3117582	CLIC1	Enhancer	0.98	Nhlf_DGF, Nhlf_Dnase	C4A, NOTCH4 …	ATF3, GR, SREBP	Y	NA
rs3*	12p13.33	rs6489769 rs10849605	RAD52	Promoter	0.57	Beas2B_IL4+_H3K4me3, PC-3_H3K4me3	RAD52	Ehf, ELF5	Y	1st most sig SNP after imputation in 12p13.33[2,3]
rs1*	12p13.33	rs6489769 rs10849605	RAD52	Promoter	0.54	Roadmap_FetalLung_H23266_DGF, AEC_2014_4_CTCF_D0, AEC_2013_3_CTCF_D0, Beas2B_CTCF_ChIPSeq_Nickel, ENCODE_HPF_DGF	RAD52	CTCF	Y	6th most sig SNP after imputation in 12p13.33[2]
rs4*	22q12.22	rs36600	MTMR3	Enhancer	0.64	Basal Cell ATAC-long Basal Cell ATAC-short	MTMR3, ASCC2 …	SOX	Y	1st most sig SNP after imputation in 22q12.2[4]
rs1*	22q12.22	rs17728461	HORMAD2	Enhancer	1	IMR90_DGF	UQCR10	Gm397	N	1st most sig SNP after imputation in 22q12.2[4]

Allele-specific effects in RAD52 locus



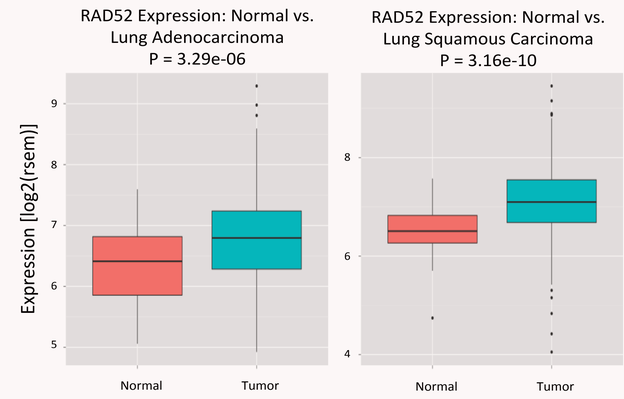
Analysis of SNPs showing allele-specific effects (ASE) in the RAD52 locus. Two SNPs (rs3* and rs1*) that are associated with lung cancer risk showed allele-specific effects in numerous sequencing experiments (A). Allelic distribution showed a preference for the alternative allele (B). Transcription factor binding sites were predicted around the two SNPs (C). Expression Quantitative Trait Loci (eQTL) showed that the alternative alleles are associated with higher expression of RAD52 (D and E). Luciferase assays showed that vectors containing the promoter region of RAD52 with alternative alleles of the two SNPs have higher activity than those with reference alleles.

Discussion

RAD52 Expression in TCGA lung cancer

RAD52

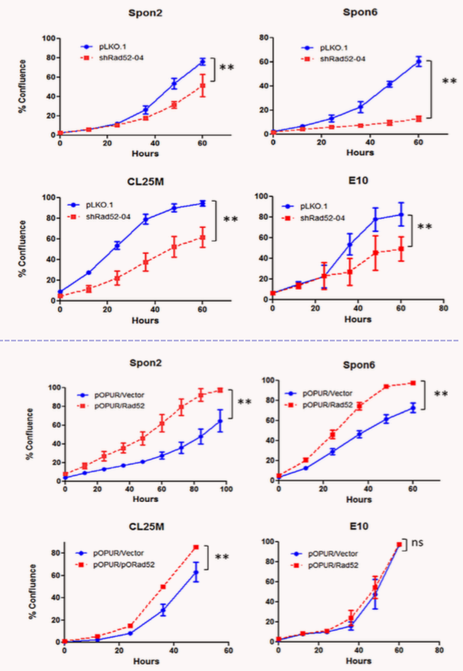
- The protein encoded by this gene shares similarity with Saccharomyces cerevisiae Rad52, a protein important for DNA double-strand break repair and homologous recombination. This gene product was shown to bind single-stranded DNA ends, and mediate the DNA-DNA interaction necessary for the annealing of complementary DNA strands. It was also found to interact with DNA recombination protein RAD51, which suggested its role in RAD51-related DNA recombination and repair. [RefSeq]



RAD52 and cell proliferation

RAD52 knockdown cells showed lower proliferation rates

RAD52 overexpressing cells showed higher proliferation rates



Lieberman, et al. 2015

References

1. Wang, Z. *et al.* Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum Mol Genet* **23**, 6616-6633, doi:10.1093/hmg/ddu363 (2014).
2. Wang, Y. *et al.* Deciphering associations for lung cancer risk through imputation and analysis of 12,316 cases and 16,831 controls. *Eur J Hum Genet* **23**, 1723-1728, doi:10.1038/ejhg.2015.48 (2015).
3. Timofeeva, M. N. *et al.* Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet* **21**, 4980-4995, doi:10.1093/hmg/dds334 (2012).
4. Hu, Z. *et al.* A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat Genet* **43**, 792-796, doi:10.1038/ng.875 (2011).
5. Genotype-Tissue Expression Portal (**GTEx**): http://www.gtexportal.org/home/
6. Pruitt, K. D. *et al.* RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* **42**, D756-763, doi:10.1093/nar/gkt1114 (2014).
7. The Cancer Genome Atlas (**TCGA**) Data Portal: https://tcga-data.nci.nih.gov/tcga/.
8. Lieberman, R. et al. Functional characterization of RAD52 as a lung cancer susceptibility gene in the 12p13.33 locus. Mol Carcinog 55, 953-963, doi:10.1002/mc.22334 (2016).