Identification of Copy Number Variants in Hereditary Lung Cancer Families

John Waldron[1], Kirsten W. Termine[1], Anthony M. Musolf[2], Mariza de Andrade[3], Colette Gaba[4], Ramaswamy Govindan[5], Ping Yang[3], Ming You[6], Marshall W. Anderson[7], Ann G. Schwartz[8], Susan M. Pinney[7], Christopher I. Amos[9], Joan E. Bailey-Wilson[2], Diptasri Mandal[1]

[1]Department of Genetics, LSU Health Sciences Center, New Orleans, LA; [2]National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; [3]Mayo Clinic, Rochester, MN; [4]University of Toledo Dana Cancer Center, Toledo, OH; [5]Division of Oncology, Washington University School of Medicine, St. Louis, MO; [6]Medical College of Wisconsin, Milwaukee, WI; [7]University of Cincinnati College of Medicine, Cincinnati, OH; [8]Karmanos Cancer Institute, Wayne State University, Detroit, MI; [9]Baylor College of Medicine, Houston, TX

INTRODUCTION

- Lung cancer (LC) is the most common cause of cancer mortality and the third most common cancer by incidence in the United States.
- While environmental factors (e.g. tobacco smoke) play an important role in its development, lung cancer risk also exhibits a high degree of heritability.
- Structural mutations, including a gain or loss of DNA (copy number variants or CNVs) are important sources of phenotypic variation and likely contribute a larger fraction of genomic variation among individuals than single nucleotide polymorphisms.
- The goal of the current project is to utilize the whole exome sequencing (WES) data in identifying the CNVs in the hereditary lung cancer families recruited by the Genetic Epidemiology of Lung Cancer Consortium (GELCC).

METHODS

1. Whole Exome Sequences (WES) were obtained from germline samples of 204 individuals (60 [29.4%] with LC diagnosis) from 25 high-risk lung cancer families (≥3 affected lung cancer cases/family).
2. Two independent CNV identification tools using read depth to infer copy number changes and genomic breakpoints were used: GATK4.0 and XHMM.
3. CNVs presented here were limited to deletions. Those shared with non-blood relatives and with an allele frequency > 0.05 in gnomAD SV were discarded. Remaining CNVs were considered reliable according to consistency with Mendelian expectations if present in at least 3 families and at least 2 individuals per family.

RESULTS

- **The vast majority of CNVs were shared by fewer than 5 individuals. 103 exonic loci affected by deletions of mean width 24kb remained following a stringent screening approach. Some of these loci have been associated with cancer in previous studies (Table 1). Larger deletions are associated with CNVs that are common in the general population.**
- **Biological function of observed variants is enriched for cell adhesion, and cytoplasm, nucleoplasm, and chromosome structure.**
- **Deletions in SSX1 were identified in only one family, but six individuals (2 with LC diagnosis) with X linked dominant inheritance.**

CNV calls varied across algorithms and filter steps

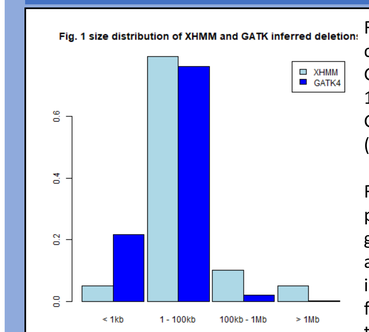


Fig 1. Size distribution of calls in XHMM and GATK4. XHMM mean = 133.91kb (SD = 397.08), GATK4 mean = 17.69kb (SD = 68.59).

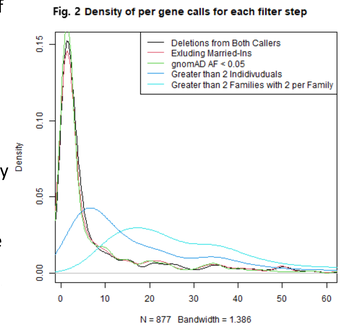
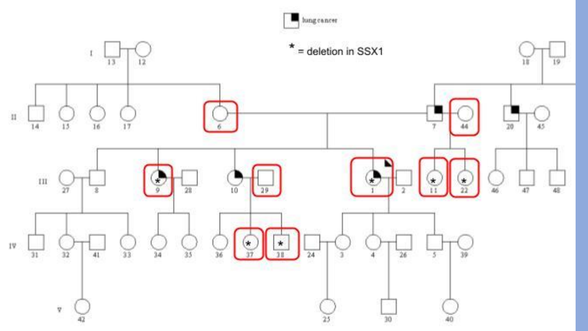


Fig. 2. Changes in density plot of protein coding genes with deletions across filter steps. There is a high density of low frequency variants prior to filtering.

One high risk family exhibits X linked dominant LC risk pattern potentially explained by a 16kb deletion

Fig 3. pedigree illustrating family members with lung cancer diagnosis (shaded) and whole exome sequences (red boxes). An asterisk marks individuals with a 16kb deletion in SSX1 which is associated with synovial sarcoma.



Conclusions

A systematic approach to identifying causal structural variants reveals non-intuitive and potentially novel loci underlying heritability of lung cancer risk. A priori analysis as conducted reveals few apparent functional associations with oncogenic mechanisms. The vast majority of identified structural variants occurred in one or a few individuals. To our knowledge, there have been no studies addressing germline structural variation and lung cancer

Current analysis is limited to deletions, because predicting the functional significance of duplication is more complex.

Family-based association testing (FBAT) for each deletion, which is more appropriate than odds ratio for family data, is the next step in analysis. Overlapping deletions within genes of interest will be subjected to linkage analysis. We will consider overall CNV burden and disease association and further functional analysis of identified deletions and duplications. We also plan to extend this work to sporadic lung cancer and assess risk in individuals with different ancestry.

Acknowledgements: We express our gratitude to the study participants, the collaborating hospitals, and the physicians. This work is funded by NIH: HHSN268201200007C, N01-HG-65404, 3U01CA076293-10S1, and 3U19CA203654-05S1. Support was also provided by the BIG Program at LSU School of Medicine, Tulane Cancer Center (NIH: P01CA214091), and LSUHSC-NO (NSF:1657895).

Table 1. CNV Analyses of Protein coding genes found in three or more families (out of 25 high-risk families) and at least two individuals per family with frequency < 0.05 in general population using XHMM and GATK4.0

Gene Name	Number of Families [affected, unaffected]	LC Dx Odds Ratio [95% CI]	Mean CNV width	Previously reported in cancer?	Gene Function
KIR2DL1	16 [11, 16]	1.77 [0.77-4.08]	18.7kb	Yes [LC]	Receptor on natural killer (NK) cells for some HLA-C alleles such as w4 and w6. Inhibits the activity of NK cells thus preventing cell lysis.
ADAM3A	17 [9, 27]	0.75 [0.33-1.71]	38.5kb	Yes [Glioma, Leukemia]	ADAM3A (ADAM Metallopeptidase Domain 3A (Pseudogene)) is a Pseudogene. Diseases associated with ADAM3A include Conjunctival Squamous Cell Carcinoma.
DMBT1	14 [15, 25]	1.56 [0.75-3.23]	13.6kb	Yes [LC]	Loss of sequences from human chromosome 10q has been associated with the progression of human cancers. This gene was originally isolated based on its deletion in a medulloblastoma cell line. May be considered as a candidate tumor suppressor gene for brain, lung, esophageal, gastric, and colorectal cancers. May play roles in mucosal defense system, cellular immune defense and epithelial differentiation.
GOLGA8B	15 [9, 27]	0.75 [0.33-1.71]	43.4kb	Yes [Hepatocellular carcinoma]	Involved in maintaining Golgi structure. Diseases associated with GOLGA8B include Smith-Mccort Dysplasia 1 and Myopia. GOLGA8B may act as a tumor suppressor in hepatocellular carcinoma.
PTPN20	17 [15, 41]	0.82 [0.42-1.63]	13.7kb	Yes [Lymphoma & other cancers]	Tyrosine-protein phosphatase targeted to sites of actin polymerization in response of varied extracellular stimuli. Has tyrosine phosphatase activity towards various tyrosyl phosphorylated substrates. PTPN gene family members is involved in cell proliferation, survival, immune response, migration, and metabolism.
HNRNPCL1	9 [3, 16]	0.41 [0.12-1.48]	54.1kb	Yes [LC & other cancers]	May play a role in nucleosome assembly by neutralizing basic proteins such as A and B core hnRNPs. Associated diseases include Mayer-Rokitansky-Kuster-Hauser Syndrome. HNRNPCL1 is co-expressed with ACSL3 (long-chain fatty acyl-Co A synthetase), expression of which has been linked with melanoma, brain, lung, breast, ovarian cancer.
SIRPB1	22 [22, 39]	1.53 [0.81-2.90]	21.7kb	Yes [LC]	A receptor-type transmembrane glycoproteins known to be involved in the negative regulation of receptor tyrosine kinase-coupled signaling processes. Associated diseases include Polycystic Lipomembranous Osteodysplasia With Sclerosing Leukoencephalopathy 1 and Immunodeficiency 50. Differential methylation was observed in promoter of SIRPB1 in lung cancer.
LILRA6	24 [27, 62]	1.22 [1.09-2.69]	3.2kb	Yes [LC]	Leukocyte Immunoglobulin Like Receptor A6 is associated with Shwartzman Phenomenon. Among its related pathways are Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell and Class I MHC mediated antigen processing and presentation. Expression of LILRB2 is adversely related to lung cancer prognosis.
HLA-DRB5	25 [35, 70]	1.64 [1.01-3.09]	5.1kb	Yes [Cancer]	HLA class II beta chain paralogues. This class II molecule is a heterodimer consisting of an alpha (DRA) and a beta (DRB) chain, both anchored in the membrane. It plays a central role in the immune system by presenting peptides derived from extracellular proteins. There are multiple pseudogenes of this gene.