

UC Office of the President

Research Grants Program Office (RGPO) Funded Publications

Title

Positional integration of lung adenocarcinoma susceptibility loci with primary human alveolar epithelial cell epigenomes

Permalink

<https://escholarship.org/uc/item/2fh6f6sf>

Journal

Epigenomics, 10(09)

ISSN

1750-1911

Authors

Yang, Chenchen

Stueve, Theresa Ryan

Yan, Chunli

et al.

Publication Date

2018-09-01

DOI

10.2217/epi-2018-0003

Peer reviewed

1 **Positional Integration of Lung Adenocarcinoma Susceptibility Loci with Primary Human Alveolar**
2 **Epithelial Cell Epigenomes**

3
4 **Chenchen Yang^{1,2,3*}, Theresa Ryan Stueve^{1,2,3,4*}, Chunli Yan^{1,2,3}, Suhm Kyong Rhie^{1,2,3}, Daniel J.**
5 **Mullen^{1,2,3}, Jiao Luo^{2,5}, Beiyun Zhou^{3,5,6}, Zea Borok^{2,3,5,6}, Crystal N. Marconett^{1,2,3}, and Ite A. Offringa^{1,2,3#}**

6 **1** Department of Surgery, **2** Department of Biochemistry & Molecular Medicine, **3** Norris Comprehensive
7 Cancer Center, **4** Department of Preventive Medicine, **5** Division of Pulmonary and Critical Care and Sleep
8 Medicine, Department of Medicine, **6** Hastings Center for Pulmonary Research, Keck School of Medicine,
9 University of Southern California 90089, USA

0 * These authors contributed equally to this work

1 # Corresponding author: ilaird@usc.edu

2
3 Short title: Epigenetic environment of lung adenocarcinoma risk SNPs

4
5 **Data deposition:** Data is provided in GEO record GSE84273

6
7 **Acknowledgements**

8 H1648 and H522 cells were the kind gifts of the Dr. Eric Haura Laboratory. This work was supported by
9 National Institute of Health (<https://www.nih.gov/>) grants R01 HL114094 (to IALO, ZB), R01 HL126877 and
0 R01 HL112638 (to ZB), the Norris Comprehensive Cancer Center core grant (National Cancer Institute
1 P30CA01408, supporting IALO), Department of Defense (<http://cdmrp.army.mil/funding/lcrp>) Concept Award
2 W81XWH-14-1-0174-1 (to IALO), the Tobacco-Related Disease Research Program (award ID 500806, to
3 IALO) and the California Community Foundation (<http://www.calfund.org/>) (to IALO, supporting CY and JL) and
4 support from the Thomas G. Labrecque Foundation (<https://tgfoundation.com/>) (to IALO), the Whittier
5 Foundation (to IALO), the Hastings Foundation (to ZB) and generous donations from Conya and Wallace
6 Pembroke (to IALO). ZB is the Ralph Edgington Chair in Medicine and Hastings Professor of Medicine. TRS
7 was supported by the National Institute of Health National Institute of Environmental Health Sciences (NIEHS,

8 NIH T32ES013678) and by the USC Provost's Postdoctoral Scholar Research Grant. DJM was supported by a
9 University of Southern California Provost Fellowship and a Roy E. Thomas Foundation graduate scholarship.
0 CNM was supported by American Cancer Society (<https://www.cancer.org/research.html>) /Canary Foundation
1 (<http://www.canaryfoundation.org/>) postdoctoral fellowship # PFTED-10-207-01-SIED and later the Department
2 of Surgery, Keck School of Medicine. The funders had no role in study design, data collection and analysis,
3 decision to publish, or preparation of the manuscript.
4

5 **Author Contributions**

6 Conceptualization: CY (Chenchen Yang), TRS, CNM, IALO. Data curation: CY, CNM. Formal analysis: CY,
7 TRS, DJM. Funding acquisition: CNM, TRS, BZ, ZB, IALO. Investigation: CY, TRS, CLY (Chunli Yang).
8 Methodology: CY, TRS, CNM, SKR, DJM. Project administration: IALO. Resources: ZB, JL, BZ, IALO.
9 Software: CY, SKR. Supervision: CNM, IALO. Validation: CY, TRS, CLY. Visualization: CY, TRS, DJM, IALO.
0 Writing original draft: CY, TRS, IALO. Writing review & editing: All authors.
1

2 **Abstract**

3 **Aim:** To identify functional lung adenocarcinoma (LUAD) risk SNPs. **Materials & methods:** Eighteen validated
4 LUAD risk SNPs ($p \leq 5 \times 10^{-8}$) and 930 SNPs in high linkage disequilibrium ($r^2 > 0.5$) were integrated with
5 epigenomic information from primary human alveolar epithelial cells. Enhancer-associated SNPs likely
6 affecting transcription factor (TF) binding sites were predicted. Three SNPs were functionally investigated
7 using luciferase assays, expression quantitative trait loci (eQTLs) and cancer-specific expression. **Results:**
8 Forty-seven SNPs mapped to putative enhancers; eleven located to open chromatin. Of these, seven altered
9 predicted TF binding motifs. Rs6942067 showed allele-specific luciferase expression and eQTL analysis
0 indicate it influences expression of *DCBLD1*, a gene that encodes an unknown membrane protein and is
1 overexpressed in LUAD. **Conclusions:** Integration of candidate LUAD risk SNPs with epigenomic marks from
2 normal alveolar epithelium identified numerous candidate functional LUAD risk SNPs including rs6942067,
3 which appears to affect *DCBLD1* expression.
4
5
6

7
8
9
0
1
2
3
4
5
6
7
8
9
0
1
2
3
4
5
6
7
8
9
0
1
2
3

Data deposition: Data are provided in GEO record GSE84273

Key words: SNP, lung adenocarcinoma, epigenomics, alveolar epithelial cells, enhancer, FAIRE, eQTL, rs6942067, *DCBLD1*

Introduction

Lung cancer is the leading cause of cancer death worldwide, and in the United States it is responsible for more deaths than breast, colorectal, and prostate cancer combined [1]. The most common histological subtype of lung cancer is lung adenocarcinoma (LUAD), accounting for over forty percent of new lung cancer cases. Never-smokers with lung cancer are more likely to have LUAD and to be women and persons of Asian descent [2, 3]. Though smoking is strongly implicated in lung cancer risk, it is estimated that half of all new cases arise in never smokers or smokers who quit many years ago [4]. Epidemiological studies indicate that in addition to environmental factors, LUAD susceptibility also has a significant genetic component [5-7].

To date, genome-wide association studies (GWAS) have uncovered 18 validated single nucleotide polymorphisms (SNPs) associated with LUAD risk ($p \leq 5 \times 10^{-8}$), distributed over 12 chromosomal regions (Table 1). Causal links between these SNPs and lung health remain largely unknown. The principal challenges to understanding how SNPs confer susceptibility to LUAD are the same for all post-GWAS studies [8-11]. First, because GWAS SNPs are landmark or 'index' SNPs that serve as positional references for phenotypic associations, hundreds of SNPs in linkage disequilibrium (LD), could be the actual functional SNP. Secondly, the majority of all GWAS SNPs, including 17 of the 18 index LUAD SNPs (Table 1), reside outside of protein-coding regions of genes and are thus hypothesized to influence gene expression by disrupting regulatory elements that could be hundreds of kilobases away from their target genes. Consistent with this hypothesis, several groups have determined that the majority of causal or 'functional' noncoding SNPs are in fact concentrated in epigenomic features characteristic of enhancers of gene transcription [8-16]. Epigenomic features make up a layer of information that is superimposed onto the genome; this does not alter the genetic sequence but affects cell type-specific gene expression [17]. Epigenomic features include chemical modifications such as DNA methylation, regulatory RNAs, transcription factors, modified histones and genome-

organizing factors [17]. Epigenomic features characteristics of enhancers include regions of open chromatin (such as DNase I hypersensitive sites) which are accessible to DNA-binding proteins, histone 3 monomethylation at lysine 4 (H3K4me1, a mark for poised or active enhancers [18]) and histone 3 acetylation at lysine 27 (H3K27ac, a mark for active enhancers [19]). Identification of epigenetic marks disrupted by SNPs has allowed exciting progress to be made in post-GWAS investigations of genetic susceptibility to autoimmune disorders, Alzheimer's disease, diabetes, and cancers of the breast, prostate, and colon [10, 12-16, 20-22]. Importantly, in order to be relevant, the epigenomic features used for positional integration with candidate functional SNPs must be obtained from cell types pertinent to the disease at hand. In the case of LUAD, alveolar epithelial cells are highly relevant because LUAD arises in the peripheral lung.

Alveolar lung epithelium consists of two epithelial cell types: squamous type 1 (AT1) cells and cuboidal type 2 (AT2) cells. AT1 cells are considered terminally differentiated and mediate gas exchange. They are large delicate cells that form the interface between the air in the alveoli and the microcapillaries that surround the alveolar sacs and are the site for gas exchange [31]. AT1 cells make up 95% of the surface area of the alveoli, and could, therefore, be disproportionately impacted by inhaled toxicants such as those present in tobacco smoke or pollution. AT2 cells produce the surfactants required to prevent the collapse of alveoli upon expiration. In adult lung, they are the progenitors of AT1 cells in response to lung injury; AT2 cells can both self-renew and differentiate into AT1 cells, and this process can be replicated *in vitro* [32-35]. Given their proliferative capacity, AT2 cells are commonly implicated as the likely progenitors of LUAD [36-39]. However, because definitive data on the roles and interactions of AT1 and AT2 cells in LUAD initiation remain to be provided, we rationalized that both cell types (here collectively referred to as alveolar epithelial cells or AEC) should be investigated in LUAD risk studies. Here we use epigenomic data from purified primary human AT2 cells, *in vitro*-differentiated AT1-like cells, and LUAD cell lines for positional integration with LUAD risk SNPs, and apply a combination of bioinformatics and molecular approaches to identify candidate *functional* risk SNPs (outlined in flowchart Supplementary Fig 1).

Results & Discussion

Classification of SNPs associated with LUAD risk

We first carried out a literature review, collecting all SNPs reported to be significantly associated with LUAD risk. We filtered for genome-wide significance ($P < 5 \times 10^{-8}$) and for those SNPs that had been validated,

2 i.e. observed to be significant in more than one data set. We thus collected 18 validated GWAS index SNPs
3 that had been significantly associated with LUAD risk (Table 1, Figure 1A). In agreement with the general
4 observation that most GWAS SNPs are located in noncoding regions [9], 17 were found to be either intronic or
5 intergenic, while the one, rs1051730, falls within an exon resulting in a synonymous mutation. We next
6 obtained 930 SNPs in high linkage disequilibrium (LD; $r^2 \geq 0.5$) with the 18 index-SNPs using 1000 Genomes
7 LD data [40] matched for the ethnicity of the population in which the index risk SNP was identified
8 (Supplementary Table 1), and grouped these into functional classes for subsequent analysis (Supplementary
9 Table 2). Of the 948 total index and LD SNPs, 583 were intronic, while 335 were intergenic. Only seven SNPs
0 mapped to coding regions, including three missense mutations and four synonymous mutations. Seven SNPs
1 were located in the 5' or 3' untranslated regions (UTRs), where microRNAs are most often targeted, and 16 in
2 putative promoter regions (within 1000 bp to either side of the transcription start site or TSS).

3 Very little published data on the potential functionality of the 948 SNPs is available. Several SNPs in
4 the 15q25.1 locus have been implicated as potential functional SNPs affecting expression of the cholinergic
5 receptor, neuronal nicotinic alpha polypeptide-5 (*CHRNA5*), a protein involved in the nicotine response [41,
6 42]. Rs16969968 resides in the fifth exon of *CHRNA5* and changes an aspartic acid codon (GAT) into an
7 asparagine codon (AAT) at amino acid 398 in the second intracellular loop. This makes the receptor less
8 responsive to nicotine, thereby increasing nicotine dependence, heavy smoking, and potentially, lung cancer
9 risk [43]. Rs55853698 and rs55781567, located in the 5'UTR of *CHRNA5*, have been reported to alter
0 *CHRNA5* promoter activity [44].

1 To preliminarily evaluate whether the remaining two missense-causing SNPs (rs2076530 in *BTNL2* and
2 rs9891146 in *C17orf58*) might alter protein function, we performed structure-activity predictions using
3 PolyPhen2 and SIFT software. Neither of the missense mutations were predicted to be deleterious by either
4 program (Supplementary Table 3), decreasing the likelihood that these SNPs might alter the function or
5 structure of the proteins. To investigate whether the 7 SNPs in UTRs might affect miRNA targeting, we
6 performed sequence-activity binding predictions using the miRNA target database TargetScanHuman. No
7 allele for any UTR SNP was highlighted as a known microRNA target (Supplementary Table 4). As noted
8 above, of the SNPs located in promoter regions, only the two *CHRNA5* SNPs have been implicated in affecting
9 promoter function (Supplementary Table 5).

0 In sum, except for several SNPs in the 15q25.1 locus, all loci appear to lack a biological mechanism
1 explaining their association with increased LUAD risk. Because enhancers have been commonly implicated in
2 risk SNP function [10, 12-16, 20-22], we explored whether one or more of the SNPs might influence risk by
3 affecting AEC enhancers. To be thorough, we included all 948 SNPs in the investigation.

4 5 **Identification of AEC enhancers**

6 As a first step in generating AEC enhancer profiles, we purified AT2 cells from remnant human
7 transplant lung. We used cells from a non-smoker to ensure optimally healthy cells, thereby limiting disease
8 confounders. AT1 cells are too delicate to allow purification in sufficient numbers, necessitating derivation of
9 AT1-like cells through *in vitro* differentiation over the course of 6 days as described [45] (Supplementary Figure
0 2A). At day 2 in culture, AT2 cells undergo dramatic changes in gene expression and epigenomic marks, and
1 by days 4 and 6, they have transdifferentiated into cells that exhibit characteristics of native AT1 cells. We
2 confirmed proper differentiation of the cells by Western blot analysis (Supplementary Figure 2B). We
3 performed chromatin immunoprecipitation and subsequent DNA sequencing (ChIP-seq) using material from
4 day 0 (D0; AT2 cells) and days 4 and 6 (D4, D6; AT1-like cells) using antibodies against enhancer marks
5 H3K27ac and H3K4me1. In D0, D4, and D6 cells respectively, we identified 39,210, 44,976, and 43,743
6 H3K27ac peaks and 87,443, 76,292, and 75,490 H3K4me1 peaks (Supplementary Table 6; saturation plots
7 are shown in Supplementary Figure 3A,B). Peak analysis revealed substantial overlap of peaks from different
8 days (Supplementary Figure 4A,B).

9 In addition to ChIP-seq for enhancer marks, we also performed Formaldehyde-Assisted Isolation of
0 Regulatory Elements-sequencing (FAIRE-seq [46]) on AECs from days 0, 4 and 6 in culture to identify
1 nucleosome-depleted regions of the genome that are accessible to DNA-binding proteins (i.e. open chromatin,
2 Supplementary Figure 3C). The D0 peaks were largely distinct from the D4/D6 peaks, while the latter showed
3 substantial overlap (Supplementary Fig 4C). Plots of the tag density versus distance to the center of the
4 FAIRE-seq peaks showed that FAIRE-seq signals were enriched in the center of FAIRE-seq peaks
5 (nucleosome depletion), while as expected, signals for the histone modifications H3K27ac and H3K4me1 (on
6 nucleosomes) were enriched in regions flanking FAIRE-seq peaks (Supplementary Figure 5).

8 Identification of SNPs present in enhancers

9 We next investigated which among the 948 candidate SNPs were located in AEC enhancer regions (i.e.
0 were located in H3K27ac and H3K4me1 ChIP-seq peaks) in D0, D4, and/or D6 cells. We filtered out SNPs
1 located on the extreme flanks of peaks by requiring that SNPs be significantly enriched in enhancer marks on
2 both of the 0.5 kb flanking sides (Supplementary Figure 6). This yielded 47 AEC enhancer-associated SNPs
3 that we classified as AT2 cell-specific (present only in D0 enhancers), AT1 cell-specific (present only in D4/6
4 enhancers) and general AEC enhancer SNPs (present in both AT1 and AT2 enhancers). Seven AT2-specific
5 SNPs, 23 AT1-specific SNPs, and 17 AEC-specific SNPs were identified respectively (Table 2 and Figure 1B).
6 Notably, none of the index-SNPs themselves were located in AEC, AT1 or AT2 cell enhancer elements. For 10
7 of the 18 index SNPs, including the two near the *CHRNA5* region, we detected no AEC enhancer-associated
8 LD SNPs. This could be because these SNPs do not function through enhancers (e.g. the previously identified
9 missense and promoter *CHRNA5* SNPs), because they affect LUAD risk through effects on other cell types
0 (such as immune cells, lung fibroblasts, etc.) or because their function is influenced by other factors such as
1 smoking, which are not investigated here.

2 The 47 candidate AEC enhancer SNPs correspond to 8 of the 18 index GWAS LUAD risk SNPs (Table
3 2 and Figure 1B). The number of candidate SNPs per region varied from one or two to 22 SNPs on 6q22.1, in
4 LD with index SNP rs9387478. The latter SNPs are located in the *DCBLD1* and *ROS1* gene region, which is
5 also an intron of a putative fusion transcript between *GOPC* and *ROS1* and contains numerous active
6 chromatin marks. Two SNPs in high LD with rs4488809 on chromosome 3q28 were found in general AEC
7 enhancer marks located in the promoter region (-1000bp - +100bp from TSS) of TP63 (Supplementary Table
8 5), suggesting an enhancer close to a transcription start site. Such an enhancer may or may not affect the
9 nearest gene [47]. In studies of risk enhancers, SNPs located in nucleosome-free, transcription factor (TF)-
0 accessible regions of the regulatory element are the most promising candidates for functional follow-up [9]. We
1 thus filtered the 47 candidate SNPs by whether they were located in FAIRE peaks, which resulted in 11 top
2 candidates (Supplementary Figure 1B). To further evaluate these 11 SNPs, we assessed the likelihood that
3 they would disrupt or create TF binding sites.
4

5 Identification of SNPs that affect transcription factor binding sites

6 We compiled motif positional weight matrixes (PWM) from ENCODE-motif [48], Factorbook [49] and
7 HOCOMOCO [50] (Supplementary Figure 7). We removed TFs that are not expressed in AT1 and/or AT2 cells
8 by excluding sites for TFs with low reads per kilobase per million (rpkm; set at $\text{rpkm} \leq 3$, from RNA-seq data
9 [51]); binding sites for TFs that are not expressed in alveolar epithelium would not be expected to have any
0 functional effects in these cells. Our analysis identified predicted transcription factor binding sites overlapping
1 with 7 of the 11 candidate risk enhancer SNPs (Figure 1C, Table 3). Next, we searched publicly available
2 ChIP-seq data from ENCODE [52] and GEO [53] for experimental evidence supporting binding of predicted
3 TFs to the SNP regions. This approach is limited by the fact that publicly available ChIP-seq datasets such as
4 those generated by the ENCODE consortium are restricted to the factors that have been examined to date. In
5 this latter analysis, we included all ENCODE cell types, including but not limited to lung-relevant cell types such
6 as A549 (a LUAD cell line), small airway epithelial cells (SAEC, normal primary cells from the more distal
7 airways), IMR90 (a cell line derived from embryonic lung fibroblast), normal human lung fibroblasts (NHLF) and
8 human pulmonary fibroblasts (HPF, obtained from a different source than NHLF). All of these cell types are
9 distinct from AEC, and it is well recognized that enhancers vary considerably among cell types. However, if a
0 predicted TF was reported to bind its target by ChIP-seq in any cell type, we took this as confirmation that the
1 factor can actually bind to the predicted target in the context of a cellular environment. For 4 of the 7 SNPs, we
2 found ChIP-seq data supporting binding of the TFs we had predicted. In addition, ChIP-seq data-based TF
3 binding was noted for two SNP locations for which we had predicted no TF binding sites (rs2057314 and
4 rs12602556, Table 3). This could happen when TFs bind in multi-protein complexes or bind indirectly to DNA.
5 In the case of indirect binding, we assumed that an effect of the SNP would be less likely. We next focused on
6 3 SNPs for which predicted TF binding sites were supported by ChIP-based evidence (Figure 1C). These were:
7 rs452384, corresponding to index SNP rs465498; and rs6942067 and rs9481728, corresponding to index SNP
8 rs9387478.

9 Rs452384 lies in a putative general AEC enhancer in a central intron of the cleft lip and palate
0 transmembrane protein 1-like (*CLPTM1L*) gene (Figure 2A). *CLPTM1L* was recently implicated in lung
1 tumorigenesis [54] and encodes a membrane protein that leads to apoptosis when overexpressed in cisplatin-
2 sensitive cells [55]. This SNP also resides in a strong DNase hypersensitive site in small airway epithelial cells

3 (SAECs, ENCODE data [48, 49]). Polymorphisms in *CLPTM1L* and the neighboring telomerase (*TERT*) gene
4 have been reported to increase susceptibility to numerous cancers, including lung, pancreatic, and breast
5 cancers [28, 56-59]. The reference (risk) allele of rs452384 forms a predicted site for NKX2-1 that is disrupted
6 by the alternate allele (Table 3). Importantly, the ability of this TF to bind to this site is supported by ChIP-seq
7 data in lung adenocarcinoma cell line H3122 (homozygous for the risk allele)[51] (Figure 2B,C). NKX2-1 plays
8 a key role in driving lung epithelial tissue differentiation from endoderm [60, 61] and is one of the most
9 significantly amplified genes in LUAD [53]. Interestingly, a binding site for the *MYC* proto-oncogene is predicted
0 for the alternate rs452384 allele (Figure 2C). *MYC* is implicated in numerous cancers including LUAD [62-64]
1 and has been ChIPed at rs452384 by the ENCODE Consortium in the MCF10A-Er-Src cell line (homozygous
2 for the risk allele) [52].

3 Rs6942067 is located in the intergenic region between the *DCBLD1* (Discoidin, CUB and LCCL
4 Domain-containing 1) and *ROS1* (ROS proto-oncogene 1, a receptor tyrosine kinase) genes (Figure 3).
5 Rs6942067 is positioned in the center of a DNase hypersensitive region between two nucleosomes carrying
6 H3K27ac and H3K4me1 marks and the reference (risk) A allele lies in a predicted binding site for E1A-binding
7 protein P300 (EP300) that is disrupted by the alternate allele (Table 3). EP300 binding to this site has been
8 observed by ChIP-seq in retinoic acid-treated neuroblastoma cell line SK-N-SH_RA (A/A genotype) (ENCODE
9 data [52], Figure 3B,C). EP300 is a histone acetyltransferase that activates transcription via chromatin
0 remodeling, and it has been implicated in a variety of cancers [65]. It is mutated and/or overexpressed in a low
1 percentage of lung adenocarcinomas [66]. The alternate allele of rs6942067 forms a TF binding site for the
2 thyroid hormone receptor alpha (THRA, Figure 3C), which is one of the several receptors for thyroid hormone
3 known to be involved in lung development and alveolar cell function [67-69]. Notably, low levels of thyrotropin,
4 a hormone that regulates thyroid hormone release, have been linked to an increased risk for prostate and lung
5 cancer [70]. Like rs6942067, rs9481728 (located in the first intron of *DCBLD1*) carries enhancer histone marks
6 and an EP300 site on the reference (risk) allele that is disrupted by the alternate allele (Figure 4, Table 3).

7 The epigenomic environment of the three SNPs, coupled with the presence of overlapping TF binding
8 sites predicted to be affected by the SNP alleles, provided a strong rationale for functional analyses. This was
9 further supported by the observation of active chromatin marks at the location of these three SNPs in LUAD
0 cell lines [71] (Supplementary Figure 8-10). The first functional assay we carried out for the three candidate

1 functional SNPs was a luciferase assay, in which we cloned the PCR-amplified putative enhancer region and
2 inserted it upstream of a minimal promoter in the luciferase gene reporter vector pGL4.26. The assay allows
3 both alleles to be tested in the same genetic and cellular environment. We then tested whether the genomic
4 segment containing either allele of the SNP would exhibit enhancer activity, and whether the alleles differed in
5 the extent to which they could enhance expression.. Because immortalized human AEC are not available, we
6 transfected each reporter construct into a LUAD cell line showing strong genomic H3K4me1 and H3K27ac
7 peaks in the SNP region (based on publicly available H3K4me1 and H3K27ac ChIP-seq datasets for 26 lung
8 cancer cell lines in DBTSS, <http://dbtss.hgc.jp> [71]), reasoning that these cell lines would be enriched for TFs
9 and coactivators required for enhancer activity at the locus of interest (Supplementary Figure 11)).

0 **Functional analyses of rs452384 on chromosome 5p15.33**

1 To functionally study rs452384 we used a 459-bp region from chr5:1330187-1331866 (hg19) containing
2 either the reference (risk) allele (T) or the alternate allele (C) upstream of the minimal promoter in pGL4.26 and
3 transfected these constructs into H1648 cells, which bear strong enhancer marks in this region
4 (<http://dbtss.hgc.jp>)[71] (Supplementary Figure S11). Both constructs elicited 5~6-fold higher luciferase activity
5 over background (empty pGL4.26), (Figure 5A), indicating an enhancer element is present. However, we did
6 not observe significant differences in enhancer activity between the two alleles in H1648 cells. To ensure that
7 we were not missing any key factors binding to adjacent sequences, we also studied a genomic fragment that
8 had been expanded to the 5' and 3' end (as indicated in Supplementary Figure 12), but it did not show allele-
9 specific activity either. We cannot exclude that differences in the TF profiles of AEC and H1648 might mask an
0 allele-specific effect, or that specific environmental conditions (such as tobacco smoke exposure) might be
1 required to reveal allele-specificity. Because this SNP had been highly ranked in a fine mapping study [72], we
2 also examined the online eQTL database GTEx [73] for evidence of lung-specific expression quantitative loci
3 (eQTLs), but detected no significant lung eQTL. We also looked for eQTLs in LUAD tissues (n=428) from
4 TCGA, correcting for copy number variation and DNA methylation (the latter might affect gene expression in
5 cancer), but detected no significant eQTLs either. The lack of an allele-specific difference in the luciferase
6 assay as well as the lack of a detectable eQTL in lung tissue suggests that, despite its presence in a TF
7 binding site in an AEC enhancer, this SNP may not be the functional SNP for index SNP rs465498. However, a
8 limitation of GTEx and other available eQTL databases is that data is typically generated from whole tissue

9 samples and rarely from purified cell populations. Thus, any true cell type-specific allelic expression will be
0 diluted by the presence of other cell types, such as lung fibroblasts, macrophages, or endothelial cells when
1 'lung eQTLs' are queried. In addition, we cannot exclude that a functional effect of this SNP may only be
2 observable under certain developmental or environmental conditions. We do note that in GTEx, the risk allele
3 was associated with higher expression of the adjacent telomerase reverse transcriptase (*TERT*) gene in
4 esophageal tissue ($P=3.9 \times 10^{-9}$, Figure 5B,D) and the cleft lip and palate transmembrane protein 1-like
5 (*CLPTM1L*) gene in sun-exposed skin ($P=2.1 \times 10^{-7}$, Figure 5D, E). Telomerase is a ribonucleoprotein
6 polymerase that is part of the complex that maintains telomere ends by addition of the telomere repeat
7 TTAGGG, and its overexpression is a key component of the transformation process in many malignant cancer
8 types including lung cancer [26-28, 57, 59, 74, 75]. *TERT* has been repeatedly implicated in lung cancer, both
9 through mutations and polymorphisms that lie in *TERT* and *CLPTM1L*. Indeed, *TERT* is significantly
0 overexpressed in LUAD compared to non-tumor lung tissues in TCGA data, in which expression is very low
1 ($P=1.2 \times 10^{-38}$, Figure 5C). *CLPTM1L* is also elevated in LUAD ($P=5.5 \times 10^{-62}$, Figure 5F) and was recently
2 functionally implicated in lung cancer susceptibility and resistance to chemotherapy [76]. One possible problem
3 in the ability to detect a *TERT* eQTL in lung tissue may be the very low expression levels of *TERT* in alveolar
4 epithelium (RPKM of around 1 in AT2 cells (D=0) and absent in cells at D4 and D6), and the possibility that the
5 SNP may only be functional under conditions that would promote cell proliferation, such as lung damage. Of
6 note, rs452384 is located in DNase hypersensitive sites in small airway epithelial cells and numerous other cell
7 types in the publicly available SNP annotation database HaploReg v4 [77].
8

9 **Functional analyses of rs6942067 and rs9481728 on chromosome 6q22.1**

0 To test for a functional role of rs6942067, we cloned an 836 bp region (chr6: 117785006-117785841,
1 hg19) containing either the reference (risk) allele (A) or the alternate allele (G) into pGL4.26. Based on the
2 presence of enhancer histone marks in the rs6942067 region [71], we performed reporter assays in LUAD cell
3 line H1648 (Supplementary Figure 11B). In H1648 cells, luciferase activity over background was 6.8-fold and
4 3.5-fold over background for the A and G alleles respectively, indicating a 94% elevated enhancer activity for
5 the reference allele ($P=2.7 \times 10^{-2}$) (Figure 6A). We replicated these findings in H522 cells and observed that the

6 reference and alternate constructs respectively elicited 6.9-fold and 4.8-fold elevated activity over background,
7 indicating a 45% stronger enhancer activity attributable to the reference (A) allele ($P=8 \times 10^{-3}$) (Supplementary
8 Figure 13).

9 Rs9481728 is located 30 kb downstream of rs6942067, in the first intron of *DCBLD1* (Figure 6D). We
0 cloned a 912 bp region (chr6: 117817044-117817955, hg19) in pGL4.26 and transfected the DNA into H1648
1 LUAD cells, which also showed an enhancer peak on this SNP (Supplementary Figure 11C). However, we
2 observed no activity over background (Supplementary Figure 14). While we cannot exclude that genomic
3 interactions with more distant factors may be required for enhancer activity of this segment, given the activity
4 observed using the region encompassing rs6942067, we pursued only rs6942067 further.

5 Examination of GTEx data for rs6942067 showed a strong genome-wide significant eQTL for *DCBLD1*
6 in lung (Figure 6B). No other genes were found to be eQTLs associated with this SNP in lung or other tissues.
7 To investigate more thoroughly, we also examined the flanking 1Mb window in the SNP region using TCGA
8 LUAD samples. This was done using rs6930292, a SNP in high LD ($r^2=1$) as a surrogate, because SNP
9 rs6942067 is not annotated in TCGA. We detected significant allele-specific expression of *DCBLD1* ($p=0.041$,
0 corrected for the number of genes in the 1Mb window, Figure 6C). Surprisingly, in contrast to the luciferase-
1 based assay, we detected lower expression of the reference (risk) allele, suggesting that in the genomic
2 context, the reference allele might reduce, not enhance expression. However, we note that in GTEx, the risk
3 allele is associated with higher *DCBLD1* expression in blood cells but reduced expression in thyroid tissue
4 (Figure 6E,F), supporting a complex role of the SNP in regulating gene expression. It is of interest that we
5 predicted binding of THRA on the alternate allele (Figure 3C) and that this receptor can repress transcription in
6 the absence of thyroid hormone and induce it in the presence of hormone [78]. Further investigations will be
7 required to elucidate the role of rs6942067 and its TFs in LUAD risk.

8 The function of *DCBLD1* is currently unclear, making it difficult to predict how up- or down-regulation of
9 this gene would affect lung cancer risk. The encoded protein is predicted to be membrane-associated. Its
0 paralog NRP2, or VEGF165R2, is a transmembrane protein that interacts with vascular endothelial growth
1 factor (VEGF) and is implicated in metastasis [79]. *DCBLD1* is significantly overexpressed in LUAD vs. non-
2 tumor lung in TCGA data ($P=1.7 \times 10^{-23}$, Figure 6G), and higher expression in LUAD is negatively associated
3 with survival (Supplementary Figure 15). It is also possible that the enhancer at rs6942067 targets other

4 gene(s) that may not be detectable as an eQTL in the mix of lung cells analyzed in GTEx. rs6942067 lies
5 between *DCBLD1* and *ROS1* (ROS proto-oncogene 1, a receptor tyrosine kinase)(Figure 6D). Translocations
6 involving *ROS1* are known “driver” events in NSCLC; approximately 1% of lung tumors harbor *ROS1* fusions
7 [80]. It has been reported that the level of certain histone modifications influences the predisposition to
8 chromosome translocations [81]. Thus, effects of this SNP on the epigenetic landscape and gene translocation
9 might need to be considered.

1 **Implications of our findings**

2 In this study, we identified 47 AEC enhancer-associated SNPs from 948 LUAD candidate risk SNPs. To
3 ensure we examined SNPs in true enhancers we focused on SNPs located in regions carrying *both* H3K4me1
4 and H3K27ac marks. However, it is possible that additional SNPs of interest are located in poised enhancers
5 that are marked only with H3K4me1 and are activated by environmental stimuli such as exposure to tobacco
6 smoke (the AECs used in this study were derived from a non-smoker). Of the 47 SNPs, we chose the 11 SNPs
7 that were in FAIRE peaks for further study. In the future, it will be important to investigate SNPs in regulatory
8 marks in AEC from smoker’s lungs.

9 Our analyses do not provide strong evidence for a functional role of rs452384 on chromosome 5p15.33
0 in LUAD, despite recent fine mapping data in the *TERT/CLPTM1L* region which identified rs452384 as one of
1 the top-ranked SNPs in the region [72]. How this SNP might function may require functional analyses using
2 primary AEC or as yet unavailable immortalized AEC or a variety of environmental conditions such as
3 exposure to tobacco smoke.

4 The reference (risk) allele of rs6942067 was associated with elevated enhancer activity in LUAD cell
5 lines, but we detected an eQTL showing lower expression of *DCBLD1* in lung tissue carrying the reference
6 allele. This emphasizes that luciferase assays may be helpful in detecting allele-specific activity but that the
7 cell type or genomic context may affect the observed regulatory effects. The versatility of this SNP is supported
8 by the differential *DCBLD1* eQTLs seen in different tissues (Figure 6B,E,F). EP300, which has been predicted
9 to bind to the reference allele and was observed to do so in ChIP studies, is a histone acetyltransferase
0 associated with gene activation. A lower activity of the reference allele might therefore be associated with
1 competition of other factors with EP300 for binding. A role for *DCBLD1* in lung cancer is supported by its

2 increased expression in LUAD vs. non-tumor lung and by its negative association with survival. A recent study
3 of non-smoking Asian women found that the association of rs9387478 (the index SNP for rs6942067) with
4 LUAD was stronger in EGFR mutation-positive cases [82], suggesting that the environment, ethnicity and
5 gender could all influence the manifestation of this genetic effect. Taken together, further studies into the
6 function of *DCBLD1* in lung cancer are certainly warranted.

7 Our work emphasizes the importance of integrating epigenomes of purified disease-relevant cell types
8 to elucidate the genetic basis for lung cancer risk. However, it also illustrates that focusing on disease-relevant
9 cells may not be sufficient to identify functional SNPs for most index SNPs. In the case of lung cancer, one
0 may require tobacco-exposed disease-relevant cells. Besides examining tobacco-smoke exposed cells and
1 tissues, purification of other lung cell types will be required to investigate the role of SNPs in other histological
2 subtypes of lung cancer [17]. A recent large-scale study of lung cancer susceptibility loci highlights the
3 differences in genetic susceptibility between histological lung cancer subtypes [83]. Basal cells, which are
4 airway epithelial cells that lie on the basement membrane and are thought to be involved in airway
5 regeneration upon injury, are implicated as progenitors of squamous cell carcinoma [84,85]. Basal cells should
6 be purified and epigenetically profiled in a similar fashion to the work described here. Further investigation of
7 SNPs and their targets may ultimately yield more effective and personalized strategies for lung cancer risk
8 assessment, prevention, and treatment.

0 **Methods & Materials**

1 **Ethics statement**

2 Remnant human transplant lung was obtained in compliance with Institutional Review Board-approved
3 protocols for the use of human source material in research (HS-07-00660) and processed within 3 days of
4 death.

5 **Cell isolation, culture and quality control**

6 Human AT2 cells were isolated from a de-identified non-smoker remnant transplant lung as described
7 [45]. The subject in this study was a non-smoking male 62-year old male who died of cardiovascular disease.
8 AT2 cells were plated and differentiated into AT1-like cells over the course of 6 days as described [45]. A549,

H1648, and H522 cells were cultured in RPMI 1640 w/L-glutamine (Lonza #12-702F, Walkersville, MD) supplemented with 10% fetal bovine serum and 100 units/mL of penicillin/streptomycin. All cell lines maintained by our laboratory are routinely tested and are negative for mycoplasma, and cell line identity was confirmed by DNA fingerprinting.

AEC RNA isolation, RNA-seq, protein isolation and Western blot analysis

RNA was obtained from day 0 (AT2), day 2 and day 4 (AT1-like) and day 6 (AT1-like) cells and sequenced. Briefly, total cell RNA was DNase I digested and then subjected to ribosomal RNA depletion with the Ribominus™ Eukaryote v2 kit (Life Technologies, # A15020, Grand Island, NY). Libraries were constructed with the TruSeq RNA Sample Prep Kit v2 (Illumina # RS-122-2001) and underwent Illumina HiSeq 2000 paired-end sequencing (2 x 50 bp) according to the manufacturer's instructions. Sequence reads were aligned to hg19 with TopHat2 v2.0.7 using default settings [86]. After mapping, data were further analyzed using Cufflinks [87] transcript assembly and quantification software (version 2.2.1) with default parameters and sequence bias detection and correction. Protein lysates obtained on days 0, 2, 4, and 6 were analyzed by Western blot to confirm proper *in vitro* differentiation, using antibodies specific for AT1 cells (anti-aquaporin 5 (AQP-005, Alomone Labs, Jerusalem, Israel) and anti-podoplanin (8.1.1, Santa Cruz Biotechnology, Santa Cruz, CA) and AT2 cells (anti-pro-surfactant protein C, AB3786, Millipore), with loading control anti-actin (AC-15 NB600-501, Novus Biologicals).

LUAD risk-associated SNPs collection

LUAD risk index SNPs were collected from published GWAS papers. A P -value cutoff of $\leq 5 \times 10^{-8}$ was applied for genome-wide significance. SNPs in high LD with LUAD index-SNPs ($r^2 > 0.5$) were retrieved using the online SNP annotation tool Haploreg v3, which calculates r^2 using data from the 1000 Genomes Project [40]. Because LD varies by ethnicity, this analysis was carried out taking into account the ethnicity of the population in which each index SNP was identified. Functional predictions for coding SNPs and miRNA targets were performed using ANNOVAR [88], which integrates data from PolyPhen-2 [89], SIFT [90] and TargetScan [91].

4 **ChIP-seq and FAIRE-seq**

5 Crosslinking and sonication for ChIP-seq and FAIRE-seq were performed using cells from day 0 (AT2),
6 and cells from day 4 and day 6 (AT1-like) cells as described [45]. For ChIP-seq, chromatin was incubated with
7 antibodies against H3K4me1 (Diagenode #pAb-037-050, Denville, NJ) and H3K27ac (Active Motif #39133,
8 Carlsbad, CA) after sonication, and enrichment of AEC ChIP targets was confirmed by qPCR for each cell
9 type. Libraries were created at the USC Epigenome Center and underwent Illumina GAI single-end
0 sequencing as previously described. We used cutadapt-1.5 to filter low-quality reads. Remaining high-quality
1 reads were aligned to reference human genome hg19 using bwa-0.7.7 with 2 mismatches allowed and
2 mapping quality thresholds set to 20. Duplicate reads were removed with picard-tools-1.107. Peaks were
3 called using SICER [92] (for FAIRE-seq, H3K27ac ChIP-seq, and H3K4me1 ChIP-seq respectively, window
4 sizes of: 50bp, 100bp, 200bp; gap sizes of 50bp, 200bp and 200bp were used with an FDR cutoff = 1×10^{-4}).
5 Saturation plots (Supplementary Figure 3) were generated using the same peak calling methods on a
6 proportion of reads. AEC ChIP-seq density plots (Supplementary Figure 5) were generated by running
7 annotatePeaks.pl in HOMER [93].

8 For featured LUAD cell lines, raw FASTQ data for H3K27ac and H3K4me1 were downloaded from
9 DBTSS (DataBase of Transcriptional Start Sites) database FTP (<http://dbtss.hgc.jp>) [71]. Quality control,
0 alignment, duplicate removal, and peak calling were performed as described for AECs.

1 **Identification of enhancer-associated SNPs**

2 Candidate enhancer-associated SNPs were identified based on their position in both H3K27ac and
3 H3K4me1 peaks in epigenomic data from day 0 (AT2), and/or day 4/day 6 (AT1-like). Candidate enhancer-
4 associated SNPs that were significantly enriched ($P < 1 \times 10^{-3}$) for both H3K27ac and H3K4me1 signals were
5 designated 'enhancer-associated SNPs' (AT1 or AT2-specific or general AEC enhancers). Enrichment analysis
6 was performed in a 0.5 kb window flanking each SNP using the HOMER script getDifferentialPeaks and the
7 R/Bioconductor packages rtracklayer and GenomicRanges (Supplementary Figure 6,
8 <https://www.bioconductor.org>).

9 **TF binding prediction**

0 Sequences (± 25 bp) around each AEC enhancer-associated SNPs containing ref/alt alleles were
1 extracted using the R package “BSgenome.Hsapiens.UCSC.hg19”. Motif positional weight matrixes (PWMs)
2 compiled from ENCODE-motif [48], Factorbook [49] and HOCOMOCO [50] were downloaded and transformed
3 into meme format. FIMO [95] was used to predict TF binding sites. We employed two approaches in TF motif
4 identification. First, we used FIMO to predict TF motifs that overlap each SNP, with a significance threshold of
5 $P=1 \times 10^{-4}$ based on Factorbook and HOCOMOCO motif databases. Secondly, we identified TFs with the
6 potential to bind specific SNPs via integrating ENCODE/GEO TF ChIP-seq peaks over the locations of SNPs
7 of interest. We then used FIMO to validate whether there are corresponding TF motifs formed/disrupted by the
8 SNP, and set a less stringent threshold of $P=1 \times 10^{-3}$ for this step because binding predictions are supported by
9 publicly-available ChIP-seq data for each factor at each SNP described and depicted in this manuscript. In
0 addition, we checked the genotype of the SNP in available ChIP-seq data to determine whether the expected
1 SNP allele forms the corresponding TF motif so that matched motif-TF events were identified. Furthermore, we
2 limited all of our TFBS analyses to those cognate TFs with an AEC RPKM ≥ 3 , ensuring predicted TFs are
3 expressed in AEC (Supplementary Figure 7).

4 **Plasmid construction and luciferase enhancer assays**

5 Putative enhancers spanning each SNP as well as the nearest DNase HSS in A549 cells (ENCODE)
6 were amplified by PCR from normal human male DNA (Promega #G1471, Madison, WI) using the primers
7 listed in Supplementary Table 7. Amplicons were subcloned into the pGL4.26 luciferase plasmid (Promega,
8 #E844A) upstream of a minimal promoter. The NEB Q5® Site-Directed Mutagenesis Kit (New England Biolabs
9 #E0554S, Ipswich, MA) was used to change the allele under study. All constructs were verified by sequencing.
0 All cell lines were transfected with the indicated constructs 48 hours prior to being harvested. A549 and H522
1 cells were transfected with Fugene HD (Promega #E2311) according to the manufacturer’s instructions and
2 H1648 cells were transfected with Lipofectamine® 3000 (ThermoFisher Scientific # L3000008, Waltham, MA).
3 Luciferase assays were performed with the Dual-Luciferase Reporter Assay System (Promega # E1960) on
4 the 96-well LUMIstar Omega Luminescence Microplate Reader (BMG LABTECH, Cary, NC) according to the
5 manufacturer’s instructions. All reported allele-specific enhancer activity represents the mean \pm S.E.M. of
6 three or more independent biological replicates assayed as technical triplicates.

eQTL tests

Tests in GTEx: eQTLs for rs452384 and rs6942067 were identified by directly searching the website of GTEx (The Genotype-Tissue Expression project) portal website [73].

Tests in TCGA: To identify eQTLs for rs6942067 in TCGA, we downloaded genomic information for 428 LUAD samples from the TCGA data portal website (<https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp>). This included level-3 RNA-seq data, genotype, and methylation data. Expression data were log₂ transformed ((log₂(rsem+1)). Level-4 copy number variation (CNV) data were downloaded directly from the website (http://gdac.broadinstitute.org/runs/analyses__2014_01_15/data/LUAD/20140115/). For each gene, we calculated the methylation level as the mean beta value of all the CpGs within ± 1 kb of the transcription start site (TSS). We used a multiple linear regression model to test the association between SNP genotype and gene expression, with adjustment for gene-level CNV and DNA methylation because CNV and DNA methylation might also influence gene expression. Since rs6942067 does not appear in the Affymetrix Genome-Wide Human SNP Array 6.0 platform (SNP 6.0), we chose rs6930292 ($r^2 = 1$, $D' = 1$ with rs6942067) as a surrogate for eQTL testing. For rs6942067, we performed eQTL analysis for rs6930292 against all genes within the flanking 1Mb region in the TCGA LUAD dataset. We also ruled out genes that are lowly expressed ((log₂(rsem + 1) < 3). FDR was calculated using the Benjamini-Hochberg method, and the FDR cutoff was set at 0.05. Plots were generated using ggplot2 in R 3.1.1.

TCGA RNA expression analysis

Gene expression FPKM data (fragments per kilobase of transcript per million reads) from 521 LUAD tumor samples and 59 normal samples were downloaded using the TCGAbiolinks package (<https://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html>) and log₂ transformed. Differential expression between tumor and normal tissue was assessed via Student's T-test. Plots were generated using ggplot2 in R 3.4.3.

AEC RNA expression analysis

RNA-seq was performed on the AEC sample for Day 0/2/4/6 respectively. Data were processed as described above in "AEC RNA isolation, RNA-seq, protein isolation and Western blot analysis". RPKMs were log₂-transformed to represent the expression level.

TCGA Kaplan-Meier analysis

5 We used the TCGAanalyze_SurvivalKM function included in the TCGAbiolinks package. After
6 removing duplicate tumor samples from the same subjects, we compared the survival of subjects in the
7 highest tertile of *DCBLD1* expression (in FKPM) to subjects in the lowest tertile of expression. This
8 resulted in a comparison of 161 subjects in the highest expression group to 161 subjects in the lowest
9 expression group.

0 Data Visualization

1 The Integrative Genomics Viewer (IGV) 2.3.34 was used to visually inspect and graph all sequencing
2 data [96]. All CHIP-seq QC plots were generated using R 3.1.1.

5 LEGENDS

6 **Figure 1. Data generation and analysis flow chart.** (A) 18 validated LUAD index SNPs with alleles indicated
7 (the risk allele is colored red). (B) 47 AEC enhancer SNPs identified by integrating AEC CHIP-seq data, with
8 alleles indicated (the inferred risk allele is colored red); presence in a FAIRE peak is marked as +. (C) SNP-
9 affected TF binding motifs predicted in AEC enhancer SNPs in FAIRE peaks; those TFs whose binding is also
0 supported by CHIP-seq evidence from publicly available datasets are colored red. SNPs for whom no TF binding
1 motifs were predicted to be affected and no TF binding by CHIP has been reported are indicated by a gray box.
2 (D) Luciferase enhancer assays for both alleles of regions containing rs452384 and rs6942067, tested in LUAD
3 cell line H1648. (E) Potential eQTLs in lung tissues of the two investigated candidate AEC enhancer SNPs.

4 **Figure 2. Epigenetic analysis of the rs452384 region.** (A) Rs452384 is located in an intron of *CLPTM1L*.
5 Histone enhancer marks and FAIRE peaks in the rs452384 region on day 0, 4 and 6 of AEC culture are
6 indicated, and underneath each track, peaks called using SICER are marked. (B) CHIP-seq data for TF NKX2-
7 1 from H3122 LUAD cells [51], which are homozygous for the reference (risk) T allele. (C) Predicted TF motifs
8 for the sequences containing the reference (risk) and alternate alleles.

9 **Figure 3. Epigenetic analysis of the rs6942067 region.** (A) Rs6942067 is located in the intergenic region
0 between gene *DCBLD1* and *ROS1*. Histone enhancer marks and FAIRE peaks in the rs6942067 region on day
1 0, 4 and 6 of AEC culture are indicated, and underneath each track, peaks called using SICER are marked. (B)

2 ChIP-seq data for TF EP300 from SK-N-SH neuroblastoma cells treated with retinoic acid [52], which are
3 homozygous for the reference (risk) A allele. (C) Predicted TF motifs for the sequences containing the
4 reference (risk) and alternate alleles.

5 **Figure 4. Epigenetic analysis of the rs9481728 region.** (A) Rs9481728 is located in an intron of *DCBLD1*.
6 Histone enhancer marks and FAIRE peaks in the rs9481728 region on day 0, 4 and 6 of AEC culture are
7 indicated, and underneath each track, peaks called using SICER are marked. (B) ChIP-seq data for TF EP300
8 from SK-N-SH neuroblastoma cells treated with retinoic acid [52], which are homozygous for the reference
9 (risk) A allele. (C) Predicted TF motifs for the sequence containing the reference allele (no sites were predicted
0 for the alternate allele).

1 **Figure 5. Functional analyses for rs452384.** (A) Enhancer luciferase assay performed in H1648 cells:
2 relative luciferase signals were compared between cells transfected with the reference (T) or alternate (C)
3 allele plasmids. (B) Association between rs452384 (T/C) and *TERT* expression in esophageal mucosa from the
4 GTEx database. (C) Boxplot showing statistically significantly elevated expression of *TERT* in TCGA LUAD vs.
5 non-tumor lung tissue. (D) Relative genomic positions of *TERT*, *CLPTM1L*, and SNP rs452384. (E) Association
6 between rs452384 (T/C) and *CLPTM1L* expression in sun exposed skin from the GTEx database. (F) Boxplot
7 showing statistically significantly elevated expression of *CLPTM1L* in TCGA LUAD vs. non-tumor lung tissue.

8 **Figure 6. Functional analyses for rs6942067.** (A) Enhancer luciferase assay performed in H1648 cells
9 (459bp fragment, Figure S11B); relative luciferase signals were compared between cells transfected with either
0 the reference (A) or alternate (G) allele plasmids. (B) Association between rs6942067 (A/G) and *DCBLD1*
1 expression in lung from the GTEx database. (C) Association between rs6942067 and *DCBLD1* expression in
2 LUAD tumors from the TCGA database. (D) Relative genomic positions of *DCBLD1* and SNP rs6942067. (E,
3 F) The association between rs6942067 (A/G) and *DCBLD1* expression in blood cells (E) and thyroid (F) from
4 the GTEx database. (G) Boxplot showing statistically significantly elevated expression of *DCBLD1* in TCGA
5 LUAD vs. non-tumor lung tissue.

7 **Supporting information**

8 **Supplementary Figure 1. Flow chart of bioinformatic and experimental analyses.**

9 **Supplementary Figure 2. *In vitro* differentiation of primary alveolar epithelial cells** (A) Schematic of *in*
0 *vitro* differentiation of AT2 to AT1-like cells and isolation of RNA and epigenomic materials. (B) Western blot to
1 confirm AT2 cell to AT1-like cell differentiation. Lysates from day 0, 2, 4 and 6 in culture were probed by
2 Western blot with antibodies for beta-actin (loading control), aquaporin 5 (AQP5), an AT1 cell protein involved
3 in water homeostasis, podoplanin (PDPN), an AT1 cell membrane glycoprotein of unknown function, and
4 surfactant protein C (SFTPC), an AT2 cell marker which is not expressed in AT1 cells.

5 **Supplementary Figure 3. Saturation plots of ChIP-seq on AEC.** Plots of peak number vs. reads number or
6 relative reads number of (A) H3K27ac ChIP-seq, (B) H3K4me1 ChIP-seq, and (C) FAIRE-seq.

7 **Supplementary Figure 4. Venn diagram of ChIP-seq peaks.** Comparison of AEC ChIP-seq peaks between
8 D0, D4 and D6 for (A) H3K27Ac, (B) H3K4me1 and (C) FAIRE.

9 **Supplementary Figure 5. Relative ChIP-seq tag density around FAIRE peaks of AEC.** D0 (A), D4 (B) and
0 D6 (C).

1 **Supplementary Figure 6. Process for identification of strong candidate functional SNPs located**
2 **internally in enhancer histone marked regions.** (A) We chose SNPs that lie in both H3K4me1 peaks and
3 H3K27ac peaks as candidate enhancer-associated SNPs. In this example, SNP-2 and SNP-3 would be
4 selected as candidate enhancer-associated SNPs. (B) To rule out low-confidence candidate enhancer-
5 associated SNPs, we tested the enrichment of H3K4me1 and H3K27ac signals around the SNP (± 500 bp)
6 versus ChIP-seq input signals. As an example, SNP2 but not SNP3 would be ruled out of the enhancer-
7 associated SNP list.

8 **Supplementary Figure 7. Process for identification of TFs that may be binding to the SNP regions.**

9 **Supplementary Figure 8. Epigenetic landscape around rs452384.** From top to bottom are indicated:
0 RefSeq genes (if present), FAIRE peaks in AEC at days 0, 4, 6 in culture, DNase hypersensitivity peaks in
1 A549 cells (ENCODE data [48, 49], H3K27ac peaks (indicating active enhancers) in AEC at days 0, 4, 6 in
2 culture, H3K4me1 peaks (indicating poised or active enhancers) in AEC at days 0, 4, 6 in culture, H3K27ac
3 peaks in several LUAD cell lines (data from DBTSS [71]). Called peaks are underlined.

4 **Supplementary Figure 9. Epigenetic landscape around rs6942067.** From top to bottom are indicated:
5 RefSeq genes (if present), FAIRE peaks in AEC at days 0, 4, 6 in culture, DNase hypersensitivity peaks in
6 A549 cells (ENCODE data [48, 49], H3K27ac peaks (indicating active enhancers) in AEC at days 0, 4, 6 in

7 culture, H3K4me1 peaks (indicating poised or active enhancers) in AEC at days 0, 4, 6 in culture, H3K27ac
8 peaks in several LUAD cell lines (data from DBTSS [71]). Called peaks are underlined.

9 **Supplementary Figure 10. Epigenetic landscape around rs9481728.** From top to bottom are indicated:
0 RefSeq genes (if present), FAIRE peaks in AEC at days 0, 4, 6 in culture, DNase hypersensitivity peaks in
1 A549 cells (ENCODE data [48, 49], H3K27ac peaks (indicating active enhancers) in AEC at days 0, 4, 6 in
2 culture, H3K4me1 peaks (indicating poised or active enhancers) in AEC at days 0, 4, 6 in culture, H3K27ac
3 peaks in several LUAD cell lines (data from DBTSS [71]). Called peaks are underlined.

4 **Supplementary Figure 11. Histone enhancer marked areas containing the four top candidate SNPs in
5 LUAD cell lines showing regions cloned for functional assays.** Data is from the DBTSS database [71]. (A)
6 Histone modifications in H1648 cells for rs452384 and the enhancer region cloned. (B) Histone modifications in
7 H1648 for rs6942067 and the enhancer region cloned. (C) Histone modifications in H1648 cells for rs9481728
8 and the enhancer region cloned. (D) Histone modifications in A549 cells for rs11364096 and the enhancer
9 region cloned.

0 **Supplementary Figure 12. Histone enhancer marked areas in LUAD cell lines showing alternate region
1 cloned for functional assay.** As in Supplementary Figure 11A, but longer fragment around rs452384.

2 **Supplementary Figure 13. Enhancer luciferase assay of rs6942067 in H522 cells.** Relative luciferase
3 signals were compared between clones with either reference (A) allele or alternative (G) allele.

4 **Supplementary Figure 14. Enhancer luciferase assay of rs9481728 in H1648 cells.** Relative luciferase
5 signals were compared between clones with either reference (A) allele or alternative (G) allele.

6 **Supplementary Figure 15.** Kaplan Meier survival analysis of LUAD patients from TCGA using FPKM data.
7 161 subjects in the highest tertile of *DCBLD1* expression (in FKPM) were compared to 161 subjects in the
8 lowest tertile of expression.

9
0 **Supplementary Table 1. List of 930 LD SNPs.** The table is based on data from HaploReg (v3, [40]) ordered
1 by chromosome and position in the genome and includes the original index SNPs, which have been indicated
2 in bold. Columns indicate, from left to right, SNP ID, chromosome number, position (based on GRCh37/hg19),
3 reference and alternate alleles, LD indicators r^2 and D , input (index) SNP, population tested ASN=Asian,

4 EUR=European, population related to index SNP information, presence of Biofeatures in HaploReg, number of
5 biofeatures, the nature of observed biofeatures.

6 **Supplementary Table 2. Annotation of LUAD risk-associated SNPs**

7 **Supplementary Table 3. Functional prediction of LUAD risk-associated SNPs located in exons**

8 **Supplementary Table 4. Functional prediction of LUAD risk-associated SNPs located in UTRs**

9 **Supplementary Table 5. SNPs located in promoter regions**

0 **Supplementary Table 6. ChIP-seq and FAIRE-seq statistics**

1 **Supplementary Table 7. Cloning primers for pGL4.26/ luciferase reporter gene constructs**

2

3

4 **Summary Points**

- 5 • Integration of 948 candidate LUAD risk-associated SNPs with alveolar epithelial cell epigenomic
6 information identified 47 SNPs in putative enhancer regions, marking them as candidate functional
7 SNPS of increased interest.
- 8 • Focusing specifically on the 11 of the 47 SNPs that were located in open DNA as indicated by FAIRE
9 analyses, 7 were predicted to disrupt transcription factor binding sites.
- 0 • Genomic fragments containing rs452384 showed increased luciferase activity over background
1 (reporter vector lacking a genomic fragment), suggesting the region contains an enhancer, but no allele
2 specific activity or lung eQTLs were detected.
- 3 • A genomic fragment carrying rs6942067 showed increased expression of the risk allele when examined
4 in a luciferase reporter assay, indicating presence of an enhancer and the potential of the SNP to affect
5 gene expression.
- 6 • eQTL of *DCBLD1* was observed in GTEx for rs6942067 in lung as well as thyroid, with higher
7 expression of the alternate allele, while blood showed eQTL of *DCBLD1* with higher expression of the
8 reference allele, suggesting the role of this SNP is likely complex.

- 9
- TCGA LUAD data showed increased *DCBLD1* expression in cells containing homozygous alternate
- 0 alleles, and *DCBLD1* is significantly overexpressed in LUAD tumor vs. non-tumor lung in TCGA,
- 1 supporting a possible role of this gene in LUAD development or progression.

2

3 References

- 4 1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J. Clin.* 61(2), 69-90
- 5 (2011).
- 6 2. Toh CK, Lim WT. Lung cancer in never-smokers. *J. Clin. Pathol.* 60(4), 337-40 (2007).
- 7 3. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers--a different disease. *Nat. Rev. Cancer.* 7(10), 778-90
- 8 (2007).
- 9 4. American Cancer Society. Cancer Facts & Figures 2015 (2015).
- 0 www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2015/index
- 1 5. Brennan P, Hainaut P, Boffetta P. Genetics of lung-cancer susceptibility. *Lancet Oncol.* 12(4), 399-408 (2011).
- 2 6. Lan Q, Hsiung CA, Matsuo K *et al.* Genome-wide association analysis identifies new lung cancer susceptibility loci
- 3 in never-smoking women in Asia. *Nat Genet.* 44(12), 1330-5 (2012).
- 4 7. Li Y, Sheu CC, Ye Y *et al.* Genetic variants and risk of lung cancer in never smokers: a genome-wide association
- 5 study. *Lancet Oncol.* 11(4), 321-30 (2010).
- 6 8. Coetzee SG, Rhie SK, Berman BP, Coetzee GA, Noushmehr H. FunciSNP: an R/bioconductor tool integrating
- 7 functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. *Nucleic*
- 8 *Acids Res.* 40(18), e139 (2012).
- 9 9. Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the
- 0 functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin.* 8, 57 (2015).
- 1 **** A thorough review about the utility of epigenomic information to identify functional risk SNPs.**
- 2 10. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory
- 3 information in the human genome. *Genome Res.* 22(9), 1748-59 (2012).
- 4 11. Freedman ML, Monteiro AN, Gayther SA *et al.* Principles for the post-GWAS functional characterization of cancer
- 5 risk loci. *Nat. Genet.* 43(6), 513-8 (2011).
- 6 12. Farh KK, Marson A, Zhu J *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants.
- 7 *Nature.* 518(7539), 337-43 (2015).
- 8 13. Maurano MT, Haugen E, Sandstrom R *et al.* Large-scale identification of sequence variants influencing human
- 9 transcription factor occupancy in vivo. *Nat. Genet.* 47(12), 1393-401 (2015).
- 0 *** An great example or large scale functional analysis of SNPs to identify functional variants implicated in TF binding.**
- 1 14. Hazelett DJ, Rhie SK, Gaddis M *et al.* Comprehensive functional annotation of 77 prostate cancer risk loci. *PLoS*
- 2 *Genet.* 10(1), e1004102 (2014).
- 3 15. Rhie SK, Coetzee SG, Noushmehr H *et al.* Comprehensive functional annotation of seventy-one breast cancer risk
- 4 Loci. *PLoS One.* 8(5), e63925 (2013).
- 5 16. Yao L, Tak YG, Berman BP, Farnham PJ. Functional annotation of colon cancer risk SNPs. *Nat. Commun.* 5, 5114
- 6 (2014).
- 7 17. Stueve TR, Marconett CN, Zhou B *et al.* The importance of detailed epigenomic profiling of different cell types
- 8 within organs. *Epigenomics* 8(6), 817-829 (2016).
- 9 *** A manuscript emphasizing the importance and applications of obtaining epigenomes from many different cell types.**
- 0 18. Heintzman ND, Stuart RK, Hon G *et al.* Distinct and predictive chromatin signatures of transcriptional promoters
- 1 and enhancers in the human genome. *Nat. Genet.* 39(3), 311-8 (2007).
- 2 19. Heintzman ND, Hon GC, Hawkins RD *et al.* Histone modifications at human enhancers reflect global cell-type-
- 3 specific gene expression. *Nature.* 459(7243), 108-12 (2009).
- 4 20. Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O *et al.* Epigenomic enhancer profiling defines a signature of colon
- 5 cancer. *Science.* 336(6082), 736-9 (2012).
- 6 21. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes
- 7 risk-associated variants. *Nat. Genet.* 46(2), 136-43 (2014).
- 8 22. Gjoneska E, Pfenning AR, Mathys H *et al.* Conserved epigenomic signals in mice and humans reveal immune
- 9 basis of Alzheimer's disease. *Nature.* 518(7539), 365-9 (2015).
- 0 23. Wang Y, McKay JD, Rafnar T *et al.* Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat*
- 1 *Genet.* 46(7), 736-41 (2014).
- 2 24. Shiraishi K, Kunitoh H, Daigo Y *et al.* A genome-wide association study identifies two new susceptibility loci for
- 3 lung adenocarcinoma in the Japanese population. *Nat. Genet.* 44(8), 900-3 (2012).

- 4 25. Miki D, Kubo M, Takahashi A *et al.* Variation in TP63 is associated with lung adenocarcinoma susceptibility in
5 Japanese and Korean populations. *Nat. Genet.* 42(10), 893-6 (2010).
- 6 26. Hsiung CA, Lan Q, Hong YC *et al.* The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-
7 smoking females in Asia. *PLoS Genet.* 6(8), e1001051 (2010).
- 8 27. Landi MT, Chatterjee N, Yu K *et al.* A genome-wide association study of lung cancer identifies a region of
9 chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.* 85(5), 679-91 (2009).
- 0 28. Hu Z, Wu C, Shi Y *et al.* A genome-wide association study identifies two new lung cancer susceptibility loci at
1 13q12.12 and 22q12.2 in Han Chinese. *Nat. Genet.* 43(8), 792-6 (2011).
- 2 29. Dong J, Hu Z, Wu C *et al.* Association analyses identify multiple new lung cancer susceptibility loci and their
3 interactions with smoking in the Chinese population. *Nat. Genet.* 44(8), 895-9 (2012).
- 4 30. Wang Z, Seow WJ, Shiraishi K *et al.* Meta-analysis of genome-wide association studies identifies multiple lung
5 cancer susceptibility loci in never-smoking Asian women. *Hum. Mol. Genet.* 25(3), 620-9 (2016).
- 6 31. Dobbs LG, Johnson MD, Vanderbilt J, Allen L, Gonzalez R. The great big alveolar TI cell: evolving concepts and
7 paradigms. *Cell. Physiol. Biochem.* 25(1), 55-62 (2010).
- 8 32. Rackley CR, Stripp BR. Building and maintaining the epithelium of the lung. *J. Clin. Invest.* 122(8), 2724-30 (2012).
- 9 **** A review that provides a clear overview of cell types within the lung, information that is required when considering
0 progenitors of lung cancer**
- 1 33. Wang J, Edeen K, Manzer R *et al.* Differentiated human alveolar epithelial cells and reversibility of their
2 phenotype in vitro. *Am. J. Respir. Cell Mol. Biol.* 36(6), 661-8 (2007).
- 3 34. Ballard PL, Lee JW, Fang X *et al.* Regulated gene expression in cultured type II cells of adult human lung. *Am. J.*
4 *Physiol. Lung Cell Mol. Physiol.* 299(1), L36-50 (2010).
- 5 35. Danto SI, Shannon JM, Borok Z, Zabski SM, Crandall ED. Reversible transdifferentiation of alveolar epithelial cells.
6 *Am. J. Respir. Cell Mol. Biol.* 12(5), 497-502 (1995).
- 7 36. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and
8 cancer. *Nature.* 507(7491), 190-4 (2014).
- 9 *** A discussion of the alveolar compartment of the lung and its contribution to lung cancer development.**
- 0 37. Xu X, Rock JR, Lu Y *et al.* Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma.
1 *Proc. Natl. Acad. Sci. USA.* 109(13), 4910-5 (2012).
- 2 38. Lin C, Song H, Huang C *et al.* Alveolar type II cells possess the capability of initiating lung tumor development.
3 *PLoS One.* 7(12), e53817 (2012).
- 4 39. Mainardi S, Mijimolle N, Francoz S, Vicente-Dueñas C, Sánchez-García I, Barbacid M. Identification of cancer
5 initiating cells in K-Ras driven lung adenocarcinoma. *Proc. Natl. Acad. Sci. USA.* 111(1), 255-60 (2014).
- 6 40. Khurana E, Fu Y, Colonna V *et al.* Integrative annotation of variants from 1092 humans: application to cancer
7 genomics. *Science.* 342(6154), 1235587 (2013).
- 8 41. Falvella FS, Galvan A, Frullanti E *et al.* Transcription deregulation at the 15q25 locus in association with lung
9 adenocarcinoma risk. *Clin. Cancer Res.* 15(5), 1837-42 (2009).
- 0 42. Nguyen JD, Lamontagne M, Couture C *et al.* Susceptibility loci for lung cancer are associated with mRNA levels of
1 nearby genes in the lung. *Carcinogenesis.* 35(12), 2653-9 (2014).
- 2 43. Improgo MR, Scofield MD, Tapper AR, Gardner PD. The nicotinic acetylcholine receptor CHRNA5/A3/B4 gene
3 cluster: dual role in nicotine addiction and lung cancer. *Prog. Neurobiol.* 92(2), 212-26 (2010).
- 4 44. Doyle GA, Wang MJ, Chou AD *et al.* In vitro and ex vivo analysis of CHRNA3 and CHRNA5 haplotype expression.
5 *PLoS One.* 6(8), e23373 (2011).
- 6 45. Marconett CN, Zhou B, Rieger ME *et al.* Integrated transcriptomic and epigenomic analysis of primary human
7 lung epithelial cell differentiation. *PLoS Genet.* 9(6), e1003513 (2013).
- 8 46. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements)
9 isolates active regulatory elements from human chromatin. *Genome Res.* 17(6), 877-85 (2007).
- 0 47. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature.* 489(7414),
1 109-13 (2012).
- 2 *** A manuscript providing information on long range interactions between regulatory elements such as enhancers and
3 their target genes.**
- 4 48. Kheradpour P, Kellis M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding
5 experiments. *Nucleic Acids Res.* 42(5), 2976-87 (2014).

- 6 49. Wang J, Zhuang J, Iyer S *et al.* Factorbook.org: a Wiki-based database for transcription factor-binding data
7 generated by the ENCODE consortium. *Nucleic Acids Res.* 41(Database issue), D171-6 (2013).
- 8 50. Kulakovskiy IV, Medvedeva YA, Schaefer U *et al.* HOCOMOCO: a comprehensive collection of human
9 transcription factor binding sites models. *Nucleic Acids Res.* 41(Database issue), D195-202 (2013).
- 0 51. Marconett CN, Zhou B, Siegmund KD, Borok Z, Laird-Offringa IA. Transcriptomic Profiling of Primary Alveolar
1 Epithelial Cell Differentiation in Human and Rat. *Genom Data.* 2, 105-9 (2014).
- 2 52. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements
3 in the human genome. *Nature.* 489(7414), 57-74 (2012).
- 4 **** A solid primer on ENCODE, which provides an important resource of public epigenomic data.**
- 5 53. Watanabe H, Francis JM, Woo MS *et al.* Integrated cistromic and expression analysis of amplified NKX2-1 in lung
6 adenocarcinoma identifies LMO3 as a functional transcriptional target. *Genes Dev.* 27(2), 197-210 (2013).
- 7 54. James MA, Vikis HG, Tate E, Rymaszewski AL, You M. CRR9/CLPTM1L regulates cell survival signaling and is
8 required for Ras transformation and lung tumorigenesis. *Cancer Res.* 74(4):1116-27 (2014).
- 9 55. James MA, Wen W, Wang Y *et al.* Functional characterization of CLPTM1L as a lung cancer risk candidate gene in
0 the 5p15.33 locus. *PLoS One.* 7(6), e36116 (2012).
- 1 56. Wang Z, Zhu B, Zhang M *et al.* Imputation and subset-based association analysis across different cancer types
2 identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum. Mol. Genet.*
3 23(24), 6616-33 (2014).
- 4 57. Chen XF, Cai S, Chen QG *et al.* Multiple variants of TERT and CLPTM1L constitute risk factors for lung
5 adenocarcinoma. *Genet. Mol. Res.* 11(1), 370-8 (2012).
- 6 58. Timofeeva MN, Hung RJ, Rafnar T *et al.* Influence of common genetic variation on lung cancer risk: meta-analysis
7 of 14 900 cases and 29 485 controls. *Hum. Mol. Genet.* 21(22), 4980-95 (2012).
- 8 59. Mocellin S, Verdi D, Pooley KA *et al.* Telomerase reverse transcriptase locus polymorphisms and cancer risk: a
9 field synopsis and meta-analysis. *J. Natl. Cancer Inst.* 104(11), 840-54 (2012).
- 0 60. Minoo P. Transcriptional regulation of lung development: emergence of specificity. *Respir. Res.* 1(2), 109-15
1 (2000).
- 2 61. Herriges M, Morrisey EE. Lung development: orchestrating the generation and regeneration of a complex organ.
3 *Development.* 141(3), 502-13 (2014).
- 4 62. Dang CV. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol. Cell. Biol.* 19(1), 1-11
5 (1999).
- 6 63. Seo AN, Yang JM, Kim H *et al.* Clinicopathologic and prognostic significance of c-MYC copy number gain in lung
7 adenocarcinomas. *Br. J. Cancer.* 110(11), 2688-99 (2014).
- 8 64. Iwakawa R, Kohno T, Kato M *et al.* MYC amplification as a prognostic marker of early-stage lung adenocarcinoma
9 identified by whole genome copy number analysis. *Clin. Cancer Res.* 17(6), 1481-9 (2011).
- 0 65. Iyer NG, Ozdag H, Caldas C. p300/CBP and cancer. *Oncogene.* 23(24), 4225-31 (2004).
- 1 66. Kan Z, Jaiswal BS, Stinson J *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers.
2 *Nature.* 466(7308), 869-73 (2010).
- 3 67. Rajatapiti P, Kester MH, de Krijger RR, Rottier R, Visser TJ, Tibboel D. Expression of glucocorticoid, retinoid, and
4 thyroid hormone receptors during human lung development. *J. Clin. Endocrinol. Metab.* 90(7), 4309-14 (2005).
- 5 68. Kolla V, Gonzales LW, Gonzales J *et al.* Thyroid transcription factor in differentiating type II cells: regulation,
6 isoforms, and target genes. *Am. J. Respir. Cell Mol. Biol.* 36(2), 213-25 (2007).
- 7 69. Hume R, Richard K, Kaptein E, Stanley EL, Visser TJ, Coughtrie MW. Thyroid hormone metabolism and the
8 developing human lung. *Biol. Neonate.* 80 Suppl 1, 18-21 (2001).
- 9 70. Hellevik AI, Asvold BO, Bjørro T, Romundstad PR, Nilsen TI, Vatten LJ. Thyroid function and cancer risk: a
0 prospective population study. *Cancer Epidemiol. Biomarkers Prev.* 18(2), 570-4 (2009).
- 1 71. Suzuki A, Wakaguri H, Yamashita R *et al.* DBTSS as an integrative platform for transcriptome, epigenome and
2 genome sequence variation data. *Nucleic Acids Res.* 43(Database issue), D87-91 (2015).
- 3 *** Description of DBTSS, a great resource for epigenomic information on lung cancer cell lines.**
- 4 72. Kachuri L, Amos CI, McKay JD *et al.* Fine mapping of chromosome 5p15.33 based on a targeted deep sequencing
5 and high density genotyping identifies novel lung cancer susceptibility loci. *Carcinogenesis.* 37(1), 96-105 (2016).
- 6 73. Consortium G. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 45(6), 580-5 (2013).
- 7 **** Description of the Genotype-Tissue Expression (GTEx) project, a public resource to identify eQTLs.**

- 8 74. Ma X, Gong R, Wang R *et al.* Recurrent TERT promoter mutations in non-small cell lung cancers. *Lung Cancer*.
9 86(3), 369-73 (2014).
- 0 75. Calado RT. Telomeres in lung diseases. *Prog. Mol. Biol. Transl. Sci.* 125, 173-83 (2014).
- 1 76. James MA, Wen W, Wang Y, *et al.* Functional characterization of CLPTM1L as a lung cancer risk candidate gene
2 in the 5p15.33 locus. *PLOS One* 7(6), e36116 (2012).
- 3 77. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target
4 genes for human complex traits and disease. *Nucleic Acids Res.* 44(D1), D877-81 (2016).
- 5 **** Description of HaploReg, a key resource to carry our data mining on candidate risk variants.**
- 6 78. Astapova I. Role of co-regulators in metabolic and transcriptional actions of thyroid hormone. *J. Mol. Endocrinol.*
7 56(3), 73-97 (2016).
- 8 79. Caunt M, Mak J, Liang WC *et al.* Blocking neuropilin-2 function inhibits tumor cell metastasis. *Cancer Cell*. 13(4),
9 331-42 (2008).
- 0 80. Takeuchi K, Soda M, Togashi Y *et al.* RET, ROS1 and ALK fusions in lung cancer. *Nat. Med.* 18(3), 378-81 (2012).
- 1 81. Burman B, Zhang ZZ, Pegoraro G, Lieb JD, Misteli T. Histone modifications predispose genome regions to
2 breakage and translocation. *Genes Dev.* 29(13), 1393-402 (2015).
- 3 82. Seow WJ, Matsuko K, Hsiung CA *et al.* Association between GWAS-identified lung adenocarcinoma susceptibility
4 loci and EGFR mutations in never-smoking Asian women, and comparison with findings from Western
5 populations. *Hum. Mol. Genet.* 26(2), 454-465 (2017).
- 6 83. McKay JD, Hung RJ, Han Y *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and
7 heterogeneity in genetic susceptibility across histological subtypes. *Nature Genet.* 49(7), 1126-1132 (2017).
- 8 *** A recent large analysis showing SNPS exhibit differential risk based lung cancer histological subtype.**
- 9 84. Sutherland KD, Berns A. Cell of origin of lung cancer. *Mol. Oncol.* 4(5), 397-403 (2010).
- 0 85. Van de Laar E, Clifford M, Hasenoeder S *et al.* Cell surface marker profiling of human tracheal basal cells reveals
1 distinct subpopulations, identifies MST1/MSP as a mitogenic signal, and identifies new biomarkers for lung
2 squamous cell carcinomas. *Respir. Res.* 15, 160 (2014).
- 3 86. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in
4 the presence of insertions, deletions and gene fusions. *Genome Biol.* 14(4), R36 (2013).
- 5 87. Trapnell C, Williams BA, Pertea G *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated
6 transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28(5), 511-5 (2010).
- 7 88. San Lucas FA, Wang G, Scheet P, Peng B. Integrated annotation and analysis of genetic variants from next-
8 generation sequencing studies with variant tools. *Bioinformatics.* 28(3), 421-2 (2012).
- 9 89. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2.
0 *Curr Protoc Hum Genet.* Chapter 7, Unit7.20 (2013).
- 1 90. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31(13),
2 3812-4 (2003).
- 3 91. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs.
4 *Genome Res.* 19(1), 92-105 (2009).
- 5 92. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains
6 from histone modification ChIP-Seq data. *Bioinformatics.* 25(15), 1952-8 (2009).
- 7 93. Heinz S, Benner C, Spann N *et al.* Simple combinations of lineage-determining transcription factors prime cis-
8 regulatory elements required for macrophage and B cell identities. *Mol Cell.* 38(4), 576-89 (2010).
- 9 94. Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9(9), R137 (2008).
- 0 95. Charles E. Grant TLB, and William Stafford Noble. FIMO: Scanning for occurrences of a given motif.
1 *Bioinformatics.* 27(7), 1017-1018 (2011).
- 2 96. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data
3 visualization and exploration. *Brief Bioinform.* 14(2), 178-92 (2013).
- 4

Table 1. Lung adenocarcinoma-associated risk index SNPs, listed by chromosome

Chr	Position (hg19)	Genes	SNPs	Population	Annotation	P values in GWAS [ref]
3	189356261	<i>TP63</i>	rs4488809	Asian	Intron	4.2×10^{-25} [28]
3	189357602	<i>TP63</i>	rs13314271	European	Intron	7.22×10^{-10} [23]
3	189383183	<i>TP63</i>	rs10937405	Asian	Intron	7×10^{-17} [24] 7×10^{-12} [25]
5	1286516	<i>TERT</i>	rs2736100	Asian, European	Intron	2.50×10^{-32} [24] 4×10^{-27} [6] 2.60×10^{-20} [26] 3×10^{-11} [25] 3.74×10^{-14} [27]
5	1287194	<i>TERT</i>	rs2853677	Asian	Intron	3×10^{-40} [24]
5	1325803	<i>CLPTM1L</i>	rs465498	Asian	Intron	1.20×10^{-13} [28]
5	146644115	<i>STK32A</i>	rs2895680	Asian	Intron	3.22×10^{-11} [29]
6	32368087	<i>BTNL2</i>	rs3817963	Asian	Intron	3×10^{-10} [24]
6	32433167	<i>HLA class II</i>	rs2395185	Asian	Intergenic	9.47×10^{-10} [6]
6	41493412	<i>FOXP4</i>	rs7741164	Asian	Intron	1.22×10^{-12} [30]
6	117786180	<i>ROS1, DCBLD1</i>	rs9387478	Asian	Intergenic	1.55×10^{-9} [6]
9	22160087	<i>CDKN2B-AS1</i>	rs72658409	Asian	Intergenic	1.94×10^{-9} [30]
10	114498476	<i>VTI1A</i>	rs7086803	Asian	Intron	1.19×10^{-11} [6]
12	52349071	<i>ACVR1B</i>	rs11610143	Asian	Intron	2.25×10^{-9} [30]
13	24293859	<i>MIPEP</i>	rs753955	Asian	Intergenic	3.90×10^{-10} [28]
15	78806023	<i>HYKK (AGPHD1)</i>	rs8034191	European	Intron	1.46×10^{-15} [27]
15	78894339	<i>CHRNA3</i>	rs1051730	European	Exon: Synonymous	7.10×10^{-19} [27]
17	65898809	<i>BPTF</i>	rs7216064	Asian	Intron	7×10^{-11} [24]

Note: To be included in our study, LUAD risk must have been specifically mentioned as affected by the SNP and the reported p-value must be $p \leq 5 \times 10^{-8}$. Chr: Chromosome; Position (hg19): Based on human genome 19 numbering; Genes: gene or nearest gene (if the SNP is not in a gene body) using Human Genome Organization name. GWAS: genome-wide association study.

Table 2. Enhancer-associated SNPs in lung adenocarcinoma, listed by chromosome

rsID	Chr	Pos (hg19)	Ref	Alt	r ² **	D' **	Index SNP Population***	AEC_Type	Nearest Gene	Annotation	FAIRE-seq Peak
rs7631358	chr3	189348411	G	A	0.75	-0.94	rs4488809_ASN	AT1_specific	TP63	Intergenic	
rs76937731	chr3	189348968	GCTCCA	G	0.61	-0.92	rs4488809_ASN	AT1_specific	TP63	Intergenic	
rs380145	chr5	1328897	C	T	0.62	0.95	rs465498_ASN	AT1_specific	CLPTM1L	Intron	
rs452932	chr5	1330253	T	C	0.92	0.99	rs465498_ASN	AT1_specific	CLPTM1L	Intron	
rs452384	chr5	1330840	T	C	0.92	0.99	rs465498_ASN	AEC_specific	CLPTM1L	Intron	+
rs370348	chr5	1331219	A	G	0.89	0.95	rs465498_ASN	AT2_specific	CLPTM1L	Intron	
rs2447853	chr5	1333077	A	G	0.66	0.88	rs465498_ASN	AT1_specific	CLPTM1L	Intron	
rs31489	chr5	1342714	C	A	0.9	0.97	rs465498_ASN	AT1_specific	CLPTM1L	Intron	
rs31490	chr5	1344458	G	A	0.9	0.97	rs465498_ASN	AT2_specific	CLPTM1L	Intron	
rs31037	chr5	146615785	T	G	0.66	0.84	rs2895680_ASN	AT1_specific	STK32A	Intron	
rs9381074	chr6	41505196	T	A	0.7	0.86	rs7741164_ASN	AEC_specific	FOXP4	Intron	
rs9401006	chr6	117732924	T	C	0.63	0.8	rs9387478_ASN	AT1_specific	ROS1	Intron	
rs9385012	chr6	117733191	T	C	0.64	0.81	rs9387478_ASN	AT1_specific	ROS1	Intron	
rs9401007	chr6	117733275	G	A	0.64	0.81	rs9387478_ASN	AT1_specific	ROS1	Intron	
rs9385013	chr6	117733452	C	T	0.64	0.81	rs9387478_ASN	AT1_specific	ROS1	Intron	
rs1407185	chr6	117733650	A	T	0.61	0.82	rs9387478_ASN	AEC_specific	ROS1	Intron	
rs1407184	chr6	117733717	G	A	0.71	0.93	rs9387478_ASN	AEC_specific	ROS1	Intron	
rs1321816	chr6	117734533	T	C	0.66	0.84	rs9387478_ASN	AT1_specific	ROS1	Intron	
rs3777981	chr6	117735255	A	C	0.67	0.84	rs9387478_ASN	AEC_specific	ROS1	Intron	
rs2243	chr6	117737390	G	A	0.72	0.94	rs9387478_ASN	AEC_specific	ROS1	Intron	
rs9374658	chr6	117741495	T	G	0.73	0.94	rs9387478_ASN	AEC_specific	ROS1	Intron	
rs4945584	chr6	117750980	T	C	0.83	0.98	rs9387478_ASN	AT1_specific	ROS1	Intergenic	+
rs9372480	chr6	117760579	C	T	0.86	0.98	rs9387478_ASN	AT2_specific	ROS1	Intergenic	
rs4946254	chr6	117765313	G	A	0.87	0.97	rs9387478_ASN	AT2_specific	ROS1	Intergenic	
rs6941337	chr6	117768448	A	G	0.89	0.98	rs9387478_ASN	AEC_specific	ROS1	Intergenic	
rs6937083	chr6	117785308	A	T	1	1	rs9387478_ASN	AT1_specific	DCBLD1	Intergenic	
rs6942067	chr6	117785696	A	G	0.54	1	rs9387478_ASN	AT1_specific	DCBLD1	Intergenic	+
rs9320604	chr6	117816045	G	A	0.76	0.93	rs9387478_ASN	AT1_specific	DCBLD1	Intron	
rs4946259	chr6	117816093	A	G	0.56	0.95	rs9387478_ASN	AT1_specific	DCBLD1	Intron	
rs9481728	chr6	117817165	C	T	0.76	0.93	rs9387478_ASN	AEC_specific	DCBLD1	Intron	+
rs929057	chr6	117818911	T	A	0.76	0.93	rs9387478_ASN	AEC_specific	DCBLD1	Intron	+
rs929058	chr6	117819198	A	G	0.67	0.93	rs9387478_ASN	AEC_specific	DCBLD1	Intron	+
rs2057314	chr6	117819357	A	G	0.76	0.93	rs9387478_ASN	AEC_specific	DCBLD1	Intron	+
rs11364096	chr10	114502022	TG	T	1	1	rs7086803_ASN	AEC_specific	VTI1A	Intron	+
rs12217440	chr10	114502218	G	A	1	1	rs7086803_ASN	AT2_specific	VTI1A	Intron	
rs7094841	chr10	114502411	T	C	1	1	rs7086803_ASN	AEC_specific	VTI1A	Intron	
rs78223856	chr10	114527225	C	T	0.63	0.88	rs7086803_ASN	AT1_specific	VTI1A	Intron	+

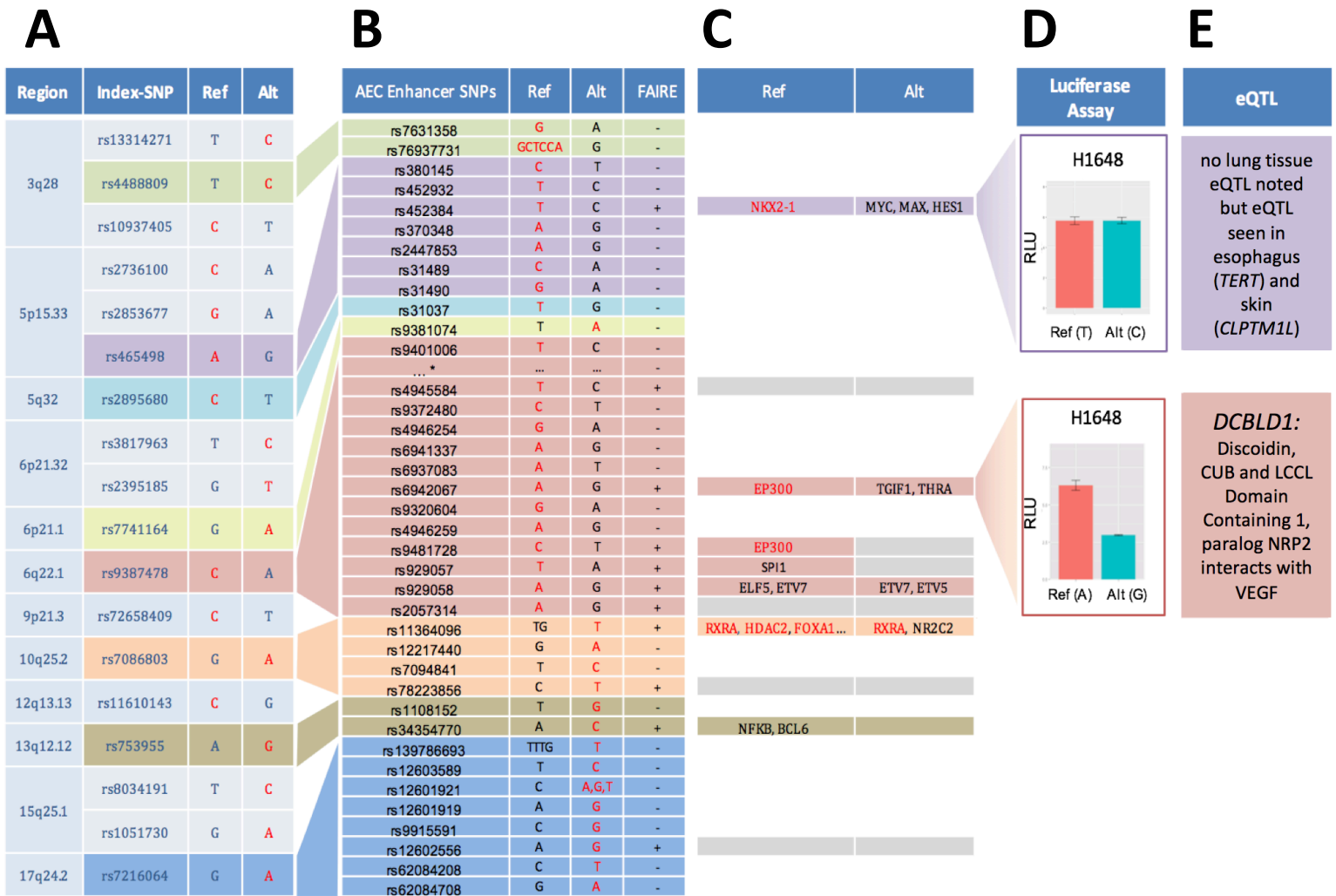
rs1108152	chr13	24312446	T	G	0.65	0.92	rs753955_ASN	AT1_specific	<i>MIPEP</i>	Intron	
rs34354770	chr13	24327618	A	C	0.59	0.88	rs753955_ASN	AEC_specific	<i>MIPEP</i>	Intron	+
rs139786693	chr17	65825228	TTTG	T	0.75	0.9	rs7216064_ASN	AT1_specific	<i>BPTF</i>	Intron	
rs12603589	chr17	65825248	T	C	0.8	0.91	rs7216064_ASN	AT1_specific	<i>BPTF</i>	Intron	+
rs12601921	chr17	65825354	C	A,G,T	0.72	0.92	rs7216064_ASN	AT1_specific	<i>BPTF</i>	Intron	
rs12601919	chr17	65825374	A	G	0.73	0.92	rs7216064_ASN	AT2_specific	<i>BPTF</i>	Intron	
rs9915591	chr17	65826090	C	G	0.81	0.92	rs7216064_ASN	AEC_specific	<i>BPTF</i>	Intron	
rs12602556	chr17	65826861	A	G	0.81	0.92	rs7216064_ASN	AEC_specific	<i>BPTF</i>	Intron	
rs62084208	chr17	65827443	C	T	0.81	0.92	rs7216064_ASN	AT1_specific	<i>BPTF</i>	Intron	
rs62084708	chr17	66049707	G	A	0.51	0.73	rs7216064_ASN	AT2_specific	<i>KPNA2</i>	Intergenic	

Chr: Chromosome; Position (hg19): Based on human genome 19 numbering; Ref: Reference allele (risk alleles indicated in bold); Alt: alternate allele (risk alleles indicated in bold); r2 and D': linkage disequilibrium information based on the population in which risk was identified for each SNP; ASN: Asian; AEC: alveolar epithelial cell; AT1: alveolar type 1 cell; AT2: alveolar type 2 cell; Nearest Gene: gene or nearest gene (if the SNP is not in a gene body) using Human Genome Organization name. GWAS: genome-wide association study. FAIRE-seq: formaldehyde-assisted identification of regulatory elements-sequencing.

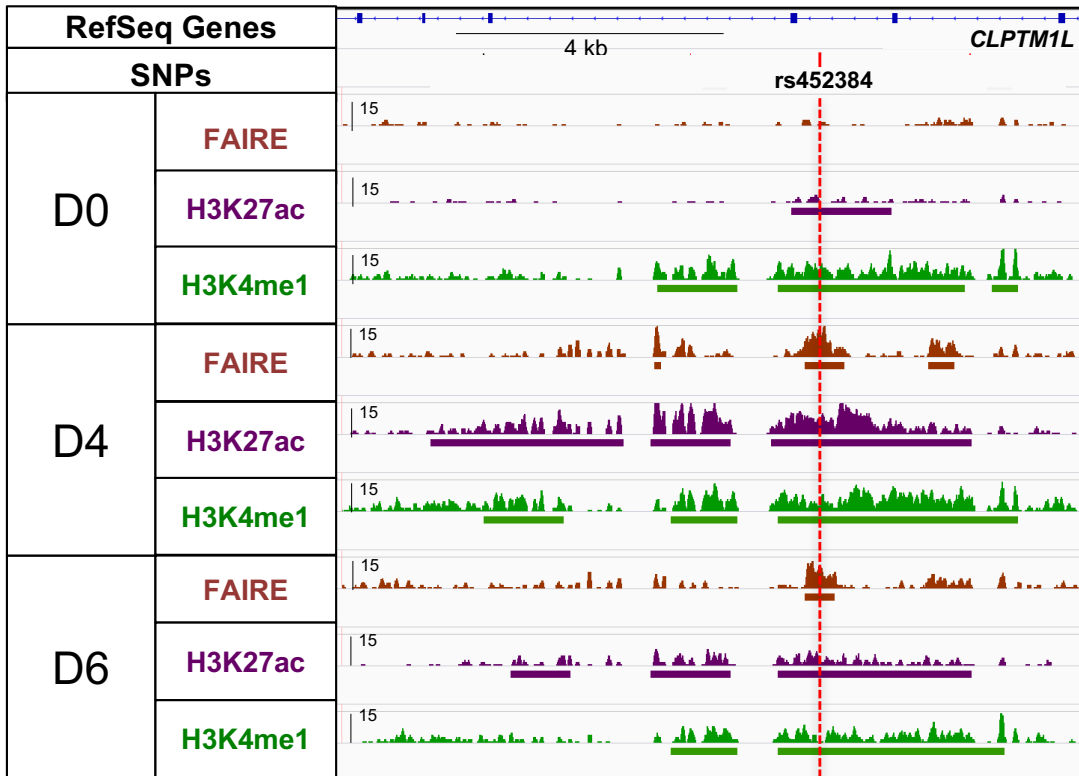
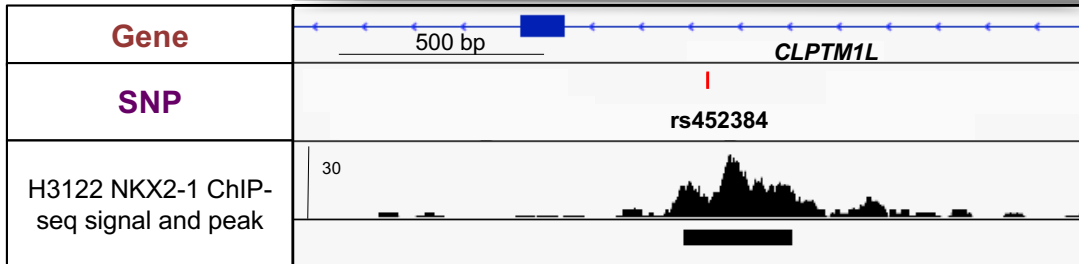
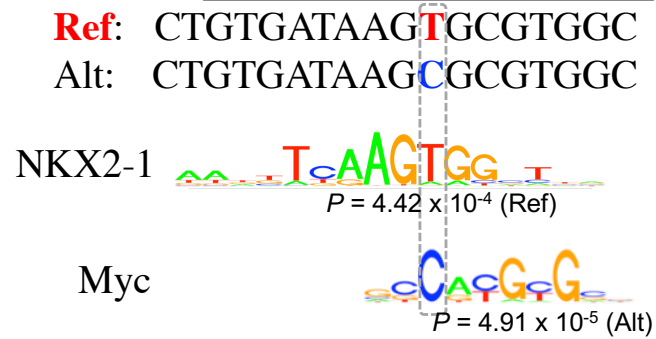
Table 3. Transcription factor binding motif prediction for AEC enhancer SNPs in FAIRE-seq peaks

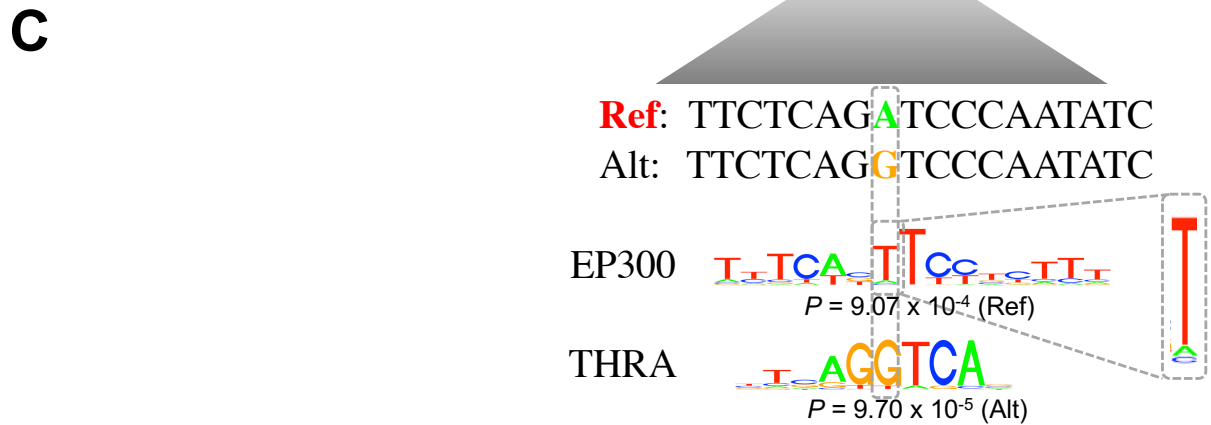
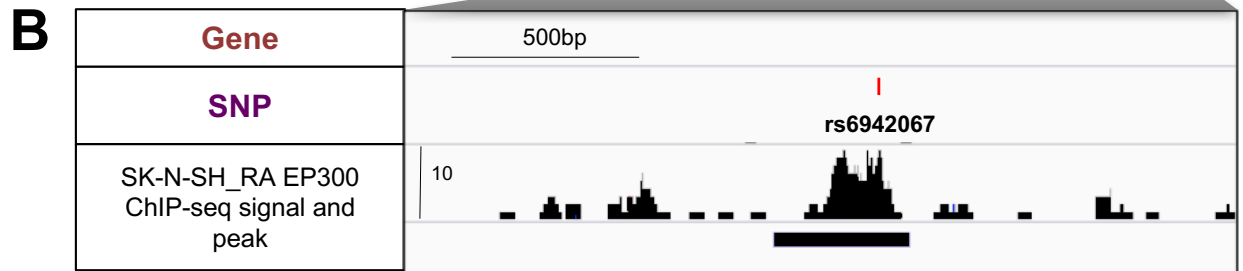
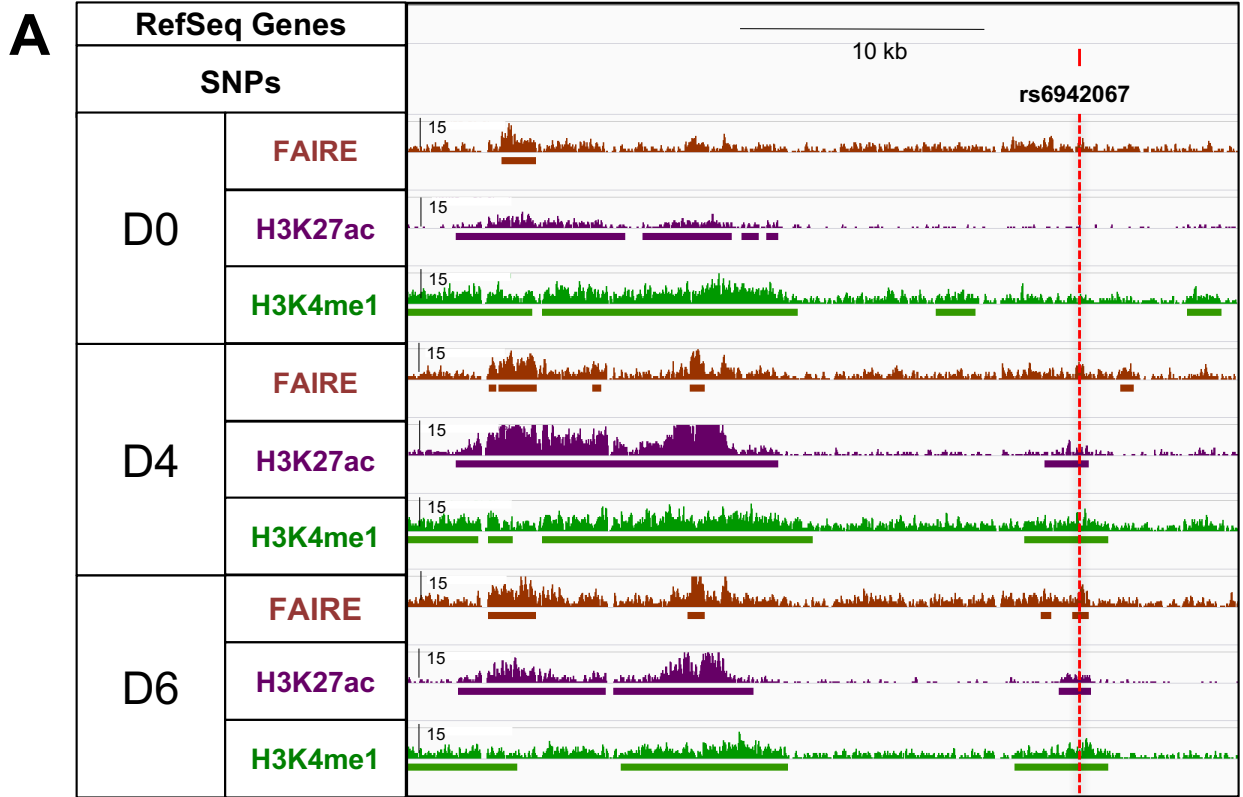
rsID	Ref	Alt	ChIP-seq TF Binding		
			Hom Ref	Het	Hom Alt
rs452384	NKX2-1	MYC, MAX, HES1	NKX2-1 , MYC, FOS		
rs6942067	EP300	TGIF1, THRA	POLR2A, EP300	POLR2A, FOSL2, GATA2, JUN, FOS	JUND
rs9481728	EP300		POLR2A, EP300	GATA2, FOS, MAFK	
rs929057	SPI1				
rs929058	ELF5, ETV7	ETV7, ETV5			
rs2057314			PolR2A	PolR2A, TCF7L2, TCF12, SPI1, FOS	NFIC, SPI1
rs11364096	RXRA, HDAC2, SP1, FOXA, NR1H2, PPARG	RXRA, NR2C2	RXRA, HDAC2, SP1, FOXA1, FOXA2, EP300, CEBPB, GATA2, TCF7L2, NFIC, ZNF217, ESR1, MYBL2, POLR2A, ARID3A, STAT1, STAT3, FOS, HNF4G, NKX2-1		
rs34354770	NFKB, BCL6		FOXA2		
rs12602556				RBBP5, EGR1	

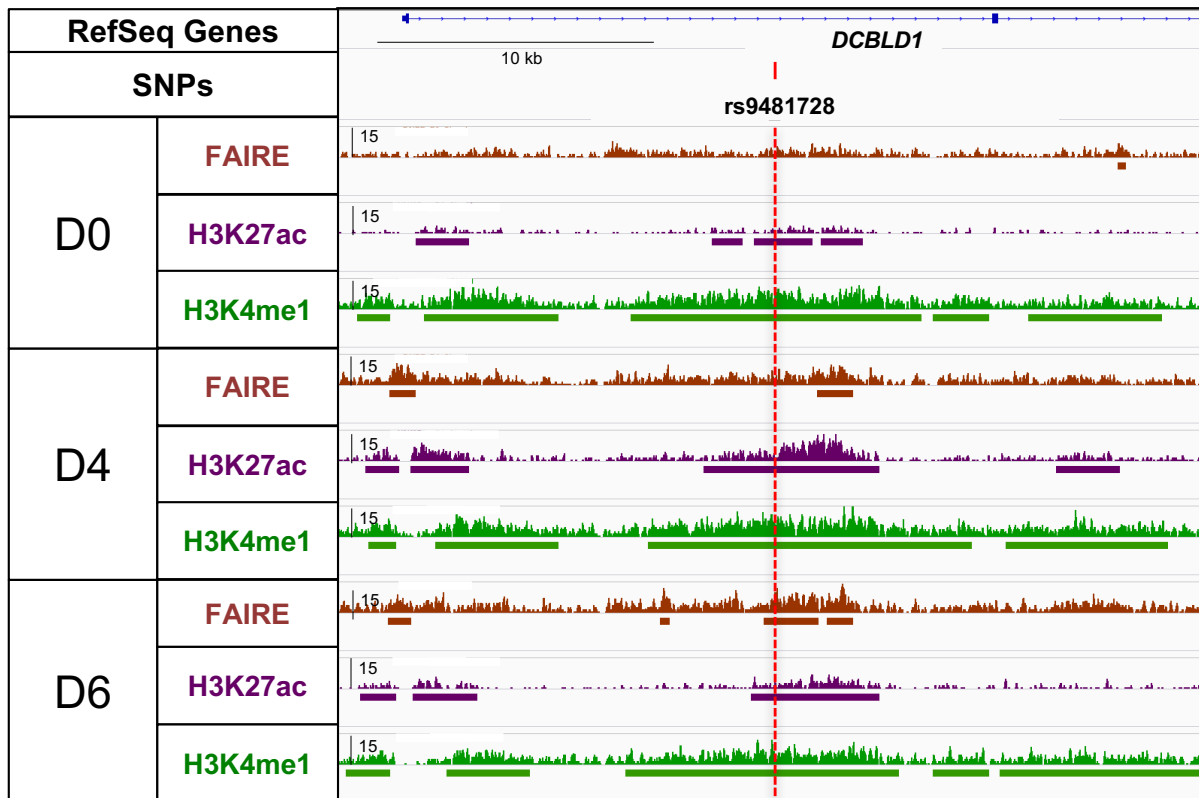
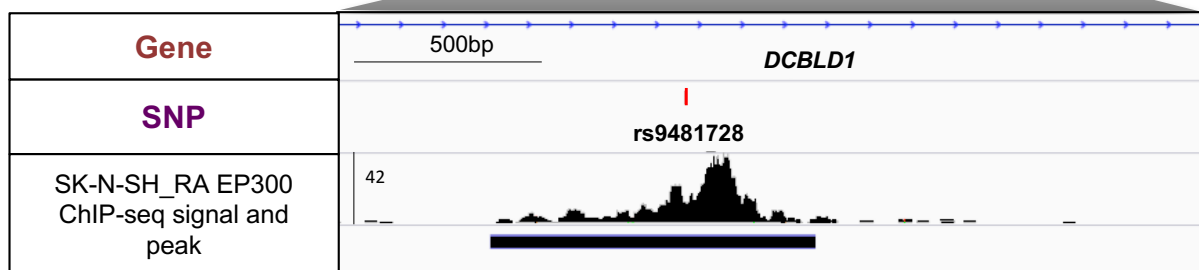
Note: Two of the 11 SNPs in FAIRE peaks (rs4945584 and rs78223856) showed no predicted TF motifs or evidence of ChIP-seq TF binding and are not listed. AEC: alveolar epithelial cell ; FAIRE-seq: formaldehyde-assisted identification of regulatory elements-sequencing; Ref: Reference allele (transcription factors (TFs) indicated in **bold** were both predicted *and* detected by ChIP-seq); Alt: alternate allele (transcription factors (TFs) indicated in **bold** were both predicted *and* detected by ChIP-seq); Hom Ref: ChIP-seq experiment carried out with cells containing a homozygous reference allele; Het: ChIP-seq experiment carried out with cells containing heterozygous alleles; Hom Alt: ChIP-seq experiment carried out with cells containing a homozygous alternate allele.



* 9 SNPs non-informative SNPs

A**B****C**

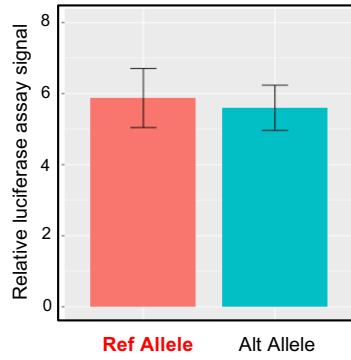


A**B****C**

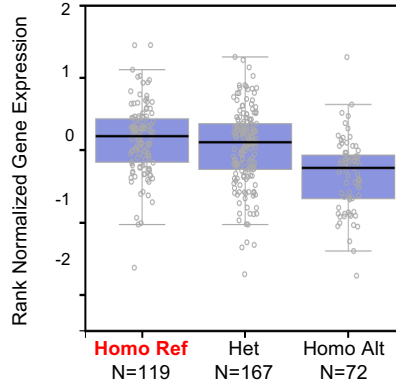
Ref: TAAGTGACACTTT
Alt: TAAGTGATACTTT
 EP300

 $P = 7.89 \times 10^{-4}$ (Ref)

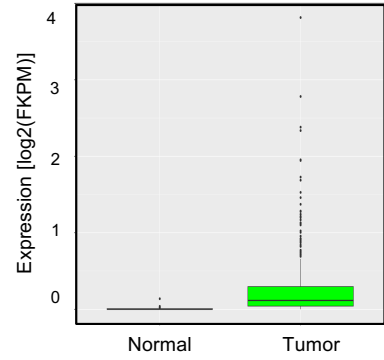
A Luciferase Assay in H1648 Cells
 $P > 0.05$



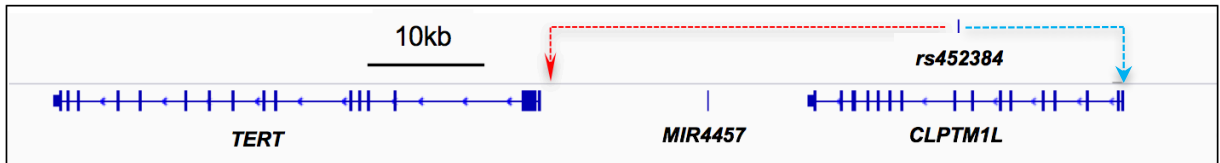
B GTEx Esophagus Mucosa eQTL
rs452384 ~ *TERT*
 $P = 3.9 \times 10^{-9}$



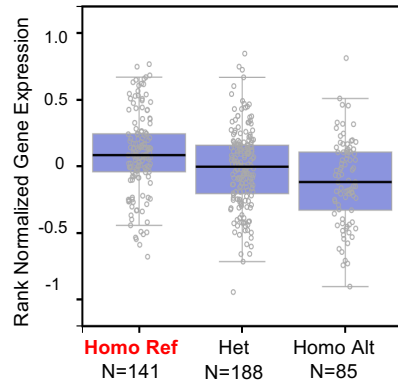
C *TERT* expression:
Normal vs. LUAD
 $P = 1.2 \times 10^{-38}$



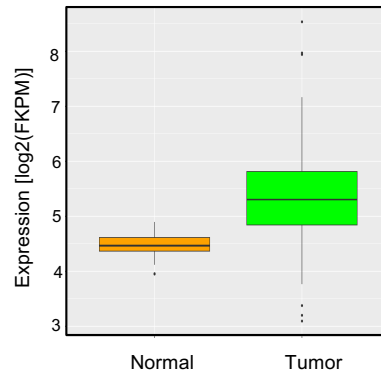
D

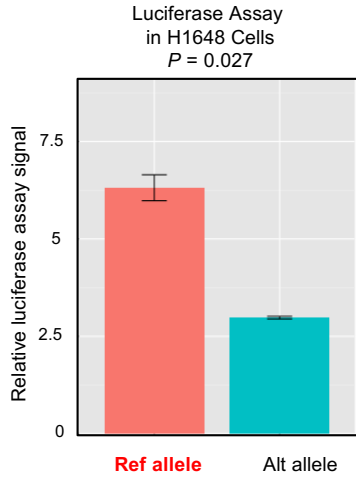
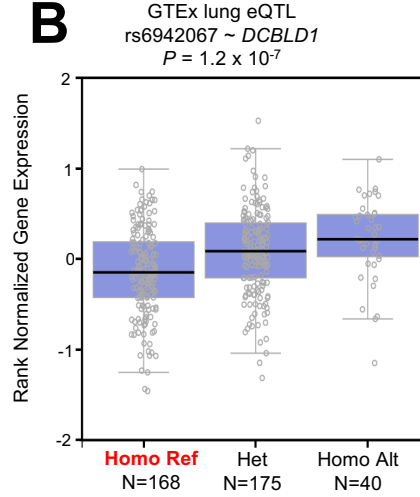
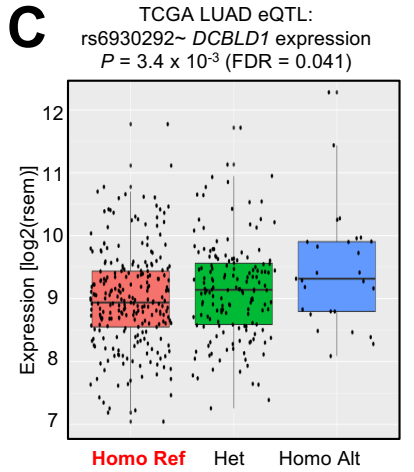
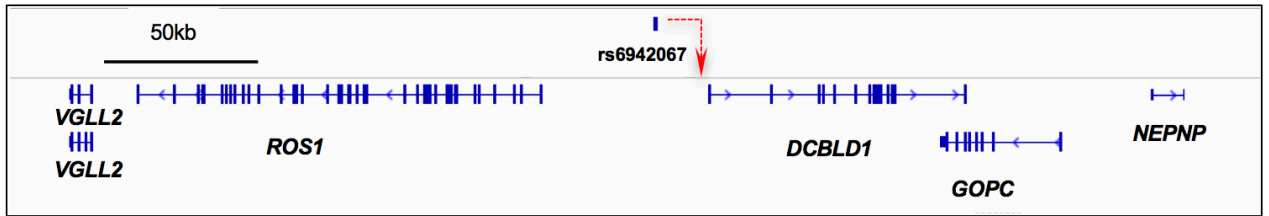
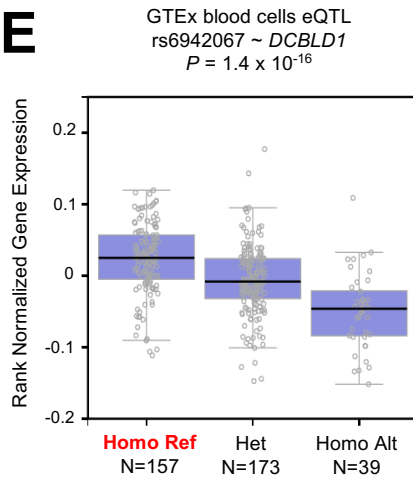
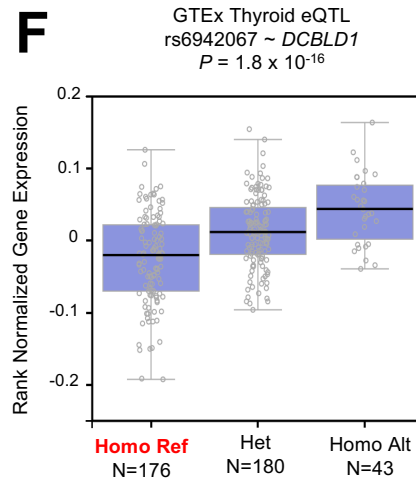
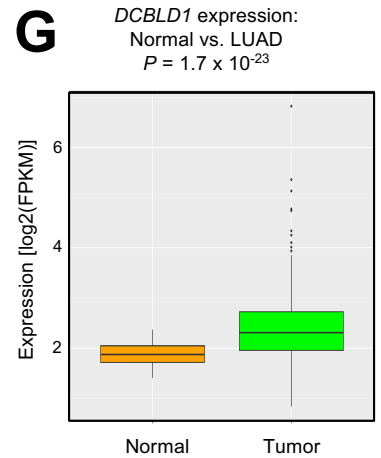


E GTEx Sun Exposed Skin eQTL
rs452384 ~ *CLPTM1L*
 $P = 2.1 \times 10^{-7}$



F *CLPTM1L* expression:
Normal vs. LUAD
 $P = 5.5 \times 10^{-62}$



A**B****C****D****E****F****G**

rsID	chr	pos	ref	alt	r2	D ₁	Input SNP	Population	IndexSNP Info	Bioteature_Overlap	Bioteature_Overlap Num	Bioteature_Overlap Names
rs73189230	chr3	189301627	C	T	0.6252	0.791033	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	FALSE	0NA	
rs3922955	chr3	189302969	A	G	0.61052	0.811033	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	FALSE	0NA	
rs17443378	chr3	189306277	T	C	0.61052	0.811033	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	FALSE	0NA	
rs9849804	chr3	189306797	C	T	0.51	-0.92	rs13314271	EUR	rs13314271;EUR	FALSE	0NA	
rs6444389	chr3	189323685	T	G	0.58	0.93	rs13314271	EUR	rs13314271;EUR	FALSE	0NA	
rs19120288	chr3	189325289	A	G	0.57	0.92	rs13314271	EUR	rs13314271;EUR	TRUE	2	FAIRE;FAIRE;D6;2013;3;MB1B2;W50;G50;FDR0.01;island
rs9822622	chr3	189326046	A	G	0.53	0.94	rs13314271	EUR	rs13314271;EUR	TRUE	1	FAIRE;D6;2013;3;MB1B2;W50;G50;FDR0.01;island
rs73194153	chr3	189327230	A	G	0.52	-0.93	rs13314271	EUR	rs13314271;EUR	FALSE	0NA	
rs4687073	chr3	189327405	A	C	0.52	-0.94	rs13314271	EUR	rs13314271;EUR	FALSE	0NA	
rs9882796	chr3	189330511	A	C	0.61	0.94	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs9812901	chr3	189331804	C	T	0.59	0.88	rs10937405	ASN	rs10937405;ASN	FALSE	0NA	
rs5626992	chr3	189334966	A	G	0.64059	0.85102	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	FALSE	0NA	
rs9851272	chr3	189335242	T	A	0.77063	0.91097	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	FALSE	0NA	
rs981000	chr3	189335606	C	A	0.65059	0.85102	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	FALSE	0NA	
rs36108040	chr3	189335844	A	C	0.82095	0.91097	rs4488809;rs13314271	ASN,EUR	rs4488809;ASN;rs13314271;EUR	FALSE	0NA	
rs10937395	chr3	189336370	G	A	0.59	-0.89	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs4687076	chr3	189336473	T	C	0.6	0.89	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs3514993	chr3	189337994	C	T	0.6	-0.89	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs28632188	chr3	189337839	G	A	0.76064	0.91097	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	FALSE	0NA	
rs200520205	chr3	189338304	MAC	A	0.51	0.82	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs6129848	chr3	189338305	AC	A	0.56	0.88	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs3456957	chr3	189339311	T	G	0.606	-0.891095	rs4488809;rs13314271	EUR	rs4488809;ASN;rs13314271;EUR	FALSE	0NA	
rs142098434	chr3	189339377	C	CAT	0.61	0.9	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs9872043	chr3	189339497	G	T	0.56	-0.97	rs13314271	EUR	rs13314271;EUR	FALSE	0NA	
rs9864659	chr3	189339649	A	G	0.61	0.9	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs1686462	chr3	189340147	G	A	0.61	-0.9	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs7622692	chr3	189340241	C	G	0.56	-0.97	rs13314271	EUR	rs13314271;EUR	FALSE	0NA	
rs7130869	chr3	189340886	G	A	0.61	-0.97	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs10937396	chr3	189341297	G	A	0.6706	-0.87092	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	FALSE	0NA	
rs7643161	chr3	189341790	G	A	0.8065	0.91097	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	FALSE	0NA	
rs4399410	chr3	189342639	C	A	0.73	0.94	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs17444559	chr3	189344011	C	T	0.81065	0.91097	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	FALSE	0NA	
rs6790258	chr3	189344126	T	A	0.55	0.93	rs10937405	ASN	rs10937405;ASN	FALSE	0NA	
rs9830151	chr3	189344149	C	T	0.63	-0.93	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs34829156	chr3	189345114	G	A	0.73	0.93	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs35133349	chr3	189345366	G	A	0.63	-0.93	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs3442923	chr3	189345380	A	T	0.63	-0.93	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs34505657	chr3	189345519	G	T	0.63	-0.93	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs71308970	chr3	189346531	G	A	0.63	-0.93	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs6797409	chr3	189346522	C	A	0.62	-0.92	rs4488809	ASN	rs4488809;ASN	TRUE	1	H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs983951	chr3	189346567	C	T	0.8015	0.91097	rs10937405;rs13314271	ASN,EUR	rs10937405;ASN;rs13314271;EUR	TRUE	1	H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs6800124	chr3	189348115	C	T	0.62	-0.92	rs4488809	ASN	rs4488809;ASN	TRUE	1	H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs35139818	chr3	189349599	C	G	0.73	-0.93	rs4488809	ASN	rs4488809;ASN	TRUE	1	H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs76313538	chr3	189348411	G	A	0.75	-0.94	rs4488809	ASN	rs4488809;ASN	TRUE	6	FAIRE;D6;2013;3;MB1B2;W50;G50;FDR0.01;island;FAIRE;D6;2013;3;MB1B2;W50;G50;FDR0.01;island;FAIRE;D6;2013;3;MB1B2;W50;G50;FDR0.01;island;FAIRE;D6;2013;3;MB1B2;W50;G50;FDR0.01;island;FAIRE;D6;2013;3;MB1B2;W50;G50;FDR0.01;island;FAIRE;D6;2013;3;MB1B2;W50;G50;FDR0.01;island
rs76937731	chr3	189348968	GCTCCA	G	0.61	-0.92	rs4488809	ASN	rs4488809;ASN	TRUE	5	island;H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs13064999	chr3	189349501	G	A	0.63	-0.93	rs4488809	ASN	rs4488809;ASN	TRUE	2	island;H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs13087793	chr3	189349742	T	C	0.63	-0.93	rs4488809	ASN	rs4488809;ASN	TRUE	2	island;H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs34234442	chr3	189349846	GA	G	0.63	-0.93	rs4488809	ASN	rs4488809;ASN	TRUE	2	island;H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs11928222	chr3	189350265	T	G	0.76	-0.95	rs4488809	ASN	rs4488809;ASN	TRUE	2	island;H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs36201505	chr3	189350483	G	A	0.65	-0.94	rs4488809	ASN	rs4488809;ASN	TRUE	2	island;H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs35832956	chr3	189350641	T	G	0.65	-0.94	rs4488809	ASN	rs4488809;ASN	TRUE	2	island;H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs35867177	chr3	189350987	A	G	0.65	-0.94	rs4488809	ASN	rs4488809;ASN	TRUE	2	island;H3Kme1;D0;2013;3;MB1B2;W50;G50;FDR0.01;island
rs12102970	chr3	189353098	G	A	0.63	-1	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs12102966	chr3	189353186	C	T	0.69	-1	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	
rs12374182	chr3	189354008	G	A	0.69	-1	rs4488809	ASN	rs4488809;ASN	FALSE	0NA	

fs5579747	CH3	189354127	A	C	0.99	0.99	1.0	0.99	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs34951828	CH3	189354655	T	C	0.69	0.69	-1	0.69	fs4488809	ASN	EUR	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs11170209	CH3	189354241	C	T	0.59	0.59	-0.99	0.59	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs4396800	CH3	189356221	G	A	0.93	0.93	0.96	0.93	fs10937405	fs13314271	ASN	EUR	fs10937405	ASN	fs13314271	EUR	fs10937405	ASN	fs13314271	EUR	FALSE	0NA
fs4488809	CH3	189356261	G	C	1.0	1.0	1.0	1.0	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs4381914	CH3	189356598	A	C	0.91	0.91	0.96	0.91	fs10937405	fs13314271	ASN	EUR	fs10937405	ASN	fs13314271	EUR	fs10937405	ASN	fs13314271	EUR	FALSE	0NA
fs4396881	CH3	189356665	G	A	0.91	0.91	0.96	0.91	fs10937405	fs13314271	ASN	EUR	fs10937405	ASN	fs13314271	EUR	fs10937405	ASN	fs13314271	EUR	FALSE	0NA
fs4342365	CH3	189356675	T	A	0.92	0.92	0.96	0.92	fs10937405	fs13314271	ASN	EUR	fs10937405	ASN	fs13314271	EUR	fs10937405	ASN	fs13314271	EUR	FALSE	0NA
fs7636839	CH3	189356941	A	T	0.99	0.99	1.0	0.99	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs13080835	CH3	189357199	G	T	0.99	0.99	1.1	0.99	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs1314271	CH3	189357602	T	C	0.99	0.99	1.1	0.99	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs1269594	CH3	189357616	A	C	0.98	0.98	1.1	0.98	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs12107251	CH3	189358862	A	T	0.95	0.95	0.96	0.95	fs10937405	fs13314271	ASN	EUR	fs10937405	ASN	fs13314271	EUR	fs10937405	ASN	fs13314271	EUR	FALSE	0NA
fs55862124	CH3	189358862	A	G	0.66	0.66	-0.96	0.66	fs4488809	ASN	EUR	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	TRUE	1FAIR
fs12100205	CH3	189359013	T	A	0.97	0.97	0.99	0.97	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	TRUE	1FAIR
fs7619157	CH3	189360235	T	C	0.97	0.97	0.99	0.97	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	TRUE	1FAIR
fs7629383	CH3	189360406	A	G	0.97	0.97	0.99	0.97	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	1FAIR
fs13084874	CH3	189361400	T	C	0.97	0.97	0.99	0.97	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs6766775	CH3	189361939	C	G	0.66	0.66	-0.97	0.66	fs4488809	ASN	EUR	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs9837544	CH3	189362620	T	C	0.95	0.95	0.98	0.95	fs10937405	fs13314271	ASN	EUR	fs10937405	ASN	fs13314271	EUR	fs10937405	ASN	fs13314271	EUR	FALSE	0NA
fs36092415	CH3	189362968	G	A	0.67	0.67	-0.98	0.67	fs4488809	ASN	EUR	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs9842008	CH3	189363001	T	C	0.85	0.85	0.97	0.85	fs10937405	ASN	EUR	EUR	fs10937405	ASN	fs13314271	EUR	fs10937405	ASN	fs13314271	EUR	FALSE	0NA
fs9827354	CH3	189363999	C	T	0.54	0.54	0.84	0.54	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs9827487	CH3	189364094	C	A	0.52	0.52	0.81	0.52	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs9827991	CH3	189364424	T	T	0.54	0.54	0.85	0.54	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs1920235	CH3	189365424	A	G	0.67	0.67	0.98	0.67	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs4687800	CH3	189365491	G	A	0.66	0.66	-0.98	0.66	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs4687881	CH3	189365570	G	A	0.66	0.66	-0.98	0.66	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs9811124	CH3	189366595	G	A	0.8	0.8	-0.98	0.8	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs9811214	CH3	189366763	C	T	0.8	0.8	-0.98	0.8	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs6807954	CH3	189366979	A	T	0.55	0.55	0.86	0.55	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs6808068	CH3	189367116	A	A	0.56	0.56	0.86	0.56	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs28612247	CH3	189367352	C	A	0.67	0.67	-0.98	0.67	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs17504410	CH3	189367551	C	A	0.67	0.67	-0.98	0.67	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs81729127	CH3	189368097	C	T	0.55	0.55	0.86	0.55	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs9821664	CH3	189368235	G	A	0.8	0.8	-0.98	0.8	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs4428661	CH3	189368337	T	A	0.8	0.8	-0.96	0.8	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs6772378	CH3	189368648	G	A	0.55	0.55	0.86	0.55	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs12330556	CH3	18936962	T	A	0.51	0.51	-0.8	0.51	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs98261717	CH3	189369390	G	A	0.51	0.51	-0.8	0.51	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs12330765	CH3	189369843	A	T	0.51	0.51	-0.8	0.51	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs9869716	CH3	189370705	A	G	0.54	0.54	0.85	0.54	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs1920221	CH3	189370780	G	A	0.5	0.5	-0.79	0.5	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs1881992	CH3	189371811	G	T	0.67	0.67	-0.97	0.67	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs9847400	CH3	189372997	G	T	0.67	0.67	-0.97	0.67	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs13082542	CH3	189373003	T	A	0.65	0.65	-0.96	0.65	fs4488809	ASN	EUR	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs1920248	CH3	189373991	T	A	1.0	1.0	1.0	1.0	fs10937405	fs13314271	ASN	EUR	fs10937405	ASN	fs13314271	EUR	fs10937405	ASN	fs13314271	EUR	FALSE	0NA
fs34279753	CH3	189376046	G	A	0.65	0.65	-0.96	0.65	fs4488809	ASN	EUR	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs43462700	CH3	189376305	A	G	1.0	1.0	1.0	1.0	fs10937405	fs13314271	ASN	EUR	fs10937405	ASN	fs13314271	EUR	fs10937405	ASN	fs13314271	EUR	FALSE	0NA
fs9835482	CH3	189376356	C	T	1.0	1.0	1.0	1.0	fs10937405	fs13314271	ASN	EUR	fs10937405	ASN	fs13314271	EUR	fs10937405	ASN	fs13314271	EUR	FALSE	0NA
fs13095135	CH3	189377530	A	C	0.65	0.65	-0.96	0.65	fs4488809	ASN	EUR	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs6809573	CH3	189378218	A	G	0.64	0.64	-0.95	0.64	fs4488809	ASN	EUR	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs1920272	CH3	189378556	A	G	0.94	0.94	0.98	0.94	fs4488809	fs13314271	ASN	EUR	fs4488809	ASN	fs13314271	EUR	fs4488809	ASN	fs13314271	EUR	FALSE	0NA
fs13321778	CH3	189379290	C	G	0.57	0.57	-0.97	0.57	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs1920273	CH3	189379493	G	A	0.67	0.67	-0.97	0.67	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs1920274	CH3	189379524	A	G	0.57	0.57	-0.97	0.57	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs7319159	CH3	189380470	C	A	0.67	0.67	-0.97	0.67	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs7319159	CH3	189380483	G	A	0.67	0.67	-0.97	0.67	fs13314271	EUR	EUR	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	fs13314271	EUR	FALSE	0NA
fs7319159	CH3	189380563	G	A	0.67	0.67	-0.97	0.67														

f51389060	ch3	18939010	AAAAG	A	0.55,0.66,0.65	0.96,0.94,0.93	f510937405,f54488809,f513314271,ASN,ASN,EUR	f510937405,ASN,f54488809,ASN,f513314271	FALSE	0NA
f51389060	ch3	18939070	AAAAG	A	1.03	1.097	f510937405,f513314271,ASN,ASN,EUR	f510937405,ASN,f513314271,EUR	FALSE	0NA
f51389060	ch3	18939029	A	G	0.64	-0.95	f54488809,ASN	f54488809,ASN	FALSE	0NA
f510054203	ch5	127964	C	C	0.58	-0.78	f52736100,ASN	f52736100,ASN	FALSE	0NA
f510054203	ch5	127964	C	C	0.62	-0.82	f52736100,ASN	f52736100,ASN	FALSE	0NA
f54975338	ch5	128018	C	C	0.65	-0.88	f52736100,ASN	f52736100,ASN	FALSE	0NA
f54002450	ch5	1280940	GAGCCAC	G	0.62	-0.89	f52736100,ASN	f52736100,ASN	FALSE	0NA
f52726159	ch5	1282319	C	C	0.81	-0.97	f52736100,ASN	f52736100,ASN	FALSE	0NA
f52725218	ch5	1282414	A	A	0.81,0.5	-0.97,-0.95	f52736100,ASN,EUR	f52736100,ASN,f52736100,EUR	FALSE	0NA
f52713218	ch5	1283112	A	G	0.52,0.91,0.5	0.81,0.96,0.71	f52853677,f52736100,ASN,ASN,EUR	f52853677,ASN,f52736100,ASN,f52736100,EUR	FALSE	0NA
f57717443	ch5	1283486	T	C	0.78	0.9	f52736100,ASN	f52736100,ASN	FALSE	0NA
f54449383	ch5	1284135	C	T	0.79	0.93	f52736100,ASN	f52736100,ASN	FALSE	0NA
f51086468	ch5	1285162	C	T	0.8	0.99	f52736100,ASN	f52736100,ASN	FALSE	0NA
f57705246	ch5	1285974	C	A	0.82	-0.95	f52736100,ASN	f52736100,ASN	FALSE	0NA
f52736100	ch5	1286516	C	A	0.53,1.1	0.83,1.1	f52853677,f52736100,ASN,ASN,EUR	f52853677,ASN,f52736100,ASN,f52736100,EUR	TRUE	1CTCF, D6 2013 3 MB1B2B3 fdrLd2 MACS2 peaks
f52853677	ch5	1287194	G	A	1.053	1.083	f52853677,ASN,ASN	f52853677,ASN,f52736100,ASN	FALSE	0NA
f52736099	ch5	1287340	A	G	0.66,0.52	0.91,0.74	f52853677,f52736100,ASN,ASN	f52853677,ASN,f52736100,ASN	FALSE	0NA
f52736102	ch5	1302144	T	C	0.72	-0.85	f5465498,ASN	f5465498,ASN	FALSE	0NA
f52853666	ch5	1302914	G	A	0.7	-0.85	f5465498,ASN	f5465498,ASN	FALSE	0NA
f52735945	ch5	1303901	T	C	0.79	-0.9	f5465498,ASN	f5465498,ASN	FALSE	0NA
f54404271	ch5	1306165	T	C	0.81	-0.91	f5465498,ASN	f5465498,ASN	FALSE	0NA
f54530805	ch5	1306331	T	C	0.79	-0.9	f5465498,ASN	f5465498,ASN	FALSE	0NA
f52746461	ch5	1306521	G	C	0.53	0.86	f5465498,ASN	f5465498,ASN	FALSE	0NA
f51133727	ch5	1306765	C	G	0.73	-0.9	f5465498,ASN	f5465498,ASN	FALSE	0NA
f520170753	ch5	1307434	T	TTA	0.78	-0.87	f5465498,ASN	f5465498,ASN	FALSE	0NA
f54653969	ch5	1308552	G	A	0.56	0.9	f5465498,ASN	f5465498,ASN	FALSE	0NA
f561574973	ch5	1309168	T	A	0.83	-0.92	f5465498,ASN	f5465498,ASN	FALSE	0NA
f56062800	ch5	1309904	G	A	0.83	-0.92	f5465498,ASN	f5465498,ASN	FALSE	0NA
f56554728	ch5	1310152	G	A	0.83	-0.92	f5465498,ASN	f5465498,ASN	FALSE	0NA
f5686294	ch5	1311693	T	C	0.84	-0.92	f5465498,ASN	f5465498,ASN	FALSE	0NA
f56866783	ch5	1312020	T	C	0.84	-0.92	f5465498,ASN	f5465498,ASN	FALSE	0NA
f53595391	ch5	1312329	C	T	0.58	0.91	f5465498,ASN	f5465498,ASN	FALSE	0NA
f51335677	ch5	1312457	G	A	0.84	-0.92	f5465498,ASN	f5465498,ASN	FALSE	0NA
f51335267	ch5	1312935	T	C	0.83	-0.91	f5465498,ASN	f5465498,ASN	FALSE	0NA
f56113365	ch5	1313242	G	C	0.58	0.91	f5465498,ASN	f5465498,ASN	FALSE	0NA
f528372921	ch5	1313201	A	G	0.84	-0.92	f5465498,ASN	f5465498,ASN	FALSE	0NA
f51007807	ch5	1314009	C	T	0.84	-0.92	f5465498,ASN	f5465498,ASN	FALSE	0NA
f54975615	ch5	1315343	G	A	0.84	-0.92	f5465498,ASN	f5465498,ASN	TRUE	3 island,H3K4me1, D0 2013 3, B1-W200-G200-FDR1e-04
f54975616	ch5	1315660	G	A	0.84	-0.92	f5465498,ASN	f5465498,ASN	TRUE	4 island,H3K4me1, D6 2013 3, MB1B2B3-W200-G200-FDR1e-04-island
f53170453	ch5	1317481	A	G	0.6	0.95	f5465498,ASN	f5465498,ASN	TRUE	5 island,H3K4me1, D4 2013 3, B1-W200-G200-FDR1e-04-island
f53816659	ch5	1317820	A	G	0.88	-0.95	f5465498,ASN	f5465498,ASN	TRUE	H3K27ac, D0C 2013 3, B1-W100-G200-FDR1e-04
f5451360	ch5	1319680	C	A	0.64	0.96	f5465498,ASN	f5465498,ASN	FALSE	island,H3K4me1, D0 2013 3, B1-W100-G200-FDR1e-04
f5421629	ch5	1320136	G	A	0.99	1	f5465498,ASN	f5465498,ASN	FALSE	island,H3K4me1, D0 2013 3, MB1B2B3-W200-G200-FDR1e-04
f5380296	ch5	1320247	G	A	0.8	0.92	f5465498,ASN	f5465498,ASN	FALSE	island,H3K4me1, D4 2013 3, B1-W200-G200-FDR1e-04
f5381949	ch5	1322468	G	A	0.92	0.97	f5465498,ASN	f5465498,ASN	FALSE	island,H3K4me1, D6 2013 3, MB1B2B3-W200-G200-FDR1e-04-island
f513178866	ch5	1323212	C	T	0.95	1	f5465498,ASN	f5465498,ASN	FALSE	H3K27ac, D4 2013 3, B1-W100-G200-FDR1e-04
f5414965	ch5	1324121	G	A	0.95	1	f5465498,ASN	f5465498,ASN	FALSE	island,H3K4me1, D0 2013 3, B1-W100-G200-FDR1e-04
f5421284	ch5	1325590	T	C	1	1	f5465498,ASN	f5465498,ASN	TRUE	island,H3K4me1, D0 2013 3, MB1B2B3-W200-G200-FDR1e-04
f54656302	ch5	1325797	A	G	1	1	f5465498,ASN	f5465498,ASN	TRUE	island,H3K4me1, D4 2013 3, B1-W200-G200-FDR1e-04-island
f5465498	ch5	1325803	A	G	1	1	f5465498,ASN	f5465498,ASN	TRUE	H3K27ac, D4 2013 3, B1-W100-G200-FDR1e-04
f51133729	ch5	1327101	C	G	0.95	1	f5465498,ASN	f5465498,ASN	TRUE	2 island,H3K4me1, D4 2013 3, B1-W200-G200-FDR1e-04-island
f5383009	ch5	1327851	C	T	0.89	1	f5465498,ASN	f5465498,ASN	TRUE	1 H3K27ac, D4 2013 3, B1-W100-G200-FDR1e-04-island
f5380145	ch5	1328897	C	T	0.62	0.95	f5465498,ASN	f5465498,ASN	TRUE	H3K27ac, D4 2013 3, B1-W100-G200-FDR1e-04
f5452932	ch5	1330253	T	C	0.92	0.99	f5465498,ASN	f5465498,ASN	TRUE	5 island,H3K4me1, D6 2013 3, MB1B2B3-W200-G200-FDR1e-04-island

16615260	C	A	0.67	0.84	ASN	16615260	C	A	0.67	0.84	ASN	16615260	C	A	0.67	0.84	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR01-Island
14662329	C	T	0.71	0.85	ASN	14662329	C	T	0.71	0.85	ASN	14662329	C	T	0.71	0.85	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14662981	T	A	0.71	0.85	ASN	14662981	T	A	0.71	0.85	ASN	14662981	T	A	0.71	0.85	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14662907	C	A	0.71	0.85	ASN	14662907	C	A	0.71	0.85	ASN	14662907	C	A	0.71	0.85	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14662759	T	C	0.77	0.88	ASN	14662759	T	C	0.77	0.88	ASN	14662759	T	C	0.77	0.88	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14663032	T	C	0.77	0.88	ASN	14663032	T	C	0.77	0.88	ASN	14663032	T	C	0.77	0.88	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14663074	C	A	0.8	0.91	ASN	14663074	C	A	0.8	0.91	ASN	14663074	C	A	0.8	0.91	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14663294	C	T	0.8	0.91	ASN	14663294	C	T	0.8	0.91	ASN	14663294	C	T	0.8	0.91	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14663468	T	C	0.85	0.92	ASN	14663468	T	C	0.85	0.92	ASN	14663468	T	C	0.85	0.92	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14663568	T	C	0.85	0.92	ASN	14663568	T	C	0.85	0.92	ASN	14663568	T	C	0.85	0.92	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14663761	A	G	0.85	0.92	ASN	14663761	A	G	0.85	0.92	ASN	14663761	A	G	0.85	0.92	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14664165	A	G	0.87	0.95	ASN	14664165	A	G	0.87	0.95	ASN	14664165	A	G	0.87	0.95	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14664247	C	T	1	1	ASN	14664247	C	T	1	1	ASN	14664247	C	T	1	1	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14664289	G	A	0.64	0.88	ASN	14664289	G	A	0.64	0.88	ASN	14664289	G	A	0.64	0.88	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14664415	C	A	1	1	ASN	14664415	C	A	1	1	ASN	14664415	C	A	1	1	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
14664672	A	G	0.59	0.98	ASN	14664672	A	G	0.59	0.98	ASN	14664672	A	G	0.59	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32354428	A	G	0.59	0.98	ASN	32354428	A	G	0.59	0.98	ASN	32354428	A	G	0.59	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32354605	G	A	0.66	0.96	ASN	32354605	G	A	0.66	0.96	ASN	32354605	G	A	0.66	0.96	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32355683	A	G	0.61	0.91	ASN	32355683	A	G	0.61	0.91	ASN	32355683	A	G	0.61	0.91	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32358942	T	A	0.64	0.92	ASN	32358942	T	A	0.64	0.92	ASN	32358942	T	A	0.64	0.92	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32361111	C	A	0.66	0.98	ASN	32361111	C	A	0.66	0.98	ASN	32361111	C	A	0.66	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32365527	C	T	0.66	0.98	ASN	32365527	C	T	0.66	0.98	ASN	32365527	C	T	0.66	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32368161	T	C	0.57	0.98	ASN	32368161	T	C	0.57	0.98	ASN	32368161	T	C	0.57	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32369955	T	C	0.66	0.98	ASN	32369955	T	C	0.66	0.98	ASN	32369955	T	C	0.66	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32367947	T	C	0.99	1	ASN	32367947	T	C	0.99	1	ASN	32367947	T	C	0.99	1	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32368097	T	C	1	1	ASN	32368097	T	C	1	1	ASN	32368097	T	C	1	1	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32428062	C	T	0.91	0.97	ASN	32428062	C	T	0.91	0.97	ASN	32428062	C	T	0.91	0.97	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32428079	C	A	0.91	0.97	ASN	32428079	C	A	0.91	0.97	ASN	32428079	C	A	0.91	0.97	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32428097	T	TG	0.5	0.91	ASN	32428097	T	TG	0.5	0.91	ASN	32428097	T	TG	0.5	0.91	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32428115	G	A	0.92	0.97	ASN	32428115	G	A	0.92	0.97	ASN	32428115	G	A	0.92	0.97	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32428715	G	A	0.92	0.97	ASN	32428715	G	A	0.92	0.97	ASN	32428715	G	A	0.92	0.97	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32428772	A	G	0.79	0.98	ASN	32428772	A	G	0.79	0.98	ASN	32428772	A	G	0.79	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32429245	C	G	0.86	0.98	ASN	32429245	C	G	0.86	0.98	ASN	32429245	C	G	0.86	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32429327	T	T	0.96	0.98	ASN	32429327	T	T	0.96	0.98	ASN	32429327	T	T	0.96	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32429643	T	C	0.96	0.98	ASN	32429643	T	C	0.96	0.98	ASN	32429643	T	C	0.96	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32429758	T	C	0.95	0.98	ASN	32429758	T	C	0.95	0.98	ASN	32429758	T	C	0.95	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32429825	C	T	0.96	0.98	ASN	32429825	C	T	0.96	0.98	ASN	32429825	C	T	0.96	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32430236	T	A	0.97	0.98	ASN	32430236	T	A	0.97	0.98	ASN	32430236	T	A	0.97	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32431124	G	A	0.97	0.98	ASN	32431124	G	A	0.97	0.98	ASN	32431124	G	A	0.97	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32431640	T	A	0.95	0.98	ASN	32431640	T	A	0.95	0.98	ASN	32431640	T	A	0.95	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32431785	A	G	0.55	0.98	ASN	32431785	A	G	0.55	0.98	ASN	32431785	A	G	0.55	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32431886	A	G	0.95	0.98	ASN	32431886	A	G	0.95	0.98	ASN	32431886	A	G	0.95	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32432077	G	C	0.98	0.99	ASN	32432077	G	C	0.98	0.99	ASN	32432077	G	C	0.98	0.99	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32432835	C	G	0.98	0.99	ASN	32432835	C	G	0.98	0.99	ASN	32432835	C	G	0.98	0.99	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32432835	C	T	1	1	ASN	32432835	C	T	1	1	ASN	32432835	C	T	1	1	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32433192	C	G	1	1	ASN	32433192	C	G	1	1	ASN	32433192	C	G	1	1	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32433276	G	A	0.62	0.95	ASN	32433276	G	A	0.62	0.95	ASN	32433276	G	A	0.62	0.95	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32433654	T	A	0.54	0.96	ASN	32433654	T	A	0.54	0.96	ASN	32433654	T	A	0.54	0.96	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32433741	G	A	0.78	0.98	ASN	32433741	G	A	0.78	0.98	ASN	32433741	G	A	0.78	0.98	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32433798	C	T	0.78	0.99	ASN	32433798	C	T	0.78	0.99	ASN	32433798	C	T	0.78	0.99	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32433907	G	A	0.73	0.99	ASN	32433907	G	A	0.73	0.99	ASN	32433907	G	A	0.73	0.99	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04
32434011	G	A	0.97	1	ASN	32434011	G	A	0.97	1	ASN	32434011	G	A	0.97	1	ASN	FAIRE_D6_2013_3_MB1B2-W200-G200-FDR1e-04

f52689690	dh6	3244130	T	C	0.73	0.87	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689691	dh6	3244135	G	A	0.72	0.87	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689692	dh6	3244149	C	T	0.72	0.87	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689693	dh6	3244240	T	A	0.97	1	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689694	dh6	3244349	C	T	1	1	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689695	dh6	3244361	A	A	1	1	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689696	dh6	3244380	C	C	1	1	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689697	dh6	3244385	G	T	0.71	0.95	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689698	dh6	3244352	A	G	0.74	1	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689699	dh6	3243556	A	G	0.67	1	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689700	dh6	3243564	T	C	0.64	0.85	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689701	dh6	3243569	C	T	0.62	0.84	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689702	dh6	3243620	A	T	0.93	0.98	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689703	dh6	3243624	A	T	0.93	0.98	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689704	dh6	3243705	A	G	0.86	0.97	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689705	dh6	3243716	T	A	0.78	0.9	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689706	dh6	3243764	G	T	0.9	0.98	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689707	dh6	3243815	C	T	0.91	0.98	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689708	dh6	3243829	A	G	0.89	0.97	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689709	dh6	3243854	T	C	0.95	0.98	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689710	dh6	3243864	C	T	0.95	0.98	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689711	dh6	3243867	G	T	0.92	0.96	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689712	dh6	3243902	T	C	0.92	0.98	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689713	dh6	3243910	A	C	0.9	0.98	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689714	dh6	3243939	A	C	0.82	0.97	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689715	dh6	3243982	G	A	0.88	0.98	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689716	dh6	3244100	C	T	0.87	0.94	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689717	dh6	3244150	C	A,G,T	0.86	0.93	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689718	dh6	3244126	C	T	0.76	0.92	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689719	dh6	3244286	T	G	0.63	0.86	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689720	dh6	3244312	C	T	0.61	0.85	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689721	dh6	3244345	I	A	0.61	0.86	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689722	dh6	3244365	T	C	0.5	0.7	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689723	dh6	3244366	G	C	0.51	0.72	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689724	dh6	3244407	G	A	0.57	0.84	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689725	dh6	3244409	C	A	0.57	0.84	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689726	dh6	3244406	C	T	0.57	0.78	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689727	dh6	3244418	G	T	0.62	0.87	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689728	dh6	3244430	A	C	0.61	0.85	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689729	dh6	3244434	C	T	0.55	0.85	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689730	dh6	3244470	T	A,G	0.51	0.72	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689731	dh6	3244731	T	G	0.6	0.84	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689732	dh6	3244509	C	T	0.59	0.83	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689733	dh6	3244514	A	G	0.56	0.8	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689734	dh6	3244517	C	T	0.56	0.8	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689735	dh6	3244530	G	A	0.56	0.79	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689736	dh6	3244560	G	A	0.73	0.86	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689737	dh6	3244591	C	A	0.59	0.86	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689738	dh6	3244592	G	T	0.6	0.86	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689739	dh6	3244601	T	C	0.55	0.85	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689740	dh6	3244605	A	C	0.6	0.85	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689741	dh6	3244671	C	G	0.6	0.85	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689742	dh6	3244676	C	T	0.78	0.78	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689743	dh6	3244630	T	C	0.5	0.76	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689744	dh6	3244642	G,A,T	C	0.55	0.76	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689745	dh6	3244645	G,A,T	A,C,T	0.57	0.79	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689746	dh6	3244658	G	A	0.71	0.84	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689747	dh6	3244658	G	A	0.71	0.84	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689748	dh6	3244674	A	T	0.5	0.71	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689749	dh6	3244683	A	T	0.55	0.81	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689750	dh6	3244692	T	C	0.58	0.78	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689751	dh6	3244694	G	A,T	0.5	0.81	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689752	dh6	3244704	G	A	0.56	0.82	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689753	dh6	3244711	A	G	0.57	0.83	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689754	dh6	3244716	A	A	0.55	0.81	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689755	dh6	3244721	G	T	0.57	0.83	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689756	dh6	3244719	G	A	0.57	0.83	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689757	dh6	3244731	G	A	0.57	0.83	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689758	dh6	3244734	G	A	0.57	0.83	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689759	dh6	3244751	T	A	0.5	0.72	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689760	dh6	3244771	A	G	0.52	0.76	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689761	dh6	3244783	A	G	0.55	0.81	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689762	dh6	3244790	T	C	0.54	0.79	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689763	dh6	3244913	C	C	0.54	0.79	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689764	dh6	3244916	G	T	0.5	0.73	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689765	dh6	3244916	G	T	0.5	0.73	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA
f52689766	dh6	3244918	G	A	0.5	0.81	f52395185	ASN	f52395185	ASN	f52395185	ASN	FALSE	ONNA

CS1853148	ch6	117750200	A	G	0.83	0.98		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS9374658	ch6	117741495	T	G	0.73	0.94		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS4945584	ch6	117750980	T	C	0.83	0.98		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS9372480	ch6	117760579	C	T	0.86	0.98		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS4946254	ch6	117765313	G	A	0.87	0.97		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS6941337	ch6	117768448	A	G	0.89	0.98		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS9374662	ch6	11777804	G	A	0.99	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS1392511	ch6	11778381	C	T	0.51	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04
CS9468919	ch6	117779546	G	ASN	0.51	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04
CS2180811	ch6	117780158	T	A	0.99	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS9481219	ch6	117780894	C	T	0.51	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS4946256	ch6	117781276	T	C	0.99	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS6941824	ch6	117782356	T	C	0.51	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04
CS9374663	ch6	11778634	A	G	0.99	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04
CS6930292	ch6	117783770	G	T	0.54	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS6910338	ch6	117783991	T	C	0.51	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS717970	ch6	117784418	A	C	0.54	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS717969	ch6	117784567	G	T	0.54	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS6937083	ch6	117785308	A	T	1	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS6942067	ch6	117785696	A	G	0.54	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS9387478	ch6	117786180	C	A	1	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS7193294	ch6	117786542	C	T	0.51	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	TRUE	3	B1-W100-G200-FDR1e-04
CS1407182	ch6	11778124	T	T	0.54	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04
CS1407181	ch6	11778151	T	G	0.54	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04
CS7149229	ch6	117787897	T	T	0.51	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04
CS13212002	ch6	117788146	A	G	0.54	1		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04
CS34882116	ch6	117789497	G	C	0.99	0.99		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04
CS1853147	ch6	117789567	A	C	0.53	0.99		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04
CS1853146	ch6	117789811	T	A	0.51	0.96		CS9387478	ASN	CS9387478	ASN	CS9387478	ASN	FALSE	3	B1-W100-G200-FDR1e-04

fs5676755	ch15	78989932	C	G	1.082	1.0.93	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs5607233	ch15	78989003	C	A	0.890.73	1.0.93	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs141518190	ch15	78900647	A	G	0.820.68	1.0.93	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs147829554	ch15	78900650	C	T	0.820.68	1.0.93	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs138544659	ch15	78900701	T	G	0.790.65	1.0.93	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs147144681	ch15	78900908	C	T	0.860.71	1.0.94	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs114205691	ch15	78901113	C	T	1.0.82	1.0.93	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs146009840	ch15	78906177	A	T	0.990.83	1.0.94	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs8040868	ch15	78911811	T	C	0.810.65	1.0.87	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	TRUE	1 CTCE DA 2013.3 MBI1B2B3 (ftr)ez MACSZ breaks
fs4243684	ch15	78911672	G	C	0.980.08	0.990.91	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs149959208	ch15	78912710	T	C	0.970.04	0.970.04	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs171448806	ch15	78913353	CGCGG6G6G	C	0.820.05	0.890.08	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs8039449	ch15	78914534	C	T	0.830.76	0.980.91	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs10851907	ch15	78915864	G	A	0.720.64	0.980.09	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs55958997	ch15	78915872	C	A	0.820.75	0.960.09	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs2869548	ch15	78926388	G	A	0.820.73	0.950.88	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs17482723	ch15	78929897	C	T	0.830.75	0.950.88	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs27243158	ch15	78929445	T	C	0.750.67	0.910.84	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs67426328	ch15	78934318	C	G	0.590.54	0.820.06	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs55988292	ch15	78936168	A	G	0.60.95	0.830.077	fs1051730/fs8034191	EUR	EUR	fs1051730 EUR/88034191	EUR	FALSE	0NA
fs3936134	ch17	65814382	C	G	0.54	0.82	fs2716064	ASN	ASN	fs2716064 ASN	ASN	FALSE	0NA
fs9295980	ch17	65815463	T	C	0.54	0.82	fs2716064	ASN	ASN	fs2716064 ASN	ASN	FALSE	0NA
fs27631329	ch17	65818374	T	C	0.54	0.82	fs2716064	ASN	ASN	fs2716064 ASN	ASN	FALSE	0NA
fs27109001	ch17	65823573	CGG	C	0.71	0.87	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	1 HK3Zac DA 2013.3 MBI1B2B3 W200-G200-FDR1-04-Island
fs139786693	ch17	65825228	TTTG	T	0.75	0.9	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	6 Island HK4knel D6 2013.3 MBI1B2B3 W200-G200-FDR1e-04-Island
fs1260389	ch17	65825248	T	C	0.8	0.91	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	6 Island HK4knel D6 2013.3 MBI1B2B3 W200-G200-FDR1e-04-Island
fs12601921	ch17	65825354	C	A,G,T	0.72	0.92	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	6 Island HK4knel D6 2013.3 MBI1B2B3 W200-G200-FDR1e-04-Island
fs12601919	ch17	65825374	A	G	0.73	0.92	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	6 Island HK4knel D6 2013.3 MBI1B2B3 W200-G200-FDR1e-04-Island
fs9915591	ch17	65826090	C	G	0.81	0.92	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	6 Island HK4knel D6 2013.3 MBI1B2B3 W200-G200-FDR1e-04-Island
fs12602556	ch17	65828661	A	G	0.81	0.92	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	7 Island HK4knel D6 2013.3 MBI1B2B3 W200-G200-FDR1e-04-Island
fs62084208	ch17	65827443	C	T	0.81	0.92	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	4 Island HK4knel D6 2013.3 MBI1B2B3 W200-G200-FDR1e-04-Island
fs202105424	ch17	65828368	AT	A	0.74	0.92	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	3 Island HK4knel D6 2013.3 MBI1B2B3 W200-G200-FDR1e-04-Island
fs2080090	ch17	65828371	T	A	0.62	0.87	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	3 Island HK4knel D6 2013.3 MBI1B2B3 W200-G200-FDR1e-04-Island
fs62084210	ch17	65828710	A	G	0.81	0.92	fs2716064	ASN	ASN	fs2716064 ASN	ASN	TRUE	2 Island HK4knel D6 2013.3 MBI1B2B3 W200-G200-FDR1e-04-Island

f511867537	ch17	65875492 A	G	0.81	0.94	f57216064	ASN	f57216064 ASN	FALSE	0NA
f511867618	ch17	65875587 G	A	0.83	0.94	f57216064	ASN	f57216064 ASN	FALSE	0NA
f535242555	ch17	65876427 G	A	0.85	0.94	f57216064	ASN	f57216064 ASN	TRUE	1H3kmetl DA 2013 3 B1-W200-G200-FDR1e-04-Island
f538122852	ch17	65876915 T	C	0.85	0.94	f57216064	ASN	f57216064 ASN	TRUE	1H3kmetl DA 2013 3 B1-W200-G200-FDR1e-04-Island
f520189733	ch17	65877610 T	IG	0.67	0.92	f57216064	ASN	f57216064 ASN	TRUE	1H3kmetl DA 2013 3 B1-W200-G200-FDR1e-04-Island
f58074644	ch17	65878217 G	C	0.87	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f562084251	ch17	65878736 G	A	0.87	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f59892803	ch17	65879243 T	G	0.87	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f5146091276	ch17	65879572 T	A	0.87	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f562084252	ch17	65879578 G	C	0.87	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f59897982	ch17	65879592 T	T	0.85	0.93	f57216064	ASN	f57216064 ASN	FALSE	0NA
f5201512564	ch17	65880256 T	TTC	0.72	0.96	f57216064	ASN	f57216064 ASN	FALSE	0NA
f5199697046	ch17	65880257 T	TC	0.73	0.96	f57216064	ASN	f57216064 ASN	FALSE	0NA
f56504545	ch17	65880367 A	G	0.75	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f56504546	ch17	65880385 A	C	0.77	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f52713292	ch17	65880443 A	G	0.87	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f53935969	ch17	65881959 A	G	0.87	0.95	f57216064	ASN	f57216064 ASN	TRUE	1H3kmetl DA 2013 3 MB1B2B3-W200-G200-FDR1e-04-Island
f54318247	ch17	65882172 A	G	0.87	0.95	f57216064	ASN	f57216064 ASN	TRUE	FAIR_De 2013 3 MB1B2-W50-G50-FDR01-2IslandH3kmetl DA 2013 3 MB1B2-W50-G50-FDR01-2IslandH3kmetl DA 2013 3 MB1B2B3-W200-G200-FDR1e-04-Island
f57386073	ch17	65882542 A	AT	0.87	0.95	f57216064	ASN	f57216064 ASN	TRUE	FAIR_De 2013 3 MB1B2-W50-G50-FDR01-2IslandH3kmetl DA 2013 3 MB1B2B3-W200-G200-FDR1e-04-Island
f58082661	ch17	65884701 G	A	0.87	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f52708663	ch17	65885121 T	C	0.87	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f57523441	ch17	65885485 A	G	0.87	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f512452823	ch17	65885783 G	A	0.87	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f511079710	ch17	65885911 C	T	0.73	0.91	f57216064	ASN	f57216064 ASN	FALSE	0NA
f56504548	ch17	65886048 T	C	0.85	0.93	f57216064	ASN	f57216064 ASN	FALSE	0NA
f58079291	ch17	65887622 A	G	0.87	0.95	f57216064	ASN	f57216064 ASN	TRUE	H3kmetl DA 2013 3 MB1B2B3-W200-G200-FDR1e-04-IslandH3kmetl DA 2013 3 B1-W200-G200-FDR1e-04-Island
f58079754	ch17	65887896 A	G	0.87	0.95	f57216064	ASN	f57216064 ASN	TRUE	CTCF DA 2013 3 MB1B2B3-FDR1e_MACS2_peaks,CTCF_De 2013_3_MBI B2B3-FDR1e2_MACS2_peaks,H3kmetl DO 2013 3 MB1B2B3-W200-G200-FDR1e-04-Island
f56504549	ch17	65888784 A	G	0.85	0.93	f57216064	ASN	f57216064 ASN	TRUE	H3kmetl DO 2013 3 MB1B2B3-W200-G200-FDR1e-04-Island
f511653507	ch17	65892064 T	A	0.5	-0.96	f57216064	ASN	f57216064 ASN	FALSE	0NA
f559646213	ch17	65892280 G	G	0.59	0.81	f57216064	ASN	f57216064 ASN	FALSE	0NA
f560858912	ch17	65892343 G	G	0.79	0.96	f57216064	ASN	f57216064 ASN	FALSE	0NA
f561676547	ch17	65892507 G	T	0.71	0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA
f561483800	ch17	65892752 T	C	0.98	0.98	f57216064	ASN	f57216064 ASN	TRUE	1H3kmetl DA 2013 3 B1-W200-G200-FDR1e-04-Island
f53206817	ch17	65892834 A	G	1	1	f57216064	ASN	f57216064 ASN	TRUE	1H3kmetl DA 2013 3 B1-W200-G200-FDR1e-04-Island
f556404791	ch17	65893019 T	G	0.98	1	f57216064	ASN	f57216064 ASN	TRUE	1H3kmetl DA 2013 3 B1-W200-G200-FDR1e-04-Island
f5145606454	ch17	65893289 AC	A	0.99	1	f57216064	ASN	f57216064 ASN	TRUE	1H3kmetl DA 2013 3 B1-W200-G200-FDR1e-04-Island
f56504550	ch17	65894542 A	G	0.83	1	f57216064	ASN	f57216064 ASN	FALSE	0NA
f57216064	ch17	65898809 G	A	1	1	f57216064	ASN	f57216064 ASN	FALSE	0NA
f58063006	ch17	65902645 G	A	0.93	0.99	f57216064	ASN	f57216064 ASN	FALSE	0NA
f56504551	ch17	65903326 T	G	0.83	1	f57216064	ASN	f57216064 ASN	FALSE	0NA
f5202228218	ch17	65906928 AT	A	0.84	0.98	f57216064	ASN	f57216064 ASN	FALSE	0NA
f512602655	ch17	65909442 A	A	0.9	0.98	f57216064	ASN	f57216064 ASN	FALSE	0NA
f58072225	ch17	65912960 A	G	0.94	0.98	f57216064	ASN	f57216064 ASN	FALSE	0NA
f555896768	ch17	65913358 T	C	0.94	0.98	f57216064	ASN	f57216064 ASN	FALSE	0NA
f512602941	ch17	65913887 A	G	0.9	0.98	f57216064	ASN	f57216064 ASN	FALSE	0NA
f562085991	ch17	65920508 G	A	0.89	0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA
f54790981	ch17	65921834 A	G	0.92	0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA
f562085992	ch17	65923083 T	A	0.89	0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA
f562085993	ch17	65923152 T	A	0.95	0.98	f57216064	ASN	f57216064 ASN	FALSE	0NA
f57338706	ch17	65926604 T	G	0.92	0.96	f57216064	ASN	f57216064 ASN	FALSE	0NA
f54271626	ch17	65930133 G	A	0.95	0.98	f57216064	ASN	f57216064 ASN	FALSE	0NA
f54791300	ch17	65930772 A	G	0.94	0.98	f57216064	ASN	f57216064 ASN	TRUE	1H3kmetl DO 2013 3 MB1B2B3-W200-G200-FDR1e-04-Island
f54791299	ch17	65930937 A	G	0.87	0.95	f57216064	ASN	f57216064 ASN	TRUE	1H3kmetl DO 2013 3 MB1B2B3-W200-G200-FDR1e-04-Island
f556393562	ch17	65933918 C	T	0.95	0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA
f58521447	ch17	65936179 TATTAA	T	0.93	0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA
f58071463	ch17	65938655 G	A	0.92	-0.96	f57216064	ASN	f57216064 ASN	FALSE	0NA
f51245121	ch17	65939075 G	A	0.86	0.95	f57216064	ASN	f57216064 ASN	TRUE	CTCF DA 2013 3 MB1B2B3-FDR1e2_MACS2_peaks,CTCF_De 2013_3_MBI B2B3-FDR1e2_MACS2_peaks
f56504555	ch17	65940727 A	G	0.88	-0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA
f51187008	ch17	65944515 A	C	0.91	0.96	f57216064	ASN	f57216064 ASN	FALSE	0NA
f57272357	ch17	65945536 C	G	0.92	0.96	f57216064	ASN	f57216064 ASN	FALSE	0NA
f5723813	ch17	65945840 A	A	0.9	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f512449442	ch17	65947640 G	A	0.88	0.95	f57216064	ASN	f57216064 ASN	FALSE	0NA
f511867401	ch17	65947970 G	A	0.91	-0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA
f59907491	ch17	65949933 G	A	0.91	-0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA
f57144214	ch17	65950230 G	G	0.91	-0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA
f54790973	ch17	65951107 A	T	0.91	-0.97	f57216064	ASN	f57216064 ASN	FALSE	0NA

rs6208462	ch17	66031521 G	C	0.54	0.76	rs7216064	ASN	rs7216064 ASN	TRUE	FARE_DO_2013_3_MBI1B2-W50-G50-FDR0-01- islandFARE_DA_2013_3_MBI1B2-W50-G50-FDR0-01- islandFARE_D6_2013_3_MBI1B2-W50-G50-FDR0-01- islandHK99c_D6_2013_3_MBI1B2-W200-G200-FDR-01-island
rs7214119	ch17	66032298 G	A	0.53	0.75	rs7216064	ASN	rs7216064 ASN	TRUE	H3K27ac_DO_2013_3_B1-W100-G200-FDRle-04- islandHK4me1_DO_2013_3_MBI1B2B3-W200-G200-FDRle-04- islandHK99c_DO_2013_3_MBI1B2B3-W200-G200-FDR-01- islandHK99c_DA_2013_3_MBI1B2B3-W200-G200-FDR-01- 5.islandHK99c_D6_2013_3_MBI1B2-W200-G200-FDR-01-island
rs8075764	ch17	66033109 C	A	0.54	0.76	rs7216064	ASN	rs7216064 ASN	TRUE	H3K9me1_DO_2013_3_MBI1B2B3-W200-G200-FDRle-04- islandHK4me1_DA_2013_3_B1-W100-G200-FDRle-04- 3.islandHK99c_D6_2013_3_MBI1B2-W200-G200-FDR-01-island
rs146286181	ch17	66035128 CTA	C	0.53	0.75	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs6208469	ch17	66035289 T	C	0.53	0.75	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs62084692	ch17	66035844 C	T	0.52	0.75	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs62084694	ch17	66037594 G	T	0.53	0.75	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs62084697	ch17	66038556 G	A	0.53	0.75	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs62084703	ch17	66045384 C	T	0.52	0.74	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs74354868	ch17	66045831 G	A	0.5	0.74	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs7593950	ch17	66048462 G	T	0.52	0.74	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs7593810	ch17	66049531 G	A,C	0.51	0.73	rs7216064	ASN	rs7216064 ASN	TRUE	H3K27ac_DO_2013_3_B1-W100-G200-FDRle-04- islandHK4me1_DO_2013_3_B1-W100-G200-FDRle-04- islandHK4me1_D6_2013_3_MBI1B2B3-W200-G200-FDRle-04- islandHK4me1_DO_2013_3_MBI1B2B3-W200-G200-FDRle-04- 5.islandHK4me1_D6_2013_3_MBI1B2B3-W200-G200-FDRle-04-island
rs62084708	ch17	66049707 G	A	0.51	0.73	rs7216064	ASN	rs7216064 ASN	TRUE	NA
rs3540037	ch17	66056682 A	G	0.51	0.73	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs74916757	ch17	66060906 G	T	0.5	0.73	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs62084740	ch17	66061184 C	T	0.51	0.73	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs74900486	ch17	66062873 C	T	0.51	0.73	rs7216064	ASN	rs7216064 ASN	FALSE	NA
rs7544834	ch17	66066438 T	C	0.5	0.72	rs7216064	ASN	rs7216064 ASN	FALSE	NA

Supplementary Table 2. Annotation of LUAD risk-associated SNPs

Region	Number
Intron	583
Intergenic region	335
Promoter region	16
5' or 3' UTR	7
Exon – Synonymous	4
Exon – Missense	3
Exon – Nonsense	0
Total	948

Supplementary Table 3. Functional prediction of LUAD risk-associated SNPs located in exons

rsID	IndexSNP- Population- r^2 *	Gene	SIFT score	SIFT pred	Polyphen HDIV score	Polyphen HDIV pred	Polyphen HVAR score	Polyphen HVAR pred
rs2076530	rs3817963 -ASN-0.57	BTNL2	NA	NA	0	Benign	0	Benign
rs9891146	rs7216064 -ASN-0.61	C17orf58	1	Tolerated	0	Benign	0	Benign
rs16969968	rs1051730 -EUR-0.98	CHRNA5	0.18	Tolerated	0.045	Benign	0.011	Benign

*ASN: Asian, EUR: European

Supplementary Table 4. Functional prediction of LUAD risk-associated SNPs located in UTRs

rsID	InputSNP-Population- r^2 *	Region	Gene	targetScanS
rs11656743	rs7216064-ASN-0.77	UTR3	BPTF	None
rs8192482	rs1051730-EUR-0.99	UTR3	CHRNA5	None
rs4887067	rs1051730-EUR-0.99	UTR3	CHRNA5	None
rs55853698	rs1051730-EUR-0.88	UTR5	CHRNA5	None
rs55781567	rs1051730-EUR-0.89	UTR5	CHRNA5	None
rs71448806	rs1051730-EUR-0.62	UTR5	CHRNA3	None
rs2854464	rs11610143-ASN-0.71	UTR3	ACVR1B	None

*ASN: Asian, EUR: European

Table S5. SNPs located in promoter regions

rsID	chr	Pos (hg19)	index SNP- population- r^2 *	D'	gene	Direction	Distance from TSS (basepairs)**
rs7631358	chr3	189348411	rs4488809-ASN-0.75	-0.94	TP63	5'	803
rs76937731	chr3	189348968	rs4488809-ASN-0.61	-0.92	TP63	5'	246
rs60622800	chr5	1309904	rs465498-ASN-0.83	-0.92	MIR4457	5	411
rs6554758	chr5	1310152	rs465498-ASN-0.83	-0.92	MIR4457	5	659
rs27996	chr5	1345474	rs465498-ASN-0.88	0.95	CLPTM1L	5	471
rs4946258	chr6	117803538	rs9387478-ASN-0.72	0.95	DCBLD1	5	280
rs34431655	chr6	117803676	rs9387478-ASN-0.74	0.98	DCBLD1	5	142
rs17483548	chr15	78730313	rs8034191-EUR-0.66	0.82	IREB2	5	203
rs8039449	chr15	78914534	rs1051730-EUR-0.83	0.98	CHRNA3	5	896
rs67426328	chr15	78934318	rs1051730-EUR-0.59	0.82	CHRNA4	5	730
rs7502307	chr17	65989961	rs7216064-ASN-0.65	0.81	C17orf58	5	195
rs113985803	chr17	65989971	rs7216064-ASN-0.71	-0.86	C17orf58	5	205
rs56305452	chr17	65990500	rs7216064-ASN-0.73	-0.87	C17orf58	5	734
rs56315139	chr17	65990571	rs7216064-ASN-0.59	-0.86	C17orf58	5	805
rs62084681	chr17	66030992	rs7216064-ASN-0.54	0.76	KPNA2	5	854
rs62084682	chr17	66031521	rs7216064-ASN-0.54	0.76	KPNA2	5	325

*ASN: Asian, EUR: European; ** TSS= transcription start site

Supplementary Table 6. ChIP-seq and FAIRE-seq statistics

	FAIRE			H3K27ac			H3K4me1		
	D0	D4	D6	D0	D4	D6	D0	D4	D6
# reads (Raw)	123,608,859	125,035,637	139,654,375	116,753,350	72,563,956	52,203,734	143,424,035	123,706,480	99,326,714
# reads (QC)	123,568,043	124,998,560	139,618,662	116,676,688	72,535,483	52,185,120	126,980,095	123,667,091	93,075,899
% High Quality Reads	99.97%	99.97%	99.97%	99.93%	99.96%	99.96%	88.53%	99.97%	93.71%
# reads (Aligned)	83,307,530	86,831,534	102,285,869	43,490,305	56,340,155	43,145,070	80,895,482	91,934,330	69,183,936
% Aligned Reads	67.42%	69.47%	73.26%	37.27%	77.67%	82.68%	63.71%	74.34%	74.33%
# reads (RemDup)	77,901,295	79,059,239	96,003,349	10,038,246	30,166,166	13,803,476	65,431,882	89,228,395	62,078,258
% Unique Reads	93.51%	91.05%	93.86%	23.08%	53.54%	31.99%	80.88%	97.06%	89.73%
# Peaks	80,845	89,177	95,117	39,210	44,976	43,743	87,443	76,292	75,490

Supplementary Table 7. Cloning primers for pGL4.26/ luciferase reporter gene constructs

SNP, construct length, & (hg19 genomic coordinates)	Forward	Reverse
rs452384, 1680 bp in <i>CLPTMIL</i> (chr5:1330187-1331866)	5'- AAATATGGTACCC <u>CACCCGAACACGCATTCCT</u> -3'	5'- TAATAAGCTAGCCTCACTGGAGCCTGCGGTAG-3'
rs452384, 459 bp in <i>CLPTMIL</i> (chr5:1330613-1331071)	5' – AAATATGGTACCCAGGTACTCCCCAGTCCACG-3'	5' – TAATAACTCGAGCGTGTCTGCGTTGTTGCTCC-3'
rs6942067, 836 bp in <i>DCBLD1/GOPC</i> (chr6:117785006-117785841)	5'- AAATATGGTACCCCTATGTGTCAATCACCTCCC -3'	5'- TAATAACTCGAGGACTTCAGCTTCTGGGAAGATG -3'
rs9481728, 912 bp in <i>DCBLD1</i> (chr6:117817044-117817955)	5'- AAATATGGTACCTCTTTGGACACTGATGGAAGTA-3'	5'- TAATAACTCGAGGGTTGTGCCTACTGAAGTGATTT3'

sequence that participates in PCR reaction underlined; additional sequence= restriction enzyme sites.

Lower case = base subject to mutagenesis