

**CRISPRa screening identifies lncRNAs regulating tumor response to CD8<sup>+</sup> T cell cytotoxicity**

by

**Yueshan Zhao**

BS, Sun Yat-sen University, 2020

Submitted to the Graduate Faculty of the  
School of Pharmacy in partial fulfillment  
of the requirements for the degree of  
Master of Science

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

SCHOOL OF PHARMACY

This thesis was presented

by

**Yueshan Zhao**

It was defended on

March 31, 2022

and approved by

Da Yang, Associate Professor, Department of Pharmaceutical Sciences

Robert Gibbs, Professor, Department of Pharmaceutical Sciences

Min Zhang, Research Assistant Professor, Department of Pharmaceutical Sciences

Thesis Advisor/Dissertation Director: Da Yang, Associate Professor, Department of  
Pharmaceutical Sciences

Copyright © by Yueshan Zhao

2022

# CRISPRa screening identifies lncRNAs regulating tumor response to CD8<sup>+</sup> T cell cytotoxicity

Yueshan Zhao, BS

University of Pittsburgh, 2022

Human genomes contain abundant non-coding RNAs which may play critical roles in all aspects of functions through diverse mechanisms in cancer cells, among which long non-coding RNAs (lncRNAs) have been reported to regulate cell proliferation, metastasis, and drug resistance. Recent studies have shown lncRNAs can act as biomarkers for immunotherapy efficacy in patients. However, the relationship between lncRNAs and immune response in cancer treatment is not well explored currently. In this thesis, we performed genome-wide CRISPR activation screening in melanoma cells with human CD8<sup>+</sup> T cells co-culture treatment. The screening identified top ranked genes including IL10RB-DT and LINC01198 that can result in resistance/sensitivity to T cells cytotoxicity respectively. IL10RB-DT/LINC01198 expression demonstrated significantly negative/positive correlation with immune cell infiltration and overall survival of patients with melanoma and breast cancer. Additionally, we performed RNA-seq in activated or inhibited IL10RB-DT and LINC01198 melanoma and breast cancer cell lines by single sgRNA transfection or siRNA treatment. Pathway analysis revealed that these two lncRNAs were associated with the regulation of Type I and Type II interferon related pathways in both activation and knockdown models. These results demonstrated the feasibility of *in vitro* high-throughput screening of potential lncRNAs associated with immune response and suggested novel targets for cancer immunotherapy.

## Table of Content

<b>1.0 Introduction.....</b>	<b>1</b>
<b>1.1 CRISPR technology and different types of screening.....</b>	<b>1</b>
<b>1.2 CRISPR screening data analysis.....</b>	<b>4</b>
<b>1.2.1 RNAi Gene Enrichment Ranking (RIGER) .....</b>	<b>5</b>
<b>1.2.2 Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK)</b> .....	<b>5</b>
<b>1.3 Long non-coding RNA .....</b>	<b>6</b>
<b>1.4 The function of lncRNA in cancer immune response.....</b>	<b>7</b>
<b>2.0 Genome-wide CRISPRa Screening Identifies Genes that Regulate the Response of</b> <b>CD8<sup>+</sup> T Cells to Tumor Cells.....</b>	<b>10</b>
<b>2.1 Methods and Materials .....</b>	<b>10</b>
<b>2.1.1 Cell lines and reagents .....</b>	<b>10</b>
<b>2.1.2 CRISPR library preparation, lentivirus production and transfection .....</b>	<b>10</b>
<b>2.1.3 Screening and sgRNA sequence.....</b>	<b>11</b>
<b>2.1.4 Genomic DNA extraction and sample preparation for NGS .....</b>	<b>12</b>
<b>2.1.5 Re-annotation of Human CRISPR lncRNA Activation Pooled Library .....</b>	<b>13</b>
<b>2.1.6 Quantitation of sgRNAs and CRISPR screening analysis .....</b>	<b>14</b>
<b>2.2 Results.....</b>	<b>15</b>
<b>2.2.1 Genome-wide screening for immune regulating lncRNA using an established</b> <b>CD8<sup>+</sup> T-tumor cells co-culture system.....</b>	<b>15</b>

2.2.2 The development of a combined analysis pipeline to analyze CRISPR screening data .....	17
<b>3.0 Novel hits were identified from both negative and positive selection .....</b>	<b>23</b>
<b>3.1 Methods and Materials .....</b>	<b>23</b>
<b>3.1.1 Public databases .....</b>	<b>23</b>
3.1.1.1 Genotype-Tissue Expression (GTEx) database.....	23
3.1.1.2 TCGA database.....	23
3.1.1.3 GEO datasets.....	24
<b>3.1.2 Cell line expression data .....</b>	<b>24</b>
3.1.2.1 RNA-seq analysis .....	24
3.1.2.2 Gene Set Enrichment Analysis (GSEA).....	25
<b>3.1.3 Statistical analysis .....</b>	<b>25</b>
<b>3.2 Results.....</b>	<b>26</b>
3.2.1 CRISPR screening identified lncRNA IL10RB-DT as a suppressor of tumor immune response.....	26
3.2.2 LncRNA LINC01198 is identified as an activator of tumor immune response .....	29
<b>4.0 Discussion.....</b>	<b>33</b>
<b>Appendix A Transcripts identified from negative selection .....</b>	<b>35</b>
<b>Appendix B Transcripts identified from positive selection.....</b>	<b>38</b>
<b>Bibliography .....</b>	<b>40</b>

## List of Figures

<b>Figure 1 General process for pooled CRISPR screens .....</b>	<b>3</b>
<b>Figure 2 Three vectors system used in CRISPRa screening .....</b>	<b>16</b>
<b>Figure 3 Overview of CRISPRa screening with a co-culture system.....</b>	<b>17</b>
<b>Figure 4 Quality control for CRISPR screening analysis .....</b>	<b>18</b>
<b>Figure 5 Relationships between lncRNAs and protein-coding genes.....</b>	<b>19</b>
<b>Figure 6 ULK1's repression of anti-tumor immunity .....</b>	<b>21</b>
<b>Figure 7 Top hits identified from both negative and positive selection .....</b>	<b>22</b>
<b>Figure 8 IL10RB-DT as a tumor immune suppressor from integrated clinical data.....</b>	<b>27</b>
<b>Figure 9 IL10RB-DT inhibits the killing effect of CD8<sup>+</sup> T cells through immunogenic pathways .....</b>	<b>28</b>
<b>Figure 10 LINC01198 as a positive regulator for immune response in breast cancer .....</b>	<b>30</b>
<b>Figure 11 LINC01198 positively regulate tumor immune reponse through type I and type II interferon signaling pathways.....</b>	<b>32</b>

## List of Tables

<b>Table 1 Transcripts included in the sgRNA library after reannotation .....</b>	<b>19</b>
-----------------------------------------------------------------------------------	-----------

## **1.0 Introduction**

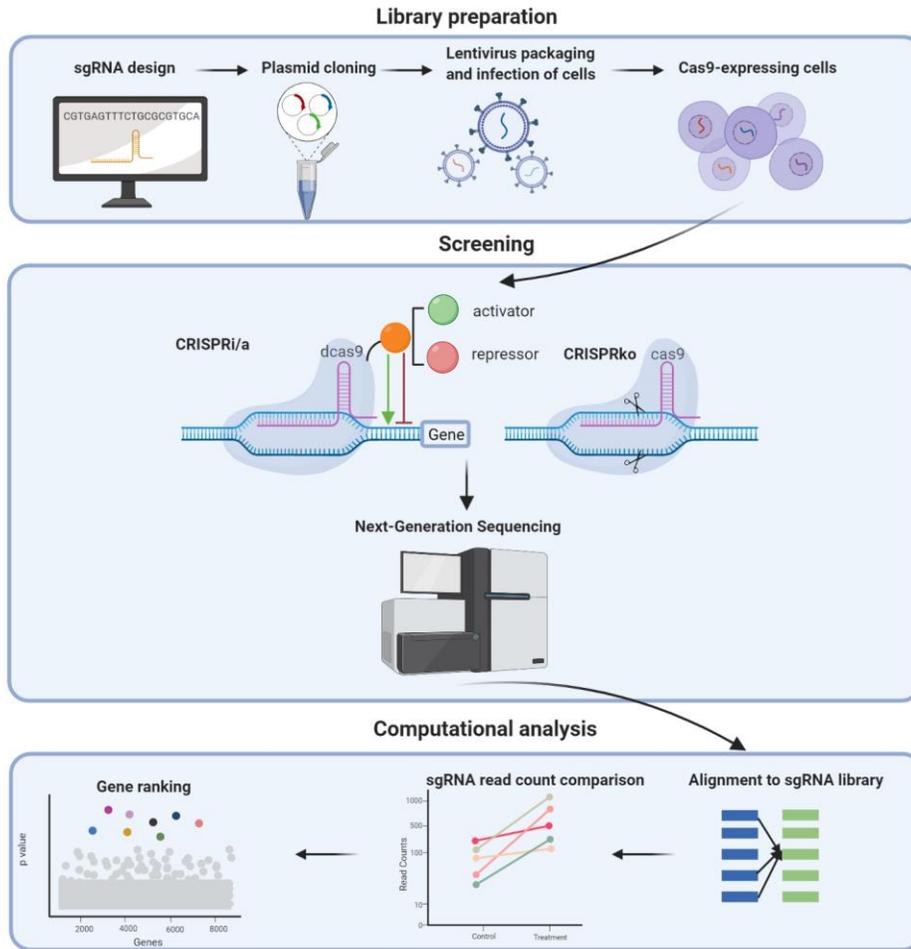
Clustered regularly interspaced palindromic repeats (CRISPR) loci with endonuclease (Cas) proteins is an immune defense system that protect bacteria from virus invasion, among which CRISPR-Cas9 is the most common one [1]. The CRISPR-Cas9 system consists of two major components: sgRNA and Cas9 nuclease. The sgRNA is made up of two parts: crRNA that is complementary to the target DNA sequences and tracrRNA serving as a binding scaffold for the Cas9 protein and guide it to the DNA region that requires editing. After directed to the target DNA locus by sgRNA, Cas9 protein will cleave the DNA and introduce a double-strand break within the sequence [2].

Winning the Nobel Prize in 2020, CRISPR technology has become an effective tool for biological research, but its application is more than gene editing. Numerous studies have used CRISPR technology to perform high throughput genome-scale screening and identified essential genes or therapeutic targets.

### **1.1 CRISPR technology and different types of screening**

The aim of genome-scale screens is to generate a population of cells with different perturbations to identify genes or regulatory regions that will play a role in specific phenotypes. Because of the wide range of potential target sequences, CRISPR system has enabled powerful pooled screens. Based on different mechanisms, CRISPR screens can be categorized into three types: CRISPR/Cas9 knockout (CRISPRko) screens, CRISPR/dCas9 activation (CRISPRa)

screens and CRISPR/dCas9 interference (CRISPRi) screens (**Figure 1**). In CRISPR knockout screens, Cas9-guided DNA double strand breaks lead to insertions or deletions when cells try to repair DNA with the non-homologous end-joining (NHEJ) pathway [3]. These mutations will result in a coding frameshift or stop codon, which ultimately silence gene expression. On the other hand, the deactivated Cas9 (dCas9) is a mutant protein which is not able to cleave DNA. Fused with transcription activators or repressors, CRISPR-dCas9 system allows regulation at gene transcription level or epigenetic level for either gene activation (CRISPRa) or interference (CRISPRi) [4]. Inhibition of gene expression can be accomplished by fusing transcription repressor such as KRAB (Kruppel-associated box) to dCas9 protein in CRISPRi screens [5]. Similarly, CRISPRa screens take advantage of transcription activator such as Synergistic Activation Mediator (SAM) system [6], which consists of four copies of VP16 from herpes simplex virus and sgRNA containing MS2-p65-HSF1 (Heat Shock Transcription Factor 1) domains. Besides protein coding genes screening, CRISPRi and CRISPRa screens can be used for functional characterization of regulatory elements [7] and long noncoding RNAs (lncRNAs) [8]. In mechanism, CRISPRko and CRISPRi facilitate loss of function while CRISPRa allows gain of function. For example, Julia Joung and her colleagues developed a genome-scale CRISPRa screening and successfully characterized that *EMICER1*'s activation led to the activation of neighboring protein-coding genes resulting in the resistance to BRAF inhibitors in melanoma cells [8]. In our case, because of the relatively low abundance of lncRNAs [9], we also decided to use CRISPRa system to achieve the activation of lncRNAs in a more 'physiological' way by designing sgRNAs targeting the promoter region of target genes to study their functions.



**Figure 1 General process for pooled CRISPR screens**

**Library preparation:** Multiple sgRNAs are designed for one target, which can be cloned into plasmids. The lentiviral library is then packaged and used to infect desired cells for CRISPR screens

**Screening:** After pooled library preparation, targets are edited by either CRISPR knockout or CRISPR interference/activation. Next-Generation sequencing is performed to collect sgRNA abundance in cell population.

**Computational analysis:** Following deep sequencing, reads are mapped to the original sgRNA library, and fold changes of sgRNA read counts are then calculated. Based on various algorithms, hit genes can be identified.

As for the phenotype, the pooled CRISPR screens were initially used to identify essential genes for cell viability [10]. Multiple researches using CRISPR screening to identify genes that will contribute to treatment sensitivity/resistance in the context of both immunotherapy [11] and

chemotherapy [12, 13]. Combining with fluorescence-activated cell sorting (FACS), the phenotype was extended to cell surface markers [14], intracellular reporters [15] or specific cell types [16]. Further utilization of single-cell RNA-seq (scRNA-seq) of the CRISPR screened samples dramatically expand the dimensions of phenotypes to the expression levels of hundreds of genes simultaneously. CRISPR screens at single cell level facilitate insights into the effect of gene perturbations on the whole transcriptome, and different methodologies such as Perturb-seq [17, 18], CRISP-seq [19] and CROP-seq [20] have been developed.

## **1.2 CRISPR screening data analysis**

As a genome wide high-throughput screening technology, whether CRISPR screens can effectively provide insights for us largely depends on the accuracy of data analysis. There have been quite some challenges for the development of CRISPR screen analysis methods. Due to the fact that multiple sgRNAs are designed for one target, we are also faced with variable sgRNA efficiency and off-target effects. The method is also expected to deal with different phenotype effects from simple cell viability to complicated transcriptome profiles. Despite the difficulty, various methods with different focus have been developed for CRISPR screen analysis. The overall workflow of those methods usually includes sequence quality assessment, read alignment, read count normalization, estimate changes of sgRNA abundance and aggregating sgRNA effects for the overall effects of targeted genes. In addition to the methods that are newly designed to analyze the CRISPR screening data. Some methods, which were previously designed to analyze the RNA interference (RNAi) screening data, can be repurposed for CRISPR screens analysis. In

order to guarantee the accuracy of our screening result, two methodologies with different mechanisms were utilized and overlapped genes from two analysis were selected as our hits.

### **1.2.1 RNAi Gene Enrichment Ranking (RIGER)**

RIGER was first developed in 2008 and used to identify essential genes in RNAi screens [21]. The core of RIGER analysis was based on Gene Set Enrichment Analysis (GSEA) which utilized a weighted Kolmogorov-Smirnov statistics to test whether a predefined set of genes skewed to the top or bottom of the whole gene list [22]. RIGER considered the entire list of sgRNAs targeting the same gene due to various efficiencies of designed sgRNAs. For CRISPR screen data, RIGER first scored sgRNAs in terms of their differential effects, such as signal-to-noise ratio, between the treatment group and control group. Raw enrichment scores were then calculated in a similar manner as for GSEA analysis. Normalization was further performed to account for different numbers of sgRNAs targeting different genes by dividing the enrichment score by the mean of a null distribution generated from random permutations. RIGER offered gene rankings computed for positive scores and negative scores separately.

### **1.2.2 Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK)**

MAGeCK [23] was the first workflow designed for CRISPR/Cas9 screen analysis and has been widely used ever since. The assumption was that if a gene is essential, then sgRNAs targeting this gene will be enriched at one side (either top or bottom) instead of being randomly distributed in the whole list of sgRNA ranking based on its effect. Read counts in different groups were first normalized for the adjustment of library sizes and count distributions. The sgRNA

abundance was over-dispersed like other high-throughput sequencing experiments [24], which means that the variance of reads is larger than the mean, thus a mean-variance model was utilized to estimate the variance. A negative binomial distribution similar to edgeR [25] package was then used to test whether there is significant difference between treatment and control groups because it works better in dealing with the over-dispersion of sequencing results. They further used  $p$ -values calculated from the negative binomial distributions to rank sgRNAs, and a robust ranking aggregation (RRA) [26] method to identify positively and negatively enriched genes simultaneously. Finally, the false discovery rate (FDR) can be calculated from permutation tests to correct for multiple comparisons. MAGeCK was also able to identify essential pathways based on the same principle.

Recently, MAGeCK have been further developed into integrated workflows, MAGeCK-VISPR [27] (Visualization for CRISPR) and MAGeCKFlute [28], which were able to provide extensive quality control at the sequence level, read count level and sample level. VISPR also provided multiple ways to explore screen results in depth including Gene Ontology (GO) enrichment analysis from Gorilla [29], gene-gene interaction network from GeneMANIA[30] and so on. The MAGeCK series is also constantly updated, and it has become the standard solution for CRISPR screen analysis.

### **1.3 Long non-coding RNA**

Long non-coding RNAs (lncRNA) is one representative group of non-coding RNAs, which are defined as transcripts longer than 200 nucleotides and not translated into functional proteins. LncRNAs can be further divided into different groups based on their biogenesis loci

including intergenic, intronic, antisense, bidirectional and enhancer lncRNAs [31]. They also are involved in numerous cellular functions such as cell differentiation, cell proliferation and activation of several types of immune cells [32, 33]. The expression level of lncRNAs is lower compared with protein-coding genes and their regulation of gene expression is often cell-type specific [9]. For example, in the context of T cell transcriptomes, 1524 intergenic lncRNAs are expressed in a lineage-specific manner at different stages for T cell maturation and differentiation [34]. Within the cell, lncRNAs can be found in multiple areas including the nucleus, cytoplasm and the mitochondria [35], and its localization will provide some hints to their functions. Functionally speaking, lncRNAs can be associated with the change of chromosome structure [36], interaction with transcription machinery [37], alternative splicing [38] and mRNA stability [39].

#### **1.4 The function of lncRNA in cancer immune response**

Tumor immune surveillance has been identified as a critical process in inhibiting tumorigenesis whereby the immune system identifies potentially cancerous cells and destroys them [40, 41], in which effector T cells play a vital role [42]. CD8<sup>+</sup> T cells are activated after T cell receptors (TCR) recognize tumor neoantigens presented on major histocompatibility complex class I (MHC-I) molecule, resulting in the killing of cancer cells [43-45]. By stimulating patients' own immune system to selectively kill cancer cells instead of attacking normal cells indiscriminately, immunotherapy has revolutionized the treatment of malignant tumors [46]. Multiple immune checkpoint inhibitors have been approved for the enhancement of CD8<sup>+</sup> T cell anti-tumor response in multiple cancer types [47-49], either as monotherapy or combination therapy [50-52]. Nevertheless, the low response rate [53] and undesired side effects [54] for some

patients called for more in-depth study of underlying mechanisms. The genomic changes in tumor tissues have been shown to be involved in tumor immune response [55, 56]. Identification of immune response biomarkers will not only expand the beneficial patients' quantity from the immunotherapy but also quality of patient treatment.

Previous studies on mechanisms leading to enhanced immune response mainly focused on protein-coding genes, yet the cause of cancer progression cannot be completely explained by the regulation of protein-coding genes [57]. It has been revealed that some non-coding RNAs may play vital roles in the development of various diseases such as cancer [58]. Recently, lncRNAs have been demonstrated to regulate the immune system in multiple ways. For instance, lncRNA ITPRIP-1 and lnc-Lsm3b are shown to be relevant to innate immune response to viral infection [59, 60]. lnc-DC is critical to dendritic cells differentiation by interacting with transcription factor STAT3 [61]. lncRNA NKILA regulates T cell sensitivity to immunological elimination by interacting with NF- $\kappa$ B [62]. Our previous analysis characterized an onco-lncRNA, EPIC1, which interacts with protein EZH2 and directly suppresses tumor cell antigen presentation and leads to anti-PD-1 treatment resistance [63]. Although these findings demonstrate the possibility of identifying novel lncRNA genes that regulate human immune response, the majority of lncRNA genes' functions and their roles in tumor immunity remain unknown.

In this study, we performed genome-scale CRISPR activation screening in the context of *ex vivo* CD8<sup>+</sup> T cell and tumor cell coculture with a library containing 96,458 sgRNAs. Our screen discovered novel lncRNAs which contribute to resistance or sensitivity to T cell cytotoxicity. Specifically, RNA-seq data in LINC01198 activation demonstrated up-regulation in type I interferon signaling pathway. Additional validation demonstrated that LINC01198 upregulated immune-related genes at both RNA and protein level after being activated. Another potential

suppressor, IL10RB-DT, downregulated immune-related pathways such as TNFA signaling and inflammatory response after being activated in melanoma cells. By further integrating predicted infiltrating lymphocytes, immune signatures and survival data from TCGA samples, we aim to identify lncRNAs that may act as biomarkers for tumor immune response in several cancer types, hoping to provide more possibilities for the development of immunotherapy.

## **2.0 Genome-wide CRISPRa Screening Identifies Genes that Regulate the Response of CD8<sup>+</sup> T Cells to Tumor Cells**

### **2.1 Methods and Materials**

#### **2.1.1 Cell lines and reagents**

HEK293FT cells were purchased from Invitrogen. PBMCs transduced with GP100 TCR, and Mel-526 cells are kindly provided by Dr. Kammula's lab. PBMCs were cultured in complete medium (RPMi-1640, 10% heat-inactivated human AB serum (Gemini Bio-Products, Woodland, CA), 2 mM L-glutamine (Invitrogen, Carlsbad, CA), 100 units/penicillin (Lonza), 100 µg/ml streptomycin (Lonza), 50 µg/ml gentamicin (Gibco), 10 mM HEPES (Lonza), and 250 ng/ml Amphotericin B (Invitrogen)). MEL-526 cells were cultured in RPMi-1640 (Hyclone) supplemented with 10% FBS (Gibco) and 1% penicillin-streptomycin (Corning). HEK293FT cell was cultured in D10 medium supplemented with GlutaMax (Gibco), 10% FBS (Gibco) and 1% penicillin-streptomycin (Corning).

#### **2.1.2 CRISPR library preparation, lentivirus production and transfection**

Human CRISPR IncRNA Activation Pooled Library (SAM - 3 plasmid system) was purchased from Addgene (#1000000106). The library contained 96,458 sgRNAs targeting 10,504 transcription start sites (TSS); also contained 500 control sgRNAs. All steps of CRISPR screening were performed according to Genome-scale CRISPR-dCas9 transcriptional activation screening

protocol [64]. Pooled sgRNA library was first amplified by transformation. Antibiotic killing curve was performed before lentivirus production for all three plasmids with three different antibiotics. The concentration for selection was indicated below: 400 µg/ml Zeocin (Invivogen, #ant-zn-05), 10 µg/ml Blasticidin (Sigma, #15205) and 250 µg/ml Hygromycin (RPI, #31282-04-9). HEK293FT cells were cultured in T225 flask and maintained at 70% confluency. Cells for transfection were seeded at 80-90% confluency; transfection used Lipofectmaine 2000 (Invitrogen, #11668-019) and Plus reagent (Invitrogen, #11514-015). For each T225 flask, 2250 µl of Opti-MEM (Gibco) was mixed with 15.3 µg pMD2.G (Addgene), 23.4 µg psPAX2 (Addgene) and 30.6 µg target plasmid. After 8 hr of transfection, the medium was changed to complete DMEM medium. Lentivirus was harvested after 48 hr of transfection and filtered with 0.45 µm filter. The MOI value was determined and calculated for library lentivirus. MEL-526 parental cells were first transduced with dCas9-VP64 and MS2-P65-HSF1 lentivirus individually and selected with antibiotics for 5 days until no viable cells in the no-virus treatment control cells. Then MEL-526 cells transduced with SAM CRISPR system were infected with lncRNA sgRNA library lentivirus at an MOI of 0.3, maintained a coverage of >500 cells expressing each sgRNA. Cells were selected with zeocin for 6 days.

### **2.1.3 Screening and sgRNA sequence**

The screening of candidate lncRNAs was performed by co-culture of human melanoma MEL-526 library cells and human CD8<sup>+</sup> T cells. The ratio of effective T cell and tumor cell (E:T) and time point was first determined by pre-experiment; the final ratio of E:T used for CRISPR screening was 0.5 and co-culture time point was 4 hr. Two biological replicates were performed (S1 and S2) and each batch used a total of  $5 \times 10^7$  MEL-526 library cells in round bottom 96-well

plate. For each well,  $10^5$  tumor cells and  $5 \times 10^4$  CD8<sup>+</sup> T cells were seeded and maintained for 4 hr. The control group was the MEL-526 library cells cultured under the same condition without co-culture the T cells. After co-culture, T cells were removed, and the surviving tumor cells were transferred to a new culture dish. The surviving tumor cells were cultured for another 7 days until the cells reached  $>5 \times 10^7$  cells for each screened batch. The sgRNA sequence for activate target lncRNAs and control sgRNAs are listed below: sgDT-1: 5'-AGCCTTGGGAGCTGGCTGGG-3'; sgDT-2: 5'-CCCAGCGTCCGTCCATGGCG-3'; sg198-1: 5'-ATGAAAGACTTGCTGTTCTT-3'; sg198-2: 5'-TCTCTCTCTAACGCATACAA-3'; sgNT-1: 5'-CTGAAAAGGAAGGAGTTGA-3'; sgNT-2: 5'-AAGATGAAAGGAAAGGCGTT-3'.

#### **2.1.4 Genomic DNA extraction and sample preparation for NGS**

Genomic DNA extraction and PCR steps were performed according to the protocol described above [64]. Genomic DNA was harvested from the surviving cells after the screen maintained a coverage of  $>500$  cells expressing each sgRNA using the Zymo Research Quick-DNA Midiprep plus kit (Zymo Research, #FD4075) according to the manufacture's protocol. Then the extracted genomic DNA was used for PCR used NEBNext High Fidelity 2 X PCR Master Mix (NEB, #M0541S). The NGS primers that amplify the sgRNA library were listed in the protocol [64]. All five samples used 10 different forward primers and each sample used 1 reverse primer with specific barcode. CRISPR library plasmid used 20 ng as template, control and screened samples used 1  $\mu$ g genomic DNA as template. The PCR reaction was performed using following condition: 98 °C for 3 min, 23 cycles of (98 °C for 10s, 63 °C for 10s, 72 °C for 25s), and 72 °C for 2 min. PCR products were pooled and then purified by QIAquick PCR purification kit (Qiagen, #28106). Concentration of the resulting cDNA was measured by Nanodrop. 2  $\mu$ g of PCR purification pooled product was

loaded onto a 2% agarose gel. The target size was around 280 bp. We then performed gel extraction by QIAquick Gel extraction kit (Qiagen, #28706) according to the manufacturer's directions.

### **2.1.5 Re-annotation of Human CRISPR lncRNA Activation Pooled Library**

The sgRNAs from Human CRISPR lncRNA Activation Pooled Library [8] targeting the lncRNAs were annotated by RefSeq [65]. To take the advantage of other sequencing data including TCGA database, we re-annotated these sgRNAs based on the GENCODE v19 annotation (GRCh37, [https://www.encodegenes.org/human/release\\_19.html](https://www.encodegenes.org/human/release_19.html)). First, we obtained the loci of original lncRNAs based on the hg19.refGene.gtf annotation from UCSC Genome Browser (<http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/genes/hg19.refGene.gtf.gz>) and CabiliSuppDataSet1\_lincRNAs\_transcripts.gtf [65]. Second, we considered each sgRNA sequence as a DNA motif and used HOMER 'find' function [66] to identify the loci of the sgRNAs. For each original target lncRNA, we identified the relative loci (shift from sgRNA center to transcription start site (TSS)) of the corresponding sgRNAs in its promoter region (-2 kb to +2 kb of TSS). For each sgRNA, we first considered the possible loci in the upstream of target gene. If there are multiple possible loci in the upstream of TSS, we kept the one closest to the TSS. Then, if there are no possible loci in the upstream of TSS, we kept the one closest to the TSS in the downstream of TSS. Then, we calculated the exact loci for most sgRNAs in the genome. With the coordinates of the sgRNAs, we annotated these sgRNAs using GENCODE v19 annotation. The transcripts with at least 2 sgRNAs mapped to the promoters (-1000 b to + 100 b of TSS) were considered as targets of the sgRNAs in the library. Besides, some original RefSeq lncRNAs were not found in the annotation file, we directly kept these lncRNAs and corresponding sgRNAs in the

new sgRNA library annotation. Nontargeting sgRNAs were also kept as control for the following analysis.

### 2.1.6 Quantitation of sgRNAs and CRISPR screening analysis

The read counts of sgRNAs were first quantified by `count_spacers.py` [64]. Then a pseudocount of 1 was added to each sgRNA's read count and a normalization was done using the total read counts in the sample. The fold change for each sgRNA was obtained by dividing the normalized read count in the CD8<sup>+</sup> T-cell co-cultured sample by the control followed by the base 2 logarithm transformation, for the purpose of normalization and making our data more normally distributed. To test the consistency of sgRNA read counts and log<sub>2</sub>-fold change in our screening, Spearman's correlation was performed in two control samples and two replicate samples. Spearman's correlation was performed between lncRNA expression and immune-related protein-coding genes.

RNAi Gene Enrichment Ranking (RIGER) [21] was used to select candidate genes with enriched or depleted sgRNAs in the CD8<sup>+</sup> T-cell co-cultured sample comparing to control sample. Kolmogorov-Smirnov method was selected, and the parameters were set to the recommended values as described in the protocol [64]. A weighted average of p value was calculated in order to select top candidate genes with p<0.05 in both CD8<sup>+</sup> T cell co-cultured replicates.

$$\text{Weighted average } p = \alpha \times \frac{-\log_{10}(p1) - \log_{10}(p2)}{2}$$

$$\alpha: \text{ consistency weight, } \begin{cases} p1 < 0.05 \text{ and } p2 < 0.05, \alpha = 2 \\ p1 \geq 0.05 \text{ and } p2 \geq 0.05, \alpha = 0.5 \\ \text{else } \alpha = 1 \end{cases}$$

Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK v0.5.9.4) [23] was also used for sgRNA quantification and MAGeCK RRA algorithm was then utilized for the candidate gene selection with 500 non-targeting control sgRNAs for normalization and generation of the null distribution. Parameters were set as defaulted.

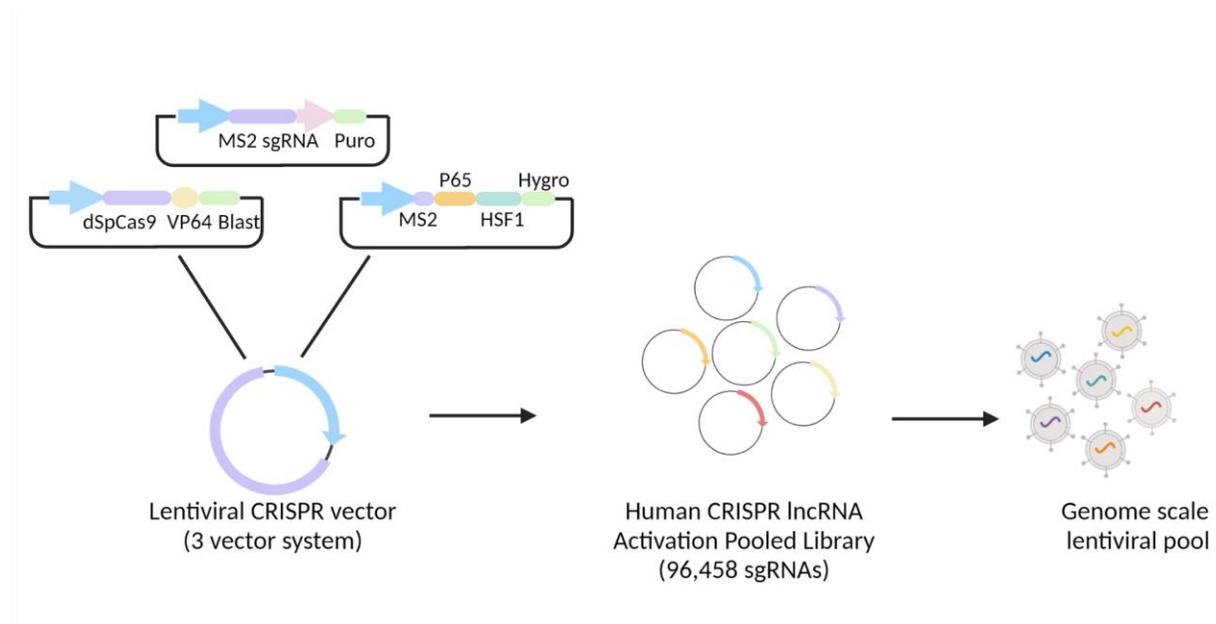
Candidate genes with  $P < 0.05$  in both CD8<sup>+</sup> T-cell co-cultured replicates by RIGER and MAGeCK were selected as significantly depleted/enriched hits from our CRISPR screening. In our CRISPR screening data, 1259 protein coding genes were ranked by the beta score from the MAGeCK result [27]. GSEA analysis was then done with the pre-ranked gene lists using C2 Curated gene set from MsigDB [67].

## **2.2 Results**

### **2.2.1 Genome-wide screening for immune regulating lncRNA using an established CD8<sup>+</sup> T-tumor cells co-culture system**

Previous publications have used CRISPR synergic activation mediator (SAM) system to identify crucial non-coding genes contributing drug resistance [12]. Here in this study, we aimed to perform genome-wide lncRNA activation screening using our established primary human CD8<sup>+</sup> T cell and melanoma cell co-culture system [63]. The primary human CD8<sup>+</sup> T cells were transduced with a specific T cell receptor (TCR) recognizing human gp100 peptide antigen. To adaptively utilizing this co-culture system for later CRISPR screening, we first generated human melanoma MEL-526 cells containing the CRISPR sgRNA library. We used three vector SAM systems which contained dCas9-VP64, MS2-P65-HSF1 and sgRNA library individually (**Figure**

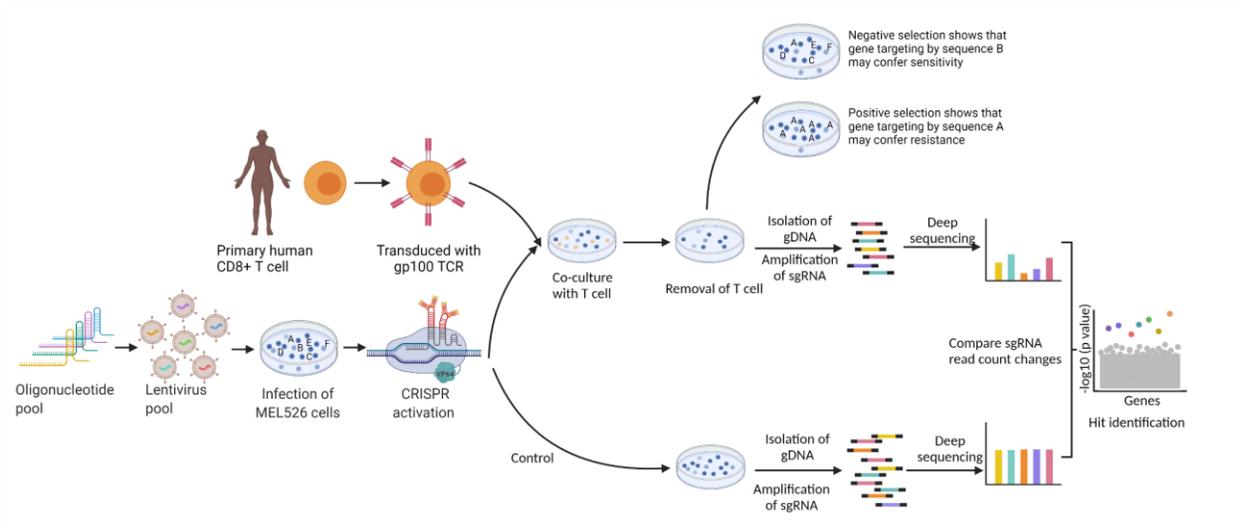
2). We transduced MEL-526-dCas9-VP64-MPH cells with lncRNA SAM sgRNA library at a multiplicity of infection (MOI) <0.3.



**Figure 2 Three vectors system used in CRISPRa screening**

The three-vector system containing dCas9-VP64, MS2-P65-HSF1 and sgRNA library was transduced into MEL526 cells to obtain genome scale activation.

Transduced MEL-526 sgRNA library cells were co-cultured with human gp100 TCR transduced CD8<sup>+</sup> T cells at E (effector): T (tumor) ratios of 0.5 for 4 hr for 2 biological replicates in independent screens (**Figure 3**). After that, T cells were removed and cancer cells were cultured for another 7 days. Then, genome DNA was isolated from MEL-526 cells and specific adaptors were added to amplify sgRNAs for deep sequencing. Through the comparison of sgRNA read counts in control and treatment group, significant depleted/enriched targets were identified.

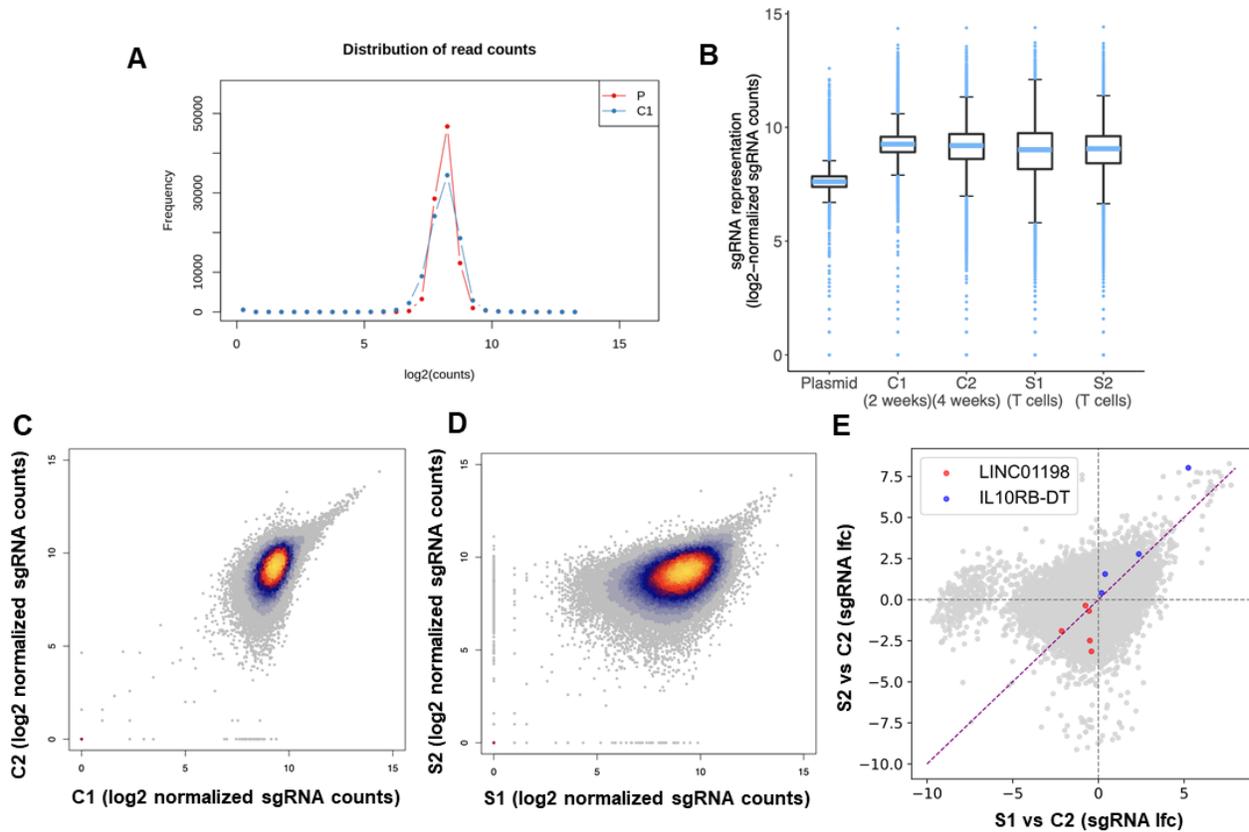


**Figure 3 Overview of CRISPRa screening with a co-culture system**

Design of the genome-wide coculture assay to identify genes regulating tumor response to CD8<sup>+</sup> T cells.

## 2.2.2 The development of a combined analysis pipeline to analyze CRISPR screening data

Initially, the read count distribution of our plasmid and control sample was checked to make sure library cells have no sgRNA distribution loss compared with original plasmid (**Figure 4A**). We then compared the sequenced distribution of sgRNAs in five samples which included the original library, two different cultured time point controls, and two biological replicates. The sgRNA distributions were of similar level, suggesting no technical issues when constructing the screening (**Figure 4B**). Next, we quantified the consistency of these two replicates through the comparison of sgRNA counts and log<sub>2</sub> fold change after T cell treatment. All of them demonstrated high correlation between replicates (**Figure 4C, D and E**).



**Figure 4 Quality control for CRISPR screening analysis**

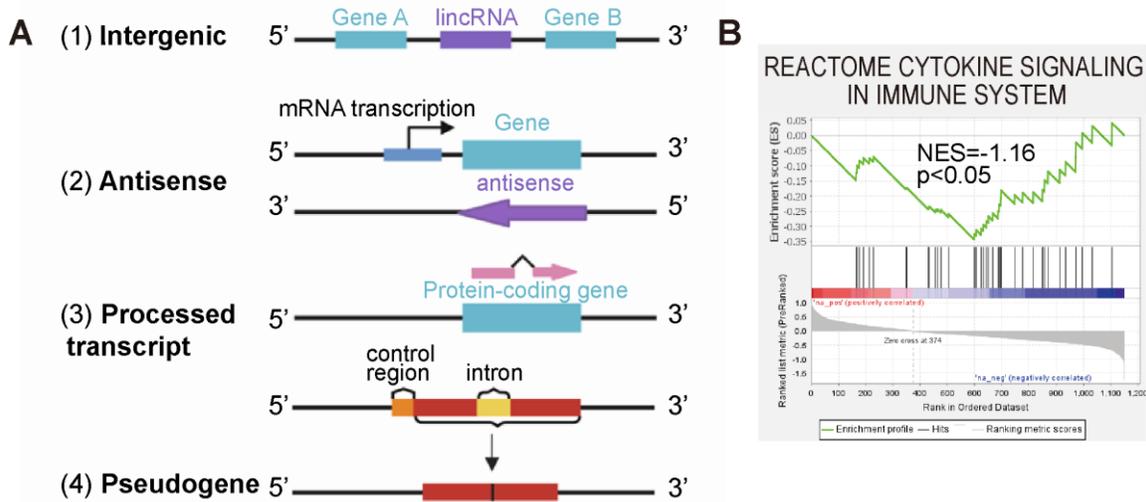
**A.** Read count distributions for sgRNAs from plasmid and control group. **B.** Box plots showing distribution of sgRNA frequencies after control or T cell coculture from  $n=2$  infection replicates. **C and D.** Scatter plot for sgRNA read counts in two replicates showing good correlation. (C1 vs C2,  $\rho=0.5$ ,  $p<0.001$ ; S1 vs S2,  $\rho=0.42$ ,  $p<0.001$ ) **E.** sgRNA read count changes after treatment ( $\rho=0.28$ ,  $p<0.05$ ). P: plasmid C1: control 1, cultured for 2 weeks, C2: control 2, cultured for 4 weeks, S1 and S2 are two treatment replicates from C2 coculture with T cells (The marked sgRNAs are targeting the lncRNA hits we do further validation)

In this study, we used a previously established lncRNA sgRNA library [8]. Since the sgRNA library was designed to target lncRNAs Transcription Start Sites (TSS), there could be protein coding genes that share TSSs with targeted lncRNAs that can be activated by the systems. To keep the rigor of our analysis, we reannotated the sgRNAs to identify all their possible targets

(see **Methods**). Generally speaking, the sgRNAs in our library can target 2236 intergenic lncRNAs, 659 antisense lncRNAs, 1259 protein-coding genes and 982 other types of transcripts (**Table 1**, **Figure 5A**).

**Table 1** Transcripts included in the sgRNA library after reannotation

Sub-library	Genes	Transcripts	sgRNAs
Intergenic lncRNA	2236	4160	46141
Antisense	659	1441	14514
Protein-coding gene	1259	5160	44252
Other	982	2290	22702
Non-targeting control			500

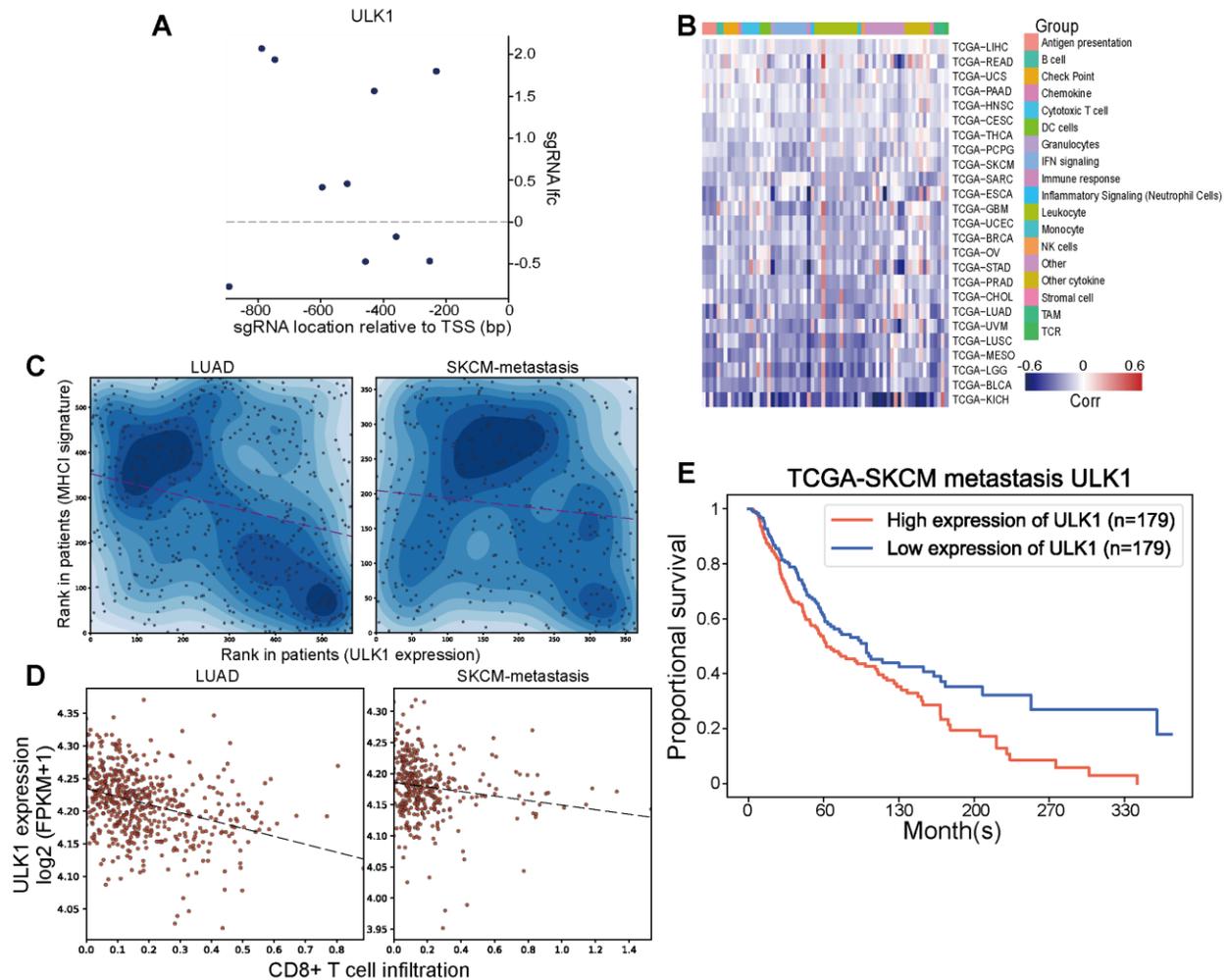


**Figure 5** Relationships between lncRNAs and protein-coding genes

**A.** Relationships between lncRNAs and protein-coding genes in the genome. **B.** GSEA analysis of MsigDB C2 gene signatures for protein-coding genes in library.

As shown above, the CRISPRa library will inevitably activate some protein-coding genes, although it has been designed to target the lncRNA TSS. The 1259 protein-coding genes which

may be activated by the library left us a chance to identify some protein-coding genes regulating cancer cell response to CD8<sup>+</sup> T cell cytotoxicity. In this regard, we performed a GSEA analysis on the genes based on their MAGeCK beta value. Overall, this analysis revealed that the negatively selected protein-coding genes were highly enriched in cytokine signaling (**Figure 5B**, NES=-1.16, p<0.05), suggesting our screening captured genes involved in the regulation of immune response. For example, the protein-coding gene ULK1 is among the significantly positive selected genes. In 10 sgRNAs targeting the promoter region of ULK1, 6 showed significant enrichment in both replicates (**Figure 6A**), which indicated that ULK1 may lead to the suppression of immune response in tumorigenesis. To further investigate ULK1's regulation of immune reaction in the tumor microenvironment, we found a negative association between ULK1's expression and 68 immune signatures [68] in 25 types of cancer in TCGA patient data (**Figure 6B**). Particularly, we discovered ULK1's expression was highly negatively correlated with MHC1 immune signature in cancer patients (metastasis SKCM: rho=-0.11, p<0.05, LUAD: rho=-0.25, p<0.0001) (**Figure 6C**). Moreover, ULK1's expression level was negatively associated with CD8<sup>+</sup> T cell tumor infiltration ratio in LUAD and SKCM (**Figure 6D**). Consistent with its function, we identified ULK1 as an immune inhibition oncogene, higher ULK1 expression demonstrated poor survival rate in metastatic SKCM patients (**Figure 6E**). Indeed, ULK1 was recently identified as an oncogene that inhibits antigen presentation and shows anti-immunity function in lung cancer [69]. Mechanistically, ULK1 suppressed immunoproteasome activity by inducing autophagic flux, and further repression of ULK1 showed synergized effect with anti PD-1 therapy. In summary, these results suggest that our CRISPR screening utilizing the CD8<sup>+</sup> T/tumor cell co-culture assay can successfully recapitulate the established immune regulator.

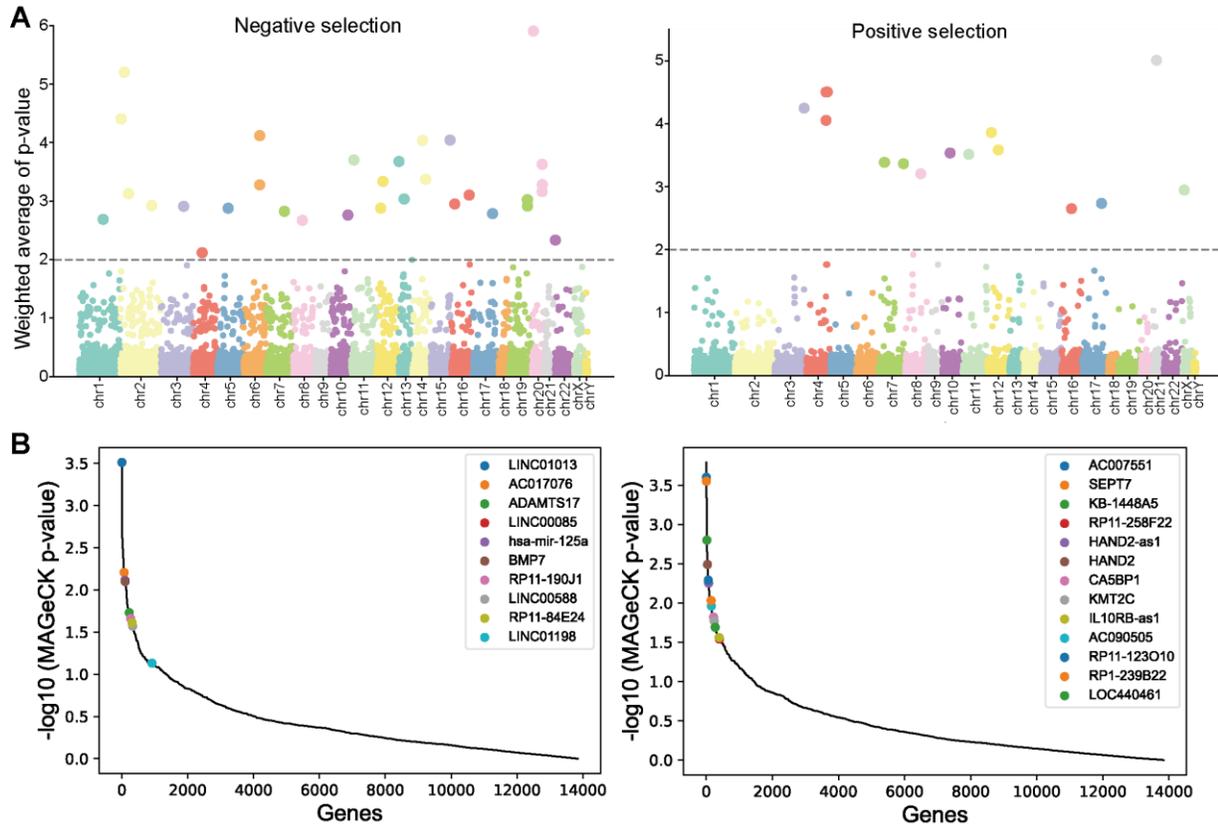


**Figure 6 ULK1's repression of anti-tumor immunity**

**A.** The location and log<sub>2</sub> fold change for 10 sgRNAs targeting ULK1 TSS. **B.** ULK1's correlation with 68 immune signatures in multiple TCGA cancer types. **C.** ULK1's correlation with MHC1 signature in LUAD and metastatic SKCM patients. **D.** ULK1's correlation with CD8<sup>+</sup> T cell infiltration (estimated from TIMER) in LUAD and metastatic SKCM patients. **E.** K-M survival curve representing the proportional survival of metastatic SKCM patients with different ULK1 expression level.

The top hits for both positive (resistance of T cell function) and negative (induce T cell function) selections were characterized by RNAi Gene Enrichment Ranking (RIGER) (**Figure 7A**) in individual biological replicates. Together, the screening identified 62 candidate gene transcripts,

including 42 negatively selected transcripts (423 sgRNAs) and 20 positively selected transcripts (166 sgRNAs). MAGeCK analysis was then used to confirm the robustness of our screening result against different statistical models for both positive and negative selections (**Figure 7B**). The p-values of high-ranking genes from RIGER result remain robust across the replicates and both algorithms used for analysis.



**Figure 7 Top hits identified from both negative and positive selection**

**A.** RIGER analysis for both negative selection and positive selection. **B.** MAGeCK analysis result showing overlap with RIGER.

### **3.0 Novel hits were identified from both negative and positive selection**

## **3.1 Methods and Materials**

### **3.1.1 Public databases**

#### **3.1.1.1 Genotype-Tissue Expression (GTEx) database**

The GTEx project is a public resource providing researchers insights of gene expression and cellular regulation in a tissue-specific way [70]. In the latest version (v8), a total number of 17382 samples of 54 tissues obtained from 948 healthy donors is included, where molecular assays including Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES) and RNA-seq are available. The RNA-seq datasets of normal 407 whole blood samples, 290 breast samples, 1203 skin samples and 162 spleen tissue samples were obtained from GTEx [71]. Data for following analysis were downloaded from GTEx Portal (<https://www.gtexportal.org/>).

#### **3.1.1.2 TCGA database**

The TCGA project focused on cancer genomics which revealed the molecular characteristic over 20,000 primary cancer and matched normal samples across 33 cancer types [72]. Clinical information and transcriptome data from TCGA of 1217 Breast cancer (BRCA) patients, 472 Skin Cutaneous Melanoma (SKCM) patients, 145 Acute Myeloid Leukemia (LAML) patients, 585 Lung Adenocarcinoma (LUAD) patients, 529 Lower Grade Glioma (LGG) patients, 563 Thyroid Cancer (THCA) patients and 173 Esophageal cancer (ESCA) patients were obtained from Genomic Data Commons (GDC) portal (<https://portal.gdc.cancer.gov/>) [73]. The annotation for

the transcriptome data was performed on human reference genome GRCh38 and gene expression level was then quantified by FPKM upper quartile.

### **3.1.1.3 GEO datasets**

RNAseq data for metastatic melanoma patients after treated with anti-PD-1 therapy was obtained from Gene Expression Omnibus (GEO): accession GSE78220 [74]. Transcriptomic data obtained from microarrays for breast cancer patients were obtained from GSE42568 [75] (ER positive survival), GSE16391 [76] (DT), GSE20685 [77], GSE20711 [78].

## **3.1.2 Cell line expression data**

### **3.1.2.1 RNA-seq analysis**

MEL-526, MCF-7 stable cells transduced with individual sgRNAs targeting the selected lncRNA loci or with non-targeting control sgRNAs (sgRNA sequence) were used for the activation group. MEL-526 cells treated with siRNA targeting selected lncRNAs (DT and 198), were transiently transfected by RNAi MAX reagent (Invitrogen, #13778-150) for 48 hr. Total RNA was extracted from activated and knockdown cells by Trizol. Purification of RNA used RNA clean & concentrator kit (Zymo Research, #11514-015). Purified RNAs were sent to UPMC Children's Hospital for quality control, RNA-seq library construction and sequencing. STAR (Spliced Transcript Alignment to a Reference) and RSEM (RNA-Seq by Expectation Maximization) were used to profile RNA-seq data of MEL526, MCF7 cell lines after IL10RB-DT, LINC01198 siRNA knockdown and IL10RB-DT, LINC01198 activation. The quantification of gene expression was log<sub>2</sub>-transformed FPKM and annotation was based on human reference genome GRCh38.

### 3.1.2.2 Gene Set Enrichment Analysis (GSEA)

In the TCGA data, 19,668 protein coding genes were ranked by their expression Spearman correlation coefficients with a lncRNA in the cancer samples. In the RNA-seq data with lncRNA activation and knock-down in the MEL526 and MCF-7 cell line, 20,442 protein coding genes were ranked by their log<sub>2</sub>-fold change of the expression in the treated sample compared to the control sample. Gene Set Enrichment Analysis [22] was then done with the pre-ranked gene lists using cancer hallmark gene set or C2 Curated gene set from MsigDB [67].

### 3.1.3 Statistical analysis

We also obtained the 68 immune signatures in 18 different groups from previously reported studies [68]. To explore the function of lncRNA hits in immune response, Spearman's correlation coefficient was calculated between lncRNA expression and immune signature enrichment scores in TCGA samples. The relative immune infiltration of CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells, DC, neutrophils, macrophages and B cells was estimated by Tumor Immune Estimation Resource (TIMER) [79] for TCGA-BRCA patients. Then Spearman's correlation was performed between immune infiltration level and lncRNA expression.

Data between the two groups were compared using a two-tailed unpaired Student's t-test or the Mann-Whitney test in different types of data (depending on normality of the distribution). Survival analysis was performed through log-rank test and Kaplan-Meier estimate analyses.

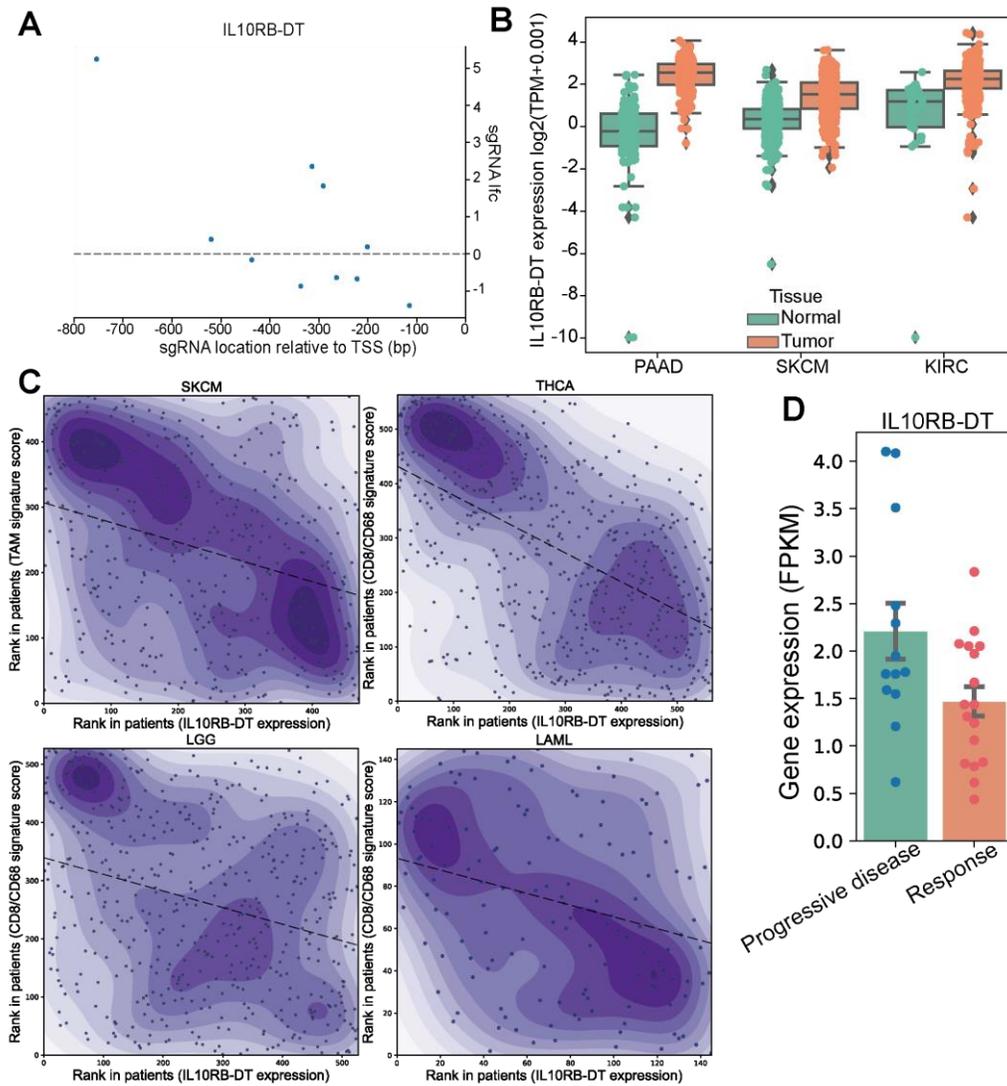
Analysis was performed with Python 3.8.0 and GraphPad Prism. Significance threshold was defined as  $p < 0.05$ . The Benjamini-Hochberg adjustment was used to the false discovery rates (FDR).

## 3.2 Results

In total, we identified 6 transcripts of protein-coding genes, 9 intergenic lncRNAs, 3 antisense lncRNAs and 2 pseudogenes among positive selection; 6 protein-coding genes, 23 intergenic lncRNAs, 9 antisense lncRNAs, 4 other types of transcripts in negative selection. The identification of ULK1's previously reported role in immune modulation motivated us to look into the top ranked lncRNAs characterized by the CRISPR screening.

### 3.2.1 CRISPR screening identified lncRNA IL10RB-DT as a suppressor of tumor immune response

Among the lncRNAs that are identified in the T cell-cancer cell co-culture assay from positive selection, a barely studied lncRNA, IL10RB-DT brought us attention. In 10 sgRNAs targeting the promoter region of IL10RB-DT, 5 showed significant enrichment (**Figure 8A**). IL10RB-DT showed significantly higher expression levels in tumor tissue especially in melanoma compared with normal tissue from GTEx database (**Figure 8B**). We found IL10RB-DT's expression also was significantly negatively associated with CD8/CD68 signature in 563 THCA patients ( $\rho=-0.53$ ,  $p<0.05$ ), 145 LAML patients ( $\rho=-0.28$ ,  $p<0.05$ ), 529 LGG patients ( $\rho=-0.29$ ,  $p<0.05$ ) and tumor associated macrophage (TAM) signature in 472 melanoma patients ( $\rho=-0.3$ ,  $p<0.05$ ) (**Figure 8C**). Additionally, a pre- and post- PD-1 treatment patient cohort [74] study revealed that IL10RB-DT expression is significantly lower in PD-1 treatment responded melanoma patients ( $p<0.05$ , **Figure 8D**). These results suggest IL10RBDT as a repressor for tumor immune response and could be a potential biomarker for immunotherapy treatment.

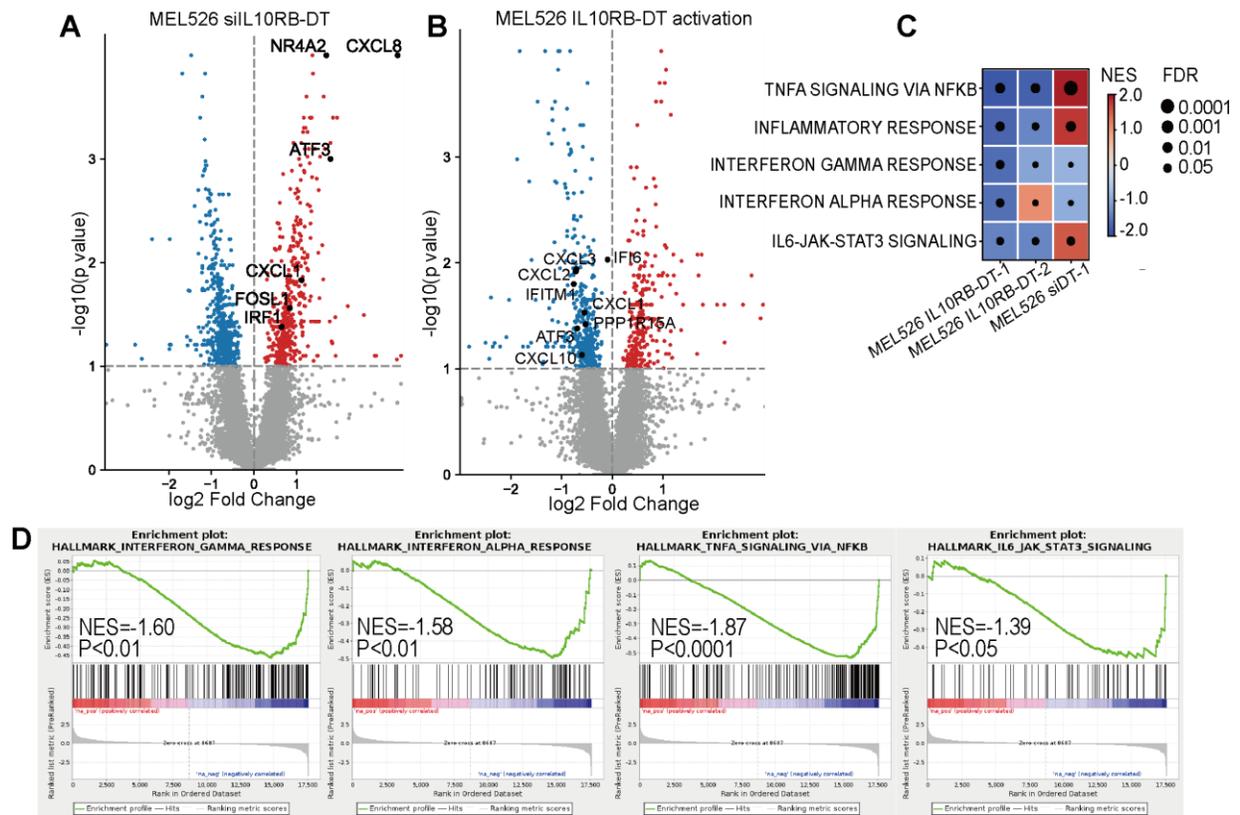


**Figure 8 IL10RB-DT as a tumor immune suppressor from integrated clinical data**

**A.** The location and log<sub>2</sub> fold change for 10 sgRNAs targeting IL10RB-DT TSS. **B.** The expression level of IL10RB-DT in several types of tumor compared with matched normal tissues. **C.** IL10RB-DT correlation with immune signature in SKCM, THCA, LGG and LAML patients. **D.** The pre-treatment expression level of IL10RB-DT for patients with different response to anti-PD1 treatment.

To further understand how IL10RB-DT leads to immune response suppression, we performed RNA-seq of IL10RB-DT activated and siRNA knockdown MEL-526 cells. The RNA-seq analysis identified 137 and 302 genes significantly upregulated; 178 and 250 genes

downregulated in MEL-526 IL10RB-DT activation and KD samples respectively with immune related genes indicated on the volcano plots (**Figure 9A and B**). Further cancer hallmark analysis demonstrated genes downregulated in IL10RB-DT activated cells were enriched in several immune related signaling pathways including both type II and type I interferon pathways such as TNF- $\alpha$  via NF- $\kappa$ B, IFN- $\alpha$  response, IFN- $\gamma$  response (**Figure 9C and D**). Immune-related genes with significant changes in DT activated and KD cells also were indicated in the heatmap (**Figure 9A and B**). To conclude, IL10RB-DT inhibited tumor response to CD8<sup>+</sup> T cell cytotoxicity through type I or type II interferon pathways in MEL-526 cells.

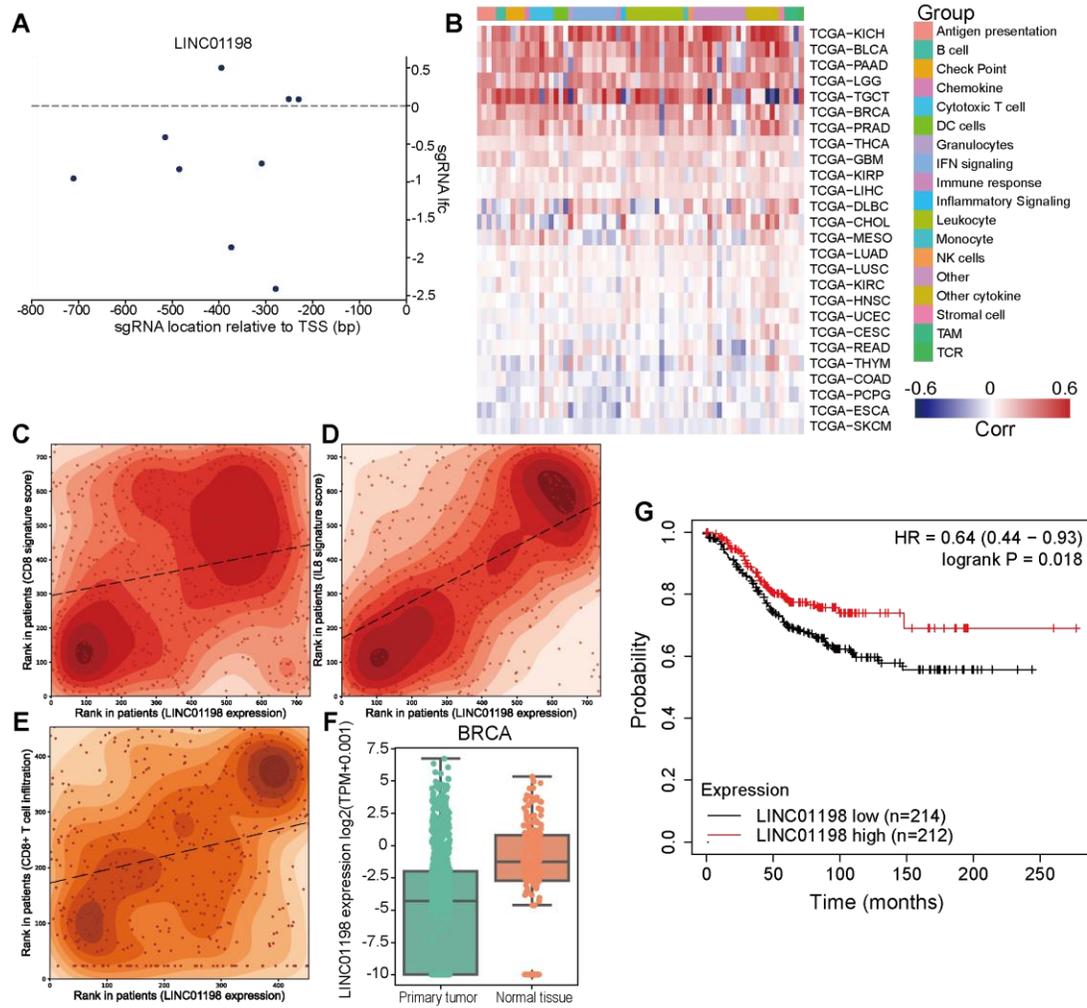


**Figure 9** IL10RB-DT inhibits the killing effect of CD8<sup>+</sup> T cells through immunogenic pathways

**A and B.** The volcano plot showing differential expression of siIL10RB-DT and IL10RB-DT activation in MEL526 cells. **C.** IL10RB-DT's regulation on immune-related pathways. **D.** Significant downregulation pathways after IL10RB-DT activation in MEL526 cells.

### 3.2.2 LncRNA LINC01198 is identified as an activator of tumor immune response

Our study identified both immune resistance and immune stimulation lncRNAs, and LINC01198 is among one of the top hits from negative selection. In 10 sgRNAs targeting the promoter region of LINC01198, 6 showed significant depletion (**Figure 10A**). LINC01198 was also positively correlated with cytotoxic T cell activity and antigen presentation gene expression in multiple cancer types included breast cancer (**Figure 10B**). Specifically, LINC01198 showed a positive correlation with CD8 ( $\rho=0.23$ ,  $p<0.05$ ) and IL8 immune signatures ( $\rho=0.55$ ,  $p<0.05$ ) in 759 BRCA patients (**Figure 10C and D**). In addition, LINC01198 also is expressed at a higher level in normal tissues compared with primary tumors, and positively correlated with CD8<sup>+</sup> T cell infiltration in breast cancer (**Figure 10E and F**). From the transcription profile in 426 breast cancer patients, higher LINC01198 also lead to higher survival in this cohort (**Figure 10G**). In this regard, we stably expressed sgRNA targeting LINC01198 TSS in MEL-526 and also breast cancer cell line MCF-7 because a more significant immune-stimulating phenotype was discovered from breast cancer patients.

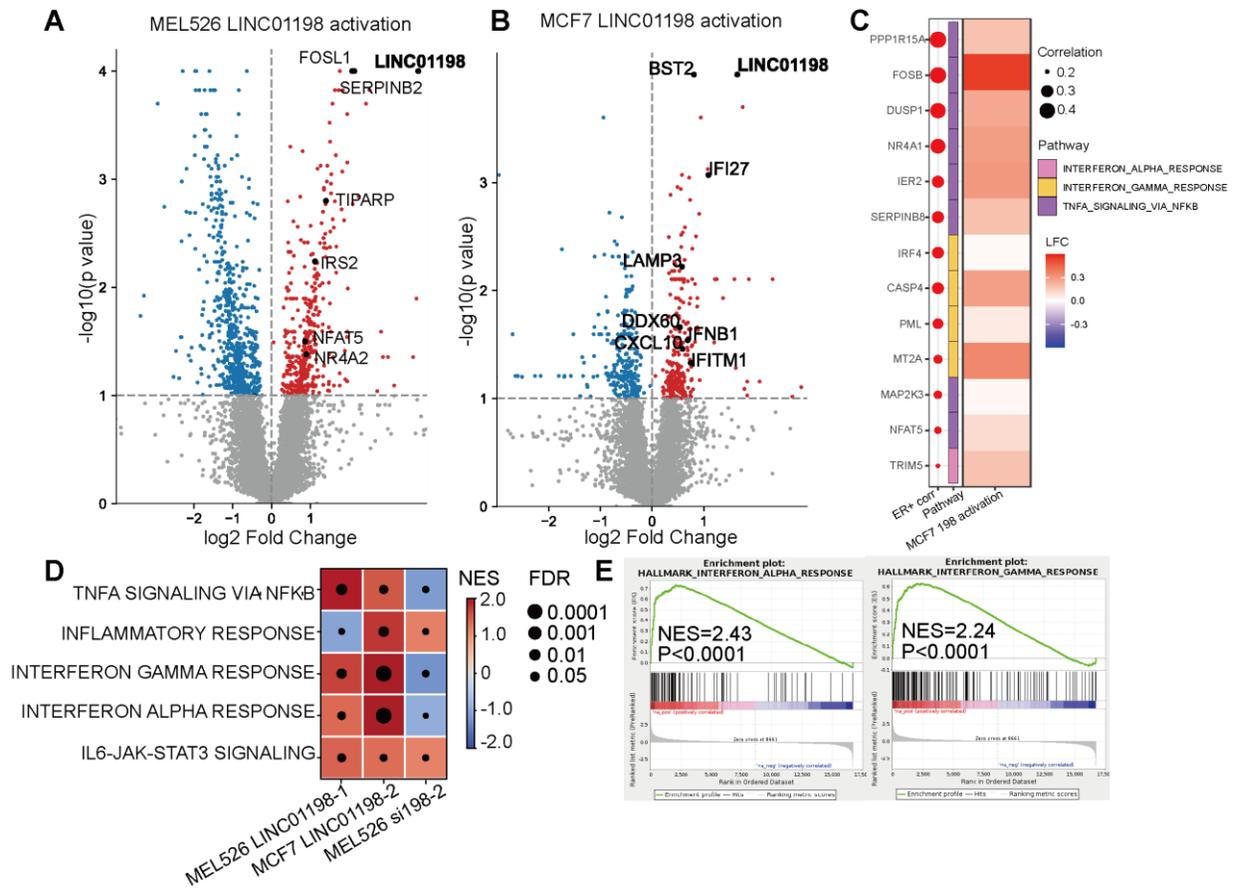


**Figure 10 LINC01198 as a positive regulator for immune response in breast cancer**

**A.** The location and log<sub>2</sub> fold change for 9 sgRNAs targeting LINC01198 TSS. **B.** LINC01198 correlation with immune signature in multiple cancer types. **C** and **D.** The correlation between LINC01198 expression and CD8 and IL8 signature in breast cancer patients. **E.** The correlation between LINC01198 expression and CD8<sup>+</sup> T cell infiltration in ER<sup>+</sup> breast cancer patients. **F.** LINC01198 expression from primary tumor and normal tissue in breast cancer. **G.** K-M survival curve representing the proportional survival of breast cancer patients with different LINC01198 expression level. LINC01198's expression is measured by Affymetrix 1553614\_a\_at (HG-U133\_Plus\_2).

To further investigate the function and mechanism of LINC01198, we also generated LINC01198 KD cells by siRNA; then performed RNA-sequencing of LINC01198 activation and

knockdown MCF-7 and MEL-526 cells. The RNA-seq analyzed volcano plot indicated 219 induced and 400 depleted genes' expression in MEL-526 cells. Meanwhile, we also identified 93 upregulated and 63 downregulated targets in MCF-7 sample; genes in Type I signaling pathway were specifically marked (**Figure 11A and B**) on the volcano plot. From the RNA-seq result, LINC01198 was one of the most significantly upregulated genes, suggesting that our sgRNA activation system can successfully activate this lncRNA in two different cell lines. Additionally, we extracted the expression of top ranked genes upregulated in LINC01198 activated cells from RNA-seq and illustrated that these genes showed high positive correlations with LINC01198 expression in ER+ breast cancer patients, showing that LINC01198 can potentially promote tumor immune response. (**Figure 11C**). Moreover, the cancer hallmark analysis (**Figure 11 D and E**) indicated the upregulated genes in 198 activated MEL-526 and MCF-7 cells were enriched in several immune response related to signaling pathways especially in type I interferon pathways (TNF- $\alpha$  signaling via NF $\kappa$ B and IFN- $\alpha$  response) and IFN- $\gamma$  response; knockdown LINC01198 showed an inhibitory effect. To conclude, LINC01198 enhanced CD8+ T cell's function and induced immune response in both melanoma and breast cancer cell lines.



**Figure 11 LINC01198 positively regulate tumor immune response through type I and type II interferon signaling pathways**

**A and B.** The volcano plot showing differential expression of LINC01198 activation in MEL526 and MCF7 cells. **C.** The positive correlation between LINC01198 and immune-related genes in ER+ breast cancer patients. **D.** LINC01198's regulation on immune-related pathways. **E.** Significant upregulation pathways after LINC01198 activation in MCF7 cells.

## 4.0 Discussion

The biological functions for the majority of long non-coding RNAs are still unknown now. To identify functional lncRNAs that may regulate tumor response to CD8<sup>+</sup> T cell cytotoxicity, we performed high-throughput CRISPRa screening with a co-culture system of melanoma cells and CD8<sup>+</sup> T cells. Given the association of these hits with antigen presentation signatures and cytotoxic T cell infiltration, this screening provides a list of novel lncRNAs that may act as targets or biomarkers to overcome the resistance and improve the outcome of immunotherapy. Since the expression level of lncRNAs is relatively low in cells and tissues compared with protein-coding genes [80], the activation system offers the advantage of overexpressing lncRNAs from the endogenous locus by designing sgRNAs targeting the promoter region of target genes. The activation system and RNA interference system we used in this project demonstrated both gain and loss of function studies related to immune response for candidate lncRNAs.

Although the candidate genes identified from our screening are in the context of melanoma cell line, we also discovered these lncRNAs showed significant correlation with immune signatures in other types of cancer patients. In particular, LINC01198 demonstrated positive association with immune signatures in breast cancer patients. In this regard, it would be better if CRISPR screenings are performed in multiple types of cell models to comprehensively understand the function of lncRNAs. Another intrinsic problem in the study of non-coding genes is the non-conservation of target genes between human and mice [81], which increases the difficulty of exploring lncRNA's function *in vivo*. It is well acknowledged that the cancer microenvironment is a complicated system, thus, to better understand the role of non-coding genes in cancer immunology, *in vivo* screening can be considered. In addition, due to the rapidly updated non-

coding RNA libraries and the restriction of the model for studying non-coding RNAs, a customized library with conserved non-coding RNAs and proper controls can be developed.

From the CRISPR screening, we have successfully identified and validated that the lncRNA IL10RB-DT is an immune response resistance gene. Future studies are necessary to show what is the relationship between IL10RB-DT and immune signature in cancer patients. Moreover, IL10RB-DT is the antisense strand of IL10RB (Interleukin 10 Receptor Subunit Beta); IL10RB-DT could potentially cis-regulate the transcription of a neighboring protein-coding gene. More experiments need to be done to prove the hypothesis of cis-regulation which may contribute to IL10RB-DT's function in immune response.

We also showed that the lncRNA LINC01198 is an immune response stimulator gene through integrated analysis. However, future studies are still needed to uncover the mechanism of its regulation for immune response. We are also interested to see if LINC01198 exerts its function through interaction with protein-coding genes such as transcription factor so that it can regulate a group of immune-related genes with a relatively low expression level in the cell.

### Appendix A Transcripts identified from negative selection

<b>Id</b>	<b>S1vsC2_ Score</b>	<b>S1vsC2_ p.value</b>	<b>S2vsC2_ Score</b>	<b>S2vsC2_ p.value</b>	<b>Transcript_type</b>	<b>MAGe CK hit?</b>
ENST00000268070	1.573	0.032	1.883	0.003	protein_coding	Y
ENST00000341141	1.591	0.028	1.487	0.046	transcribed_unprocessed_pseudogene	N
ENST00000395863	1.597	0.024	1.708	0.010	protein_coding	Y
ENST00000395864	1.538	0.038	1.640	0.018	protein_coding	N
ENST00000400596	1.636	0.020	1.726	0.010	processed_pseudogene	N
ENST00000412321	1.547	0.038	1.694	0.014	lincRNA	N
ENST00000414795	1.919	0.001	1.818	0.005	antisense	Y
ENST00000416521	1.547	0.038	1.606	0.024	lincRNA	N
ENST00000419883	1.568	0.033	1.494	0.045	lincRNA	N
ENST00000426580	1.547	0.038	1.694	0.014	lincRNA	N
ENST00000426725	1.920	0.002	1.623	0.023	antisense	N
ENST00000428036	1.673	0.016	1.493	0.046	antisense	N
ENST00000433911	1.547	0.038	1.694	0.014	protein_coding	N
ENST00000438436	1.919	0.002	1.623	0.023	antisense	N
ENST00000440887	1.805	0.005	1.512	0.041	antisense	N
ENST00000442182	1.527	0.044	1.487	0.047	lincRNA	N
ENST00000443303	1.712	0.013	1.818	0.006	lincRNA	Y
ENST00000444501	1.673	0.016	1.493	0.046	antisense	N

ENST00000450594	1.597	0.024	1.708	0.010	protein_coding	Y
ENST00000452212	1.673	0.016	1.493	0.046	antisense	N
ENST00000458028	1.712	0.012	1.501	0.044	lincRNA	N
ENST00000458489	1.512	0.047	1.532	0.037	lincRNA	Y
ENST00000469268	1.532	0.038	1.536	0.033	antisense	N
ENST00000506250	1.569	0.033	1.515	0.040	lincRNA	N
ENST00000521663	1.512	0.047	1.493	0.045	lincRNA	Y
ENST00000538693	1.498	0.039	1.465	0.042	lincRNA	Y
ENST00000548210	1.632	0.022	1.626	0.021	lincRNA	N
ENST00000552712	1.632	0.022	1.625	0.021	lincRNA	N
ENST00000553485	1.702	0.013	1.558	0.033	lincRNA	N
ENST00000553682	1.801	0.005	1.651	0.018	lincRNA	N
ENST00000556466	1.702	0.013	1.558	0.033	lincRNA	N
ENST00000557689	1.702	0.013	1.558	0.032	lincRNA	N
ENST00000557899	1.945	0.001	1.948	0.001	lincRNA	Y
ENST00000558960	1.573	0.032	1.883	0.003	nonsense_mediated_decay	Y
ENST00000568030	1.585	0.030	1.527	0.038	lincRNA	N
ENST00000574072	1.538	0.041	1.574	0.030	retained_intron	Y
ENST00000576494	1.557	0.036	1.590	0.027	retained_intron	Y
ENST00000594153	1.547	0.038	1.606	0.024	lincRNA	N
ENST00000599175	1.547	0.038	1.606	0.024	lincRNA	Y
ENST00000601250	1.535	0.043	1.657	0.018	lincRNA	N
ENST00000602324	1.557	0.036	1.590	0.027	lincRNA	Y

ENST0000 0607993	1.391	0.031	1.368	0.039	antisense	N
---------------------	-------	-------	-------	-------	-----------	---

**Appendix B Transcripts identified from positive selection**

<b>Id</b>	<b>S1C2_Score</b>	<b>S1C2_p.value</b>	<b>S2C2_Score</b>	<b>S2C2_p.value</b>	<b>Transcript_type</b>	<b>MAGECK hit?</b>
ENST00000359562	1.638	0.015	1.836	0.002	protein_coding	Y
ENST00000380333	1.522	0.033	1.498	0.034	processed_transcript	Y
ENST00000399035	1.564	0.034	1.713	0.012	protein_coding	Y
ENST00000411998	1.557	0.036	2.072	0.000	antisense	Y
ENST00000412084	1.666	0.017	1.762	0.008	lincRNA	N
ENST00000412856	1.564	0.034	1.713	0.012	lincRNA	Y
ENST00000424194	1.564	0.034	1.713	0.012	lincRNA	Y
ENST00000432530	1.607	0.026	1.722	0.011	lincRNA	Y
ENST00000437235	1.564	0.034	1.713	0.012	lincRNA	Y
ENST00000452749	1.554	0.030	1.641	0.015	protein_coding	Y
ENST00000454508	1.759	0.006	1.732	0.009	lincRNA	Y
ENST00000475109	1.564	0.034	1.713	0.012	processed_transcript	Y
ENST00000493626	1.564	0.034	1.713	0.012	processed_transcript	Y
ENST00000502700	1.516	0.034	1.692	0.008	lincRNA	Y
ENST00000505300	1.607	0.022	1.805	0.004	processed_transcript	Y
ENST00000510268	1.638	0.015	1.837	0.002	antisense	Y
ENST00000522767	1.682	0.015	1.509	0.042	lincRNA	Y
ENST00000566260	1.395	0.048	1.398	0.046	lincRNA	N

ENST000005 68280	1.717	0.011	1.585	0.028	antisense	Y
ENST000005 89904	1.509	0.048	1.525	0.038	processed_pseud ogene	Y

## Bibliography

1. Mali, P., et al., *RNA-Guided Human Genome Engineering via Cas9*. Science, 2013. **339**(6121): p. 823.
2. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system*. Nature Protocols, 2013. **8**(11): p. 2281-2308.
3. Sander, J.D. and J.K. Joung, *CRISPR-Cas systems for editing, regulating and targeting genomes*. Bioinformatics, 2014. **32**(4): p. 347-355.
4. Qi, Lei S., et al., *Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression*. Cell, 2013. **152**(5): p. 1173-1183.
5. Gilbert, Luke A., et al., *CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes*. Cell, 2013. **154**(2): p. 442-451.
6. Konermann, S., et al., *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex*. Nature, 2015. **517**(7536): p. 583-588.
7. Klann, T.S., et al., *CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome*. Nature Biotechnology, 2017. **35**(6): p. 561-568.
8. Joung, J., et al., *Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood*. Nature, 2017. **548**(7667): p. 343-346.
9. Iyer, M.K., et al., *The landscape of long noncoding RNAs in the human transcriptome*. Nature genetics, 2015. **47**(3): p. 199-208.
10. Shalem, O., et al., *Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells*. Science, 2014. **343**(6166): p. 84.
11. Joung, J., et al., *CRISPR activation screen identifies BCL-2 proteins and B3GNT2 as drivers of cancer resistance to T cell-mediated cytotoxicity*. Nature Communications, 2022. **13**(1): p. 1606.
12. Bester, A.C., et al., *An Integrated Genome-wide CRISPRa Approach to Functionalize lncRNAs in Drug Resistance*. Cell, 2018. **173**(3): p. 649-664.e20.
13. Kurata, M., et al., *Using genome-wide CRISPR library screening with library resistant DCK to find new sources of Ara-C drug resistance in AML*. Scientific Reports, 2016. **6**(1): p. 36199.
14. Gurusamy, D., et al., *Multi-phenotype CRISPR-Cas9 Screen Identifies p38 Kinase as a Target for Adoptive Immunotherapies*. Cancer Cell, 2020. **37**(6): p. 818-833.e9.
15. Zhang, H., et al., *Functional interrogation of HOXA9 regulome in MLLr leukemia via reporter-based CRISPR/Cas9 screen*. . Elife, 2020. **9**(2050-084X (Electronic)): p. e57858.
16. Parnas, O., et al., *A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks*. Cell, 2015. **162**(3): p. 675-686.
17. Adamson, B., et al., *A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response*. Cell, 2016. **167**(7): p. 1867-1882.e21.
18. Dixit, A., et al., *Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens*. Cell, 2016. **167**(7): p. 1853-1866.e17.

19. Jaitin, D.A., et al., *Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq*. Cell, 2016. **167**(7): p. 1883-1896.e15.
20. Datlinger, P., et al., *Pooled CRISPR screening with single-cell transcriptome readout*. Nature Methods, 2017. **14**(3): p. 297-301.
21. Luo, B., et al., *Highly parallel identification of essential genes in cancer cells*. Proceedings of the National Academy of Sciences, 2008. **105**(51): p. 20380-20385.
22. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*. Proc Natl Acad Sci U S A, 2005. **102**(0027-8424 (Print)): p. 15545-15550.
23. Li, W., et al., *MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens*. Genome Biology, 2014. **15**(12): p. 554.
24. Anders, S. and W. Huber, *Differential expression analysis for sequence count data*. Genome Biology, 2010. **11**(10): p. R106.
25. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics, 2010. **26**(1): p. 139-140.
26. Kolde, R., et al., *Robust rank aggregation for gene list integration and meta-analysis*. Bioinformatics (Oxford, England), 2012. **28**(4): p. 573-580.
27. Li, W., et al., *Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR*. Genome Biology, 2015. **16**(1474-760X (Electronic)): p. 281.
28. Wang, B., et al., *Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute*. Nature Protocols, 2019. **14**(1750-2799 (Electronic)): p. 756-780.
29. Eden, E., et al., *GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists*. BMC Bioinformatics, 2009. **10**(1): p. 48.
30. Warde-Farley, D., et al., *The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function*. Nucleic Acids Research, 2010. **38**(suppl\_2): p. W214-W220.
31. Fernandes, J.C.R., et al., *Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease*. Non-coding RNA, 2019. **5**(1): p. 17.
32. Wu, P., et al., *Roles of long noncoding RNAs in brain development, functional diversification and neurodegenerative diseases*. Brain Research Bulletin, 2013. **97**: p. 69-80.
33. Kitagawa, M., et al., *Cell cycle regulation by long non-coding RNAs*. Cellular and Molecular Life Sciences, 2013. **70**(24): p. 4785-4794.
34. Hu, G., et al., *Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation*. Nature immunology, 2013. **14**(11): p. 1190-1198.
35. Cabili, M.N., et al., *Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution*. Genome Biology, 2015. **16**(1): p. 20.
36. Rinn, J.L. and H.Y. Chang, *Genome regulation by long noncoding RNAs*. Annual review of biochemistry, 2012. **81**: p. 145-166.
37. Chen, Y.G., A.T. Satpathy, and H.Y. Chang, *Gene regulation in the immune system by long noncoding RNAs*. Nature Immunology, 2017. **18**(9): p. 962-972.
38. Hutchinson, J.N., et al., *A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains*. BMC genomics, 2007. **8**: p. 39-39.
39. Gong, C. and L.E. Maquat, *lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements*. Nature, 2011. **470**(7333): p. 284-288.

40. Swann, J.B. and M.J. Smyth, *Immune surveillance of tumors*. The Journal of clinical investigation, 2007. **117**(5): p. 1137-1146.
41. Hanahan, D. and Robert A. Weinberg, *Hallmarks of Cancer: The Next Generation*. Cell, 2011. **144**(5): p. 646-674.
42. Zhang, N. and Michael J. Bevan, *CD8+ T Cells: Foot Soldiers of the Immune System*. Immunity, 2011. **35**(2): p. 161-168.
43. van der Leun, A.M., D.S. Thommen, and T.N. Schumacher, *CD8+ T cell states in human cancer: insights from single-cell analysis*. Nature Reviews Cancer, 2020. **20**(4): p. 218-232.
44. Farhood, B., M. Najafi, and K. Mortezaee, *CD8+ cytotoxic T lymphocytes in cancer immunotherapy: A review*. Journal of Cellular Physiology, 2019. **234**(6): p. 8509-8521.
45. Schumacher Ton, N. and D. Schreiber Robert, *Neoantigens in cancer immunotherapy*. Science, 2015. **348**(6230): p. 69-74.
46. Koury, J., et al., *Immunotherapies: Exploiting the Immune System for Cancer Treatment*. Journal of immunology research, 2018. **2018**: p. 9585614-9585614.
47. Pardoll, D.M., *The blockade of immune checkpoints in cancer immunotherapy*. Nature Reviews Cancer, 2012. **12**(4): p. 252-264.
48. Brahmer, J.R., et al., *Phase I Study of Single-Agent Anti-Programmed Death-1 (MDX-1106) in Refractory Solid Tumors: Safety, Clinical Activity, Pharmacodynamics, and Immunologic Correlates*. Journal of Clinical Oncology, 2010. **28**(19): p. 3167-3175.
49. Kvistborg, P., et al., *Anti-CTLA-4 therapy broadens the melanoma-reactive CD8+ T cell response*. Science Translational Medicine, 2014. **6**(254): p. 254ra128-254ra128.
50. Larkin, J., et al., *Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma*. New England Journal of Medicine, 2015. **373**(1): p. 23-34.
51. Rotte, A., *Combination of CTLA-4 and PD-1 blockers for treatment of cancer*. Journal of Experimental & Clinical Cancer Research, 2019. **38**(1): p. 255.
52. Wei, S.C., et al., *Combination anti-CTLA-4 plus anti-PD-1 checkpoint blockade utilizes cellular mechanisms partially distinct from monotherapies*. Proceedings of the National Academy of Sciences, 2019. **116**(45): p. 22699.
53. Kelderman, S., T.N.M. Schumacher, and J.B.A.G. Haanen, *Acquired and intrinsic resistance in cancer immunotherapy*. Molecular Oncology, 2014. **8**(6): p. 1132-1139.
54. Bajwa, R., et al., *Adverse Effects of Immune Checkpoint Inhibitors (Programmed Death-1 Inhibitors and Cytotoxic T-Lymphocyte-Associated Protein-4 Inhibitors): Results of a Retrospective Study*. Journal of clinical medicine research, 2019. **11**(4): p. 225-236.
55. Wang, Q., et al., *Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment*. Cancer Cell, 2017. **32**(1): p. 42-56.e6.
56. Charoentong, P., et al., *Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade*. Cell Reports, 2017. **18**(1): p. 248-262.
57. Huarte, M., *The emerging role of lncRNAs in cancer*. Nature Medicine, 2015. **21**(11): p. 1253-1261.
58. Schmitt, A.M. and H.Y. Chang, *Long Noncoding RNAs in Cancer Pathways*. Cancer Cell, 2016. **29**(4): p. 452-463.
59. Jiang, M., et al., *Self-Recognition of an Inducible Host lncRNA by RIG-I Feedback Restricts Innate Immune Response*. Cell, 2018. **173**(4): p. 906-919.e13.

60. Xie, Q., et al., *Long Noncoding RNA ITPRIP-1 Positively Regulates the Innate Immune Response through Promotion of Oligomerization and Activation of MDA5*. *Journal of Virology*, 2018. **92**(17): p. e00507-18.
61. Wang, P., et al., *The STAT3-Binding Long Noncoding RNA Inc-DC Controls Human Dendritic Cell Differentiation*. *Science*, 2014. **344**(6181): p. 310-313.
62. Huang, D., et al., *NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death*. *Nature Immunology*, 2018. **19**(10): p. 1112-1125.
63. Guo, W., et al., *LincRNA-immunity landscape analysis identifies EPIC1 as a regulator of tumor immune evasion and immunotherapy resistance*. *Science Advances*, 2021. **7**(7): p. eabb3555.
64. Joung, J., et al., *Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening*. *Nature Protocols*, 2017. **12**(4): p. 828-863.
65. Cabili, M.N., et al., *Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses*. *Genes Dev.*, 2011. **25**(18): p. 1915-1927.
66. Heinz, S., et al., *Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities*. *Molecular Cell*, 2010. **38**(4): p. 576-589.
67. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0*. *Bioinformatics*, 2011. **27**(12): p. 1739-1740.
68. Thorsson, V., et al., *The Immune Landscape of Cancer*. *Immunity*, 2018. **48**(4): p. 812-830.e14.
69. Deng, J., et al., *ULK1 inhibition overcomes compromised antigen presentation and restores antitumor immunity in LKB1-mutant lung cancer*. *Nature Cancer*, 2021. **2**(5): p. 503-514.
70. Consortium, G.T., et al., *Genetic effects on gene expression across human tissues*. *Nature*, 2017. **550**(7675): p. 204-213.
71. Lonsdale, J., et al., *The Genotype-Tissue Expression (GTEx) project*. *Nature Genetics*, 2013. **45**(6): p. 580-585.
72. Cancer Genome Atlas Research, N., et al., *The Cancer Genome Atlas Pan-Cancer analysis project*. *Nature genetics*, 2013. **45**(10): p. 1113-1120.
73. Chang, K., et al., *The Cancer Genome Atlas Pan-Cancer analysis project*. *Nature Genetics*, 2013. **45**(10): p. 1113-1120.
74. Hugo, W., et al., *Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma*. *Cell*, 2016. **165**(1): p. 35-44.
75. Clarke, C., et al., *Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis*. *Carcinogenesis*, 2013. **34**(10): p. 2300-2308.
76. Desmedt, C., et al., *The Gene expression Grade Index: a potential predictor of relapse for endocrine-treated breast cancer patients in the BIG 1–98 trial*. *BMC Medical Genomics*, 2009. **2**(1): p. 40.
77. Kao, K.-J., et al., *Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization*. *BMC Cancer*, 2011. **11**(1): p. 143.
78. Dedeurwaerder, S., et al., *DNA methylation profiling reveals a predominant immune component in breast cancers*. *EMBO Molecular Medicine*, 2011. **3**(12): p. 726-741.
79. Li, T., et al., *TIMER2.0 for analysis of tumor-infiltrating immune cells*. *Nucleic Acids Research*, 2020. **48**(W1): p. W509-W514.

80. Ransohoff, J.D., Y. Wei, and P.A. Khavari, *The functions and unique features of long intergenic non-coding RNA*. Nature reviews. Molecular cell biology, 2018. **19**(3): p. 143-157.
81. Chen, J., et al., *Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs*. Genome biology, 2016. **17**: p. 19-19.