Single Cell RNA Sequencing of Lung Adenocarcinoma Reveals Heterogeneity of Immune Response-related Genes

Ke-Yue Ma[1], Alexandra A. Schonnesen[2], Amy Brock[1,2], Carla Van Den Berg[1,3,6], S. Gail Eckhardt[3], Zhihua Liu[4,5], Ning Jiang[1,2,3*]

[1]Institute for Cellular and Molecular Biology, College of Natural Sciences, The University of Texas at Austin, Austin, Texas, USA.

[2]Department of Biomedical Engineering, Cockrell School of Engineering, The University of Texas at Austin, Austin, Texas, USA.

[3]Department of Oncology, LIVE**STRONG** Cancer Institutes, Dell Medical School, The University of Texas at Austin, Austin, TX, USA

[4]Department of AnoRectal Surgery, the Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, 510799, China

[5]Department of Center Laboratory, the Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, 510799, China

[6]Division of Pharmacology/Toxicology, College of Pharmacy, The University of Texas at Austin, Austin, Texas, USA.

[*]Corresponding author

Correspondence:

Ning Jiang, Ph.D.

Department of Biomedical Engineering

Cockrell School of Engineering

The University of Texas at Austin

107 W Dean Keeton Street, room 1.108C

Austin, TX 78712

jiang@austin.utexas.edu

512-471-4860

Conflict of interest:

NJ is a scientific advisor of ImmuDX, LLC and Immune Arch, Inc.

ABSTRACT

Immunotherapy has emerged as a promising approach to treat cancer. However, partial responses across multiple clinical trials support the significance of characterizing inter- and intra-tumoral heterogeneity to achieve better clinical results and as potential tools in selecting patients for different types of cancer immunotherapies. Yet, the type of heterogeneity that informs clinical outcome and/or patient selection has not been fully explored. In particular, the lack of characterization of immune response-related genes in cancer cells hinders the further development of metrics to select and optimize immunotherapy. Therefore, we analyzed single cell RNA-seq data from lung adenocarcinoma patients and cell lines to characterize the intra-tumoral heterogeneity of immune response-related genes and demonstrated their potential impact on the efficacy of immunotherapy. We discovered that IFNγ signaling pathway genes are heterogeneously expressed and co-regulated with other genes in single cancer cells, including MHC class II (*MHCII*) genes. The downregulation of genes in IFNγ signaling pathways in cell lines corresponds to an acquired resistance phenotype. Moreover, analysis of two groups of tumor-restricted antigens, neoantigens and cancer testis antigens, revealed heterogeneity in their expression in single cells. These analyses provide a rationale for applying multi-antigen combinatorial therapies to prevent tumor escape and establish a basis for future development of prognostic metrics based on intra-tumoral heterogeneity.

INTRODUCTION

The recent decade has witnessed the advent and significant advancement of immunotherapy as an effective anticancer strategy. Its demonstrated efficacy against multiple cancer types has attracted more attention to predict the outcomes of various immunotherapies alone or in combination with other therapeutic modalities. One of the most promising approaches to activate immune responses is the immune checkpoint blockade, such as anti-CTLA4 and anti-PD-1 that block inhibitory molecules on immune cells to unleash anti-tumor immunity. Moreover, recent studies have begun shedding light on the role of IFNγ pathway genes on immune checkpoint blockade therapy, demonstrating the effective anti-tumor immune response induced by IFNγ when recognizing its cognate receptor on cancer cells or antigen presenting cells (1, 2). Another immunotherapy approach gaining more interest is targeting neoantigens or cancer testis antigens (CTAs), whose expressions are cancer cell specific. Both therapeutic vaccines as well as T cell receptor re-directed adoptive cell transfer therapy have been demonstrated to boost T cell responses against tumors expressing these antigenic targets. Several such treatments have entered clinical trials (3-5). However, many of these monotherapies are effective in only a fraction of patients. Although, studies have illustrated the inter-tumoral heterogeneity (between patients) of immune signatures (6, 7), as well as intra-tumoral heterogeneity (within the tumor) of immune cells in multiple cancer types (8-10), it remains elusive how the underlying intra-tumoral heterogeneity of immune response-related gene expression in tumor cells will impact responses and ability to predict outcome. In the current study, we demonstrate that single cell RNA-sequencing (scRNA-seq) of tumor cells can be used to identify such intra-tumoral heterogeneity.

Lung cancer is one of the most highly mutated cancer types (11) and despite the improved success of immunotherapies in lung cancer, a low response rate (≤20%) is still observed (12). Herein, we used previously published scRNA-seq data from lung adenocarcinoma (LUAD) and

two LUAD cell lines, LC2/ad-R (vandetanib resistant) and LC2/ad (vandetanib susceptible) (13), as a test set to demonstrate the functional intra-tumoral heterogeneity of immune-response related genes that might impact immunotherapy responses (13-15). We found that *MHCII* genes are heterogeneously expressed among tumor cells obtained from LUAD patients and their expression correlates with a favorable prognosis. Interestingly, *MHCII* genes are heterogeneously expressed within single cells from individual patients. *MHCII* genes can be induced by IFNγ (16). We then sought to identify the intra-tumoral heterogeneity of IFNγ signaling pathway and observed co-expression of IFNγ signaling pathway genes in a fraction of lung adenocarcinoma single cells that had a higher level of *MHCII* expression. Similar results were found to be enriched in LC2/ad cell line. Further analysis showed that the opposite trend, where discoordinated expression of IFNγ signaling pathway genes was associated with a lower level of *MHCII* expression, was enriched in LC2/ad-R cell line that acquired a vandetanib resistance phenotype. This relationship between IFNγ signaling pathway genes and *MHCII* genes could also be important in determining resistance to immunotherapy in lung adenocarcinoma. We also uncovered heterogeneity in the expression of predicted cancer neoantigens and CTAs in single cells from both LUAD patients and cell lines. Interestingly, the decrease in the number of neoantigens is also correlated with the acquired resistance phenotype. Our study suggests that using combinatorial strategy to target multiple tumor antigens in select patients could improve immunotherapy efficacy.

RESULTS

Prognostic prediction of lung adenocarcinoma by the expression pattern of cell cycle and *MHCII* genes

Identifying patients at higher risk of tumor progression or recurrence is crucial for making individualized treatment plans. Despite recently recognized intra-tumoral heterogeneity, there is a lack of understanding of how this is associated with prognosis. In this study, we aimed to characterize the heterogeneity of prognostic predictors in single cancer cells. We first identified pathways that are of potential prognostic predictors in LUAD cohorts from The Cancer Genome Atlas Research Network (TCGA) (**Supplementary Table S1**). We found that the top pathways associated with an unfavorable prognosis are enriched for cell cycle related pathways, while the top pathways associated with a favorable prognosis are enriched for T cell signaling related pathways (**Figure 1A, Supplementary Table S1**). Interestingly, the common genes shared by the top favorable prognostic pathways are *MHCII* genes. Further survival analysis validated the association of upregulated *MHCII* genes with a better overall survival rate (**Figure 1B, Supplementary Table S2**). Surprisingly, *MHCI* genes do not have a significant association with overall survival in LUAD (**Supplementary Table S2**). Compared to the expression of MHC genes in normal tissues, *MHCII* genes are more downregulated compared with that of *MHCI* (**Figure 1C**). Previously, it has been shown that higher *MHCII* expression was also associated with better prognosis in multiple other tumor types, such as melanoma and triple negative breast cancer (17, 18). Especially in melanoma patients, the expression of *MHCII* can predict response to anti-PD-1/anti-PD-L1 therapy (18).

Single cancer cells express distinct prognosis-associated gene modules

Next, we assessed the expression level of prognosis-associated genes, discovered from analyzing bulk cancer sample RNA-seq data, in LUAD patients at single cell level (**Supplementary Table S2**). We reanalyzed previously reported scRNA-seq data from LUAD

patient tumor xenografts (PDXs) (14, 15). The LC-PT-45 tumor was taken from a treatment-naïve 60-year-old male patient with an irregular primary lung lesion, while the LC-MBT-15 tumor was taken from a 57-year-old female with a metachronous brain metastasis after standard chemotherapy and erlotinib treatment. A total of 77 and 49 single cells were sequenced for LC-PT-45 and LC-MBT-15, respectively. An average of 5 million reads were sequenced for each single cell, which is needed to saturate the mutation detection (**Supplementary Figure S1A**). Additional quality control was performed to remove low quality single cells, which account for a small percentage, based on sequencing metrics, including genome mapping percentage, multi-mapped read percentage, mitochondrial DNA mapping percentage, cell-to-mean correlation and transcriptome variance (**Supplementary Figure S1B**), following a published method (19). High quality single cells were used in following analyses.

Min et al. previously reported the distinct subpopulations of LUAD single cells with respect to the expression of cell cycle genes (15). Here, we further sought to determine whether prognosis-associated genes/pathways, including cell cycle genes and antigen presenting pathway genes, are heterogeneously expressed in single LUAD cells and whether there are any expression patterns among these genes. We applied a self-organizing map (SOM), which adopts unsupervised machine learning approach to map genes into coordinately expressed groups (metagenes) (20). A second-level SOM was then processed by mapping all samples together into a two-dimensional mosaic pattern based on metagene expression. K-means clustering grouped metagenes into six clusters based on co-expression in scRNA-seq from LC-PT-45 patient (**Figure 2A**). Gene set enrichment analysis on genes in these six clusters showed that Cluster A was enriched in cell cycle genes, while Cluster B was enriched for antigen presentation pathways, specifically *MHCII* presentation (**Figure 2A, Supplementary Table S3**). We then observed that cells were separated into three groups based on their distinct metagene expression patterns: cell cycle pathway high (I), antigen presentation pathway high (II), and both

pathways low (III) (**Figure 2B**). We also analyzed the PDX tumor cells from the other LUAD patient (LC-MBT-15). As previously reported, LC-MBT-15 exhibits a less heterogeneous transcriptional profile (14). Nevertheless, SOM analysis revealed that metagenes representing cell cycle and *MHCII* genes were mapped into different metagene modules (**Supplementary Figure S2, Supplementary Table S3**).

We further applied above analysis on scRNA-seq data from two human LUAD cell lines, LC2/ad-R and LC2/ad. LC2/ad-R is a subclone of LC2/ad that has acquired resistance to vandetanib (13). Similar to LUAD patient samples, metagenes representing cell cycle and *MHCII* genes were also mapped into different metagene modules in the two cell line data (**Figure 2C, Supplementary Table S3**). Additionally, single cells from two cell lines were also separated into three groups: cell cycle pathway high (I), antigen presentation pathway high (II), and both pathways low (III) (**Figure 2D**). We noticed that more single cells from LC2/ad cell line belong to group II (51 out of 89) compared to single cells from LC2/ad-R cell line (1 out of 70) ($p < 10^{-5}$, Fisher's exact test), which suggests that a significant downregulation of antigen presentation pathway in LC2/ad-R cell line could render this cell line resistance to immunotherapy in additional to vandetanib.

Heterogeneity of IFNγ stimulated genes in single cells

Since it is known that *MHCII* genes, a class of genes in antigen presentation pathway, can be regulated through IFNγ signaling pathway (16) and we showed that antigen presentation pathway genes are heterogeneously expressed by LUAD cells, we next examined the expression pattern of IFNγ signaling pathway genes. Recent works also began to uncover the role of IFNγ signaling pathway on cancer immune checkpoint blockade therapies. Gao et al. identified the genomic variations, such as copy number variations and single-nucleotide polymorphisms of IFNγ pathway genes in melanoma patients as a resistance mechanism to

anti-CTLA4 therapy (1). By contrast, Benci et al. suggested an IFN-driven PD-L1-independent resistance to immune checkpoint blockade (21). These different conclusions on the roles that IFNγ signaling pathway plays in cancer resistance to immunotherapy could be due to distinct clinical trial cohorts as well as the underlying intra-tumoral heterogeneity of IFNγ signaling pathway genes. However, it is not known whether this cancer-immunity interaction network exhibits differential gene expression at the single cell level, which could be an important prognostic factor and shed light on the mechanism of cancer immune resistance.

Here, we used the LUAD and cell line scRNA-seq data sets to examine the intra-tumoral heterogeneity of IFNγ signaling pathway. IFNγ signaling pathway genes were curated from Gene Ontology Database and Reactome Pathway Database. Single cells from LUAD patients were first clustered based on expression of *MHCII* genes into two groups: *MHCII* low and *MHCII* high (**Figure 3A, Supplementary Figure S3A**). Gene Set Variation Analysis (GSVA) score of IFNγ signaling pathway, which indicates the extent of coordinated expression among pathway genes, was then calculated for each individual cell. We found a significant decrease of IFNγ signaling pathway GSVA scores in *MHCII* low group (**Figure 3B, Supplementary Figure S3B**). We also investigated expression of IFNγ signaling pathway genes in LC2/ad-R and LC2/ad cell lines. GSVA analysis showed that significantly more single cells in LC2/ad-R cell line downregulated IFNγ signaling pathway genes compared to that in LC2/ad cell line (**Figure 3C,D, Supplementary Figure S4**).

The different pattern we observed on the coordinated expression of IFNγ signaling pathway genes is consistent with their role in directing anti-proliferative and pro-apoptotic effects on tumor cells (2). However, their roles in activating *MHCII* expression, enhancing tumor antigen presentation, and inducing the recruitment of other immune cells suggest that the lack of coordinated expression of IFNγ signaling pathway genes in sub-set of cancer cells within a

tumor would render these cells resistance to immunotherapy in addition to small molecule inhibitors (22). Further studies are necessary to demonstrate how the heterogeneity of IFNγ signaling pathway expression influences the efficacy of immunotherapy.

Heterogeneous expression of predicted cancer neoantigens in single cells

Neoantigens are a group of promising targets to induce anti-tumor immunity through recognition of neoantigen-specific T cells and are advantageous in their selective expression in cancer cells without the risk of causing autoimmunity. It has been proposed that targeting multiple neoantigens simultaneously will likely be important to prevent tumor escape by editing of the mutated epitope (5, 11). Here we aimed to analyze whether or not neoantigens are heterogeneously expressed. If so, then this would warrant further investigation of a new strategy in cancer immunotherapy where a combination of neoantigens could be administered as therapeutic vaccines or a combination of neoantigen specific T cell receptors could be used to modify patients' own T cells in adoptive cell transfer therapy. Consequently, the analysis of the heterogeneity of neoantigen expression in single cancer cells would facilitate the development of these therapeutic regimens.

To answer this question, somatic mutations in each single cell were further assessed for neoantigen prediction (**See methods**). Only neoantigens detected in more than three cells were selected. We found more than half of the neoantigens exhibited a bimodal expression (**Figure 4A and Supplementary Figure S5**), suggesting the possibility of tumor escape with single neoantigen epitope based therapy. Surprisingly, we not only found the same bimodal expression of neoantigens in LC2/ad-R and LC2/ad cell lines, but also revealed that LC2/ad-R cell lines have a significant decrease of neoantigen load compared with the non-resistant parental cell line (**Supplementary Figure S6A, B**). This suggested the degree of neoantigen load might affect cancer cell drug-responses. In addition to detecting neoantigens, we also

identified both wild-type alleles and/or variant containing alleles (single nucleotide variants (SNV) and insertion/deletion (INDEL)) for many genes in many single cells, indicating that cells without neoantigen identified is not mainly due to drop-out of corresponding genes (**Supplementary Figure S7**).

Heterogeneous expression of cancer testis antigen in single cells

CTAs are a group of tumor antigens with normal expression restricted to male germ cells in the testis. In cancer, alteration of gene regulation results in aberrant expression of CTAs in various tumor types (23). Previous studies have demonstrated the extensive heterogeneity of CTAs among patients of different cancer types (6, 7). However, little is known about the heterogeneity of all possible CTAs expressed within each patient. Thus, we examined the expression of all possible CTAs in LUAD single cells and demonstrated the extensive heterogeneity of their expression at the single cell level. To restrict our analysis on only CTAs expressed in tumors, we selected a subset of 276 known CTAs that are transcriptionally silent in normal non-germline tissues based on Genotype-Tissue Expression (GTEx) data (7). Only CTAs with normalized counts > 0 in more than 2 cells were selected. In both LUAD patients, we observed both significant inter-tumoral heterogeneity and inter-tumoral heterogeneity of expressed CTAs (**Figure 4B**). In addition, compared with neoantigens, CTAs exhibit a lower expression level (**Figure 4 and Supplementary Figure S5**). Further analysis of CTAs in cell lines revealed that LC2/ad-R and LC2/ad cell lines can also be separated based on CTA expression. Expression of MAGEA6 and MAGEA2 is significantly higher in LC2/ad-R compared with LC2/ad (**Supplementary Figure. S6C**). These results suggest that CTAs could be therapeutic targets for cancers that are resistant to small molecule inhibitor therapy.

DISCUSSION

In this study, we applied both LUAD and cell line scRNA-seq to characterize the intra-tumoral heterogeneity of immune response-related genes. We identified that (1) prognosis-related genes, especially cell cycle pathway and antigen presentation pathway genes, are independently co-expressed among single cells; (2) IFNγ signaling pathways are heterogeneously expressed within cancer cells and downregulation is correlated to a drug-resistant phenotype; and (3) promising cancer vaccination targets, such as neoantigens and CTAs, are also heterogeneously expressed.

Previous studies showed that IFNγ induced the expression of *MHCII* genes in multiple myeloma cells (24) and normal epidermal melanocytes (25). Since *MHCII* expression on cancer cells is important for CD4+ T cell immunity as well as T cell exhaustion (26), our findings on the heterogeneity expression in cancer cells, especially its lost coordinated expression with IFNγ signaling pathway genes in LC2/ad-R cell line, could have important implications in cancer immunotherapy. Although it is yet to be determined whether cell cycle pathways and *MHCII* genes in tumor cells are mechanistically associated, recent findings that CDK4/6 inhibitors not only induce cell cycle arrest, but also promote anti-tumor immunity through activation of interferon signaling and suppression of regulatory T cells (Treg), indicates a connection between these two gene modules (27). Besides, recent study investigating resistance to immune checkpoint blockade in melanoma revealed a cancer resistance program, which is enriched for upregulation of E2F targets and downregulation of antigen presentation, is associated with T cell exclusion (28). While it is known that *MHCII* expression can be induced by IFNγ (16), the analysis of immune response-related genes in single cancer cells revealed that genes of the IFNγ signaling pathway are co-expressed, including many IFNγ stimulated genes in addition to *MHCII*. One of the co-expressed IFNγ stimulated genes, IDO1, is responsible for the conversion of tryptophan and other indole derivatives to kynurenine and has

been widely studied as an important suppressor of anti-tumor immunity. Several IDO inhibitors have entered clinical trials as monotherapies and in combination with CTLA-4 and PD-1 immune checkpoint blockade (29). Our analysis also indicated that heterogeneity in IFNγ signaling pathways might impact the responses of IFNγ-pathway-directed therapies.

Another aspect of intratumoral heterogeneity unveiled by our analysis is in cancer vaccine targets, including neoantigens and CTAs. Many completed clinical trials have failed to observe significant responses to CTA vaccines. For example, clinical trials in non-small cell lung carcinoma, identified no significant difference in melanoma-associated antigen-A3 (MAGE-A3) compared to control groups (30). Yet another recent clinical trial study in melanoma patients reported 4/6 patients had no recurrence after neoantigen vaccination (3). Our study further demonstrates the individuality and intra-tumoral heterogeneity of neoantigens and CTAs. The analysis of heterogeneous expression of neoantigens and CTAs together highlights the challenge of cancer vaccine mono-epitope therapy to elicit effective anti-tumor immunity and supports the possible advantages of targeting multiple epitopes in neoantigen vaccine or neoantigen specific T cell based immunotherapy.

In summary, we demonstrated that the intra-tumoral heterogeneity of immune module expression will help to develop better prognosis of immunotherapies, Examining the gene expression in single cancer cells not only provides a rationale for combinatorial immunotherapies, in particular neoantigen/CTA-directed therapies, but also paves the road for future analyses on how intra-tumoral heterogeneity impacts immunotherapy efficacy.

MATERIAL AND METHODS

Single-cell RNA-seq Dataset

Raw reads of scRNA-seq and whole exome sequencing were all downloaded from NCBI GEO DataSets website under accession number GSE69405 (https://www.ncbi.nlm.nih.gov/geo). Two PDX samples were processed for scRNA-seq, LC-PT-45 and LC-MBT-15 (14). Raw reads of scRNA-seq data from LC2/ad-R and LC2/ad cell lines were downloaded from the DNA Data Bank of Japan under accession DRA001287. Raw reads of scRNA-seq were first mapped to human genome reference GRCh37 with RSEM (31) and then normalized using SCnorm (32). Genes with read counts fewer than 2 in more than 10% of all single cells are filtered out.

TCGA Dataset

Quantile normalized read counts of RNA-seq from LUAD patients and corresponding clinical data were downloaded from The Broad Institute TCGA GDAC Firehose website (http://gdac.broadinstitute.org/). Aggregated somatic mutation data processed by MuTect2 was downloaded from NCI GDC Data Portal (https://portal.gdc.cancer.gov/).

Prognostic Gene and Pathway Analysis of Lung Adenocarcinoma

Each gene in the TCGA dataset was classified into “high-expression” and “low-expression” by comparing to its mean expression. Then R package “survival” was used to calculate the Cox Proportional Hazards regression hazard ratio (HR) and p-value of log-rank test. R package “survminer” was used to plot Kaplan-Meier plots. Adjusted p-value of log-rank test was applied using FDR correction method. R package “GSA” was used to identify pathways that are potential prognosis predictor of overall survival in LUAD patients (33). Curated canonical pathways were downloaded from Molecular Signatures Database (MSigDB) v6.1 (http://software.broadinstitute.org/gsea/msigdb/index.jsp).

Somatic Mutation detection

Mapping of whole exome sequencing reads and preprocessing of mappable reads were processed as described previously (14). Then somatic Mutations in both PDX samples were called using MuTect2 with default settings.

Somatic Mutations of LC2/ad cell lines were curated from Catalogue Of Somatic Mutations In Cancer (COSMIC) database. Over 97% of somatic mutations are overlapping between LC2/ad-R and LC2/ad cell lines (13).

Self Organization Map Analysis on scRNA-seq

Prognostic genes with FDR < 0.05 were selected and genes with mean normalized counts < 2 were filtered. Gene expression data were then log transformed, centralized and clustered using self-organizing map (SOM). Genes were clustered onto a 12x12 grid or 15x15 grid for LUAD scRNA-seq and cell line scRNA-seq, respectively. R package “oposSOM” was implemented for SOM processing and downstream analysis, including k-means clustering of metagenes and second-level SOM (20).

Supervised Analysis of IFNγ signaling pathways in scRNA-seq

IFNγ signaling pathways genes were curated from Gene Ontology and Reactome pathway database, under GO:0034341 and R-HAS-877300.1, respectively. GSVA was used to calculate sample-wise gene set enrichment score of IFNγ signaling pathways in each individual single cells (34). K-means clustering was used to group single cells into *MHCII* low and *MHCII* high groups based on *MHCII* gene expressions. Non-parametric Wilcoxon test was used to perform significant test of GSVA scores between different groups. R package “pheatmap” was used to generate heatmaps.

Tumor-Specific HLA Typing, HLA-Binding Neoepitope Prediction and Expression in scRNA-seq

Raw reads of whole exome sequencing were processed with OptiType (35). Then VCF files of each individual single cell generated by GATK HaplotypeCaller and MHC class I alleles (HLA-A, HLA-B and HLA-C) predicted by OptiType were used to predict neoepitope with topiary (https://github.com/hammerlab/topiary). Netmhcpan were selected in topiary as the MHC binding predictor. Cells were included if a neoantigen is detected regardless of its corresponding wildtype antigen detection status. Only genes that have predicted neoepitopes in more than three cells were examined for expression.

Cancer Testis Antigen Expression in scRNA-seq

CTAs genes are selected as previously reported (7). Briefly, known CTAs with negligible expression in GTEx normal tissues (95^{th} percentile value < 1 normalized counts in all somatic tissue types) are analyzed for scRNA-seq data.

Statistics

All p-values are false discovery rate (FDR) corrected and FDR < 0.05 is treated as significant.

AUTHOR CONTRIBUTIONS

K-Y M. wrote the scripts, performed analysis, interpreted results, and wrote the manuscript; A.A.S. helped with SNV analysis; A.B. helped interpret the results; C.V.D.B. helped interpret the results, S.G.E. helped designed the study; Z.L. helped designed study; N.J. designed study, interpreted results, and wrote the manuscript with K-Y M. with input from all authors.

ACKNOLEDGEMENTS

This work was supported by NIH grants R00AG040149 (N.J.), NSF CAREER Award 1653866 (N.J.), the Welch Foundation grant F1785 (N.J.), and National Natural Science Foundation of China grants 1147222 and 11672246 (H.Y). N.J. is a Cancer Prevention and Research Institute of Texas (CPRIT) Scholar and a Damon Runyon-Rachleff Innovator. SGE is also a CPRIT Scholar.

REFERENCES

1. Gao J, Shi LZ, Zhao H, Chen J, Xiong L, He Q, et al. Loss of IFN-gamma Pathway Genes in Tumor Cells as a Mechanism of Resistance to Anti-CTLA-4 Therapy. *Cell.* 2016;167(2):397-404 e9.
2. Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, Hu-Lieskovan S, et al. Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N Engl J Med.* 2016;375(9):819-29.
3. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature.* 2017;547(7662):217-21.
4. Robbins PF, Kassim SH, Tran TL, Crystal JS, Morgan RA, Feldman SA, et al. A pilot trial using lymphocytes genetically engineered with an NY-ESO-1-reactive T-cell receptor: long-term follow-up and correlates with response. *Clin Cancer Res.* 2015;21(5):1019-27.
5. Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia L, et al. T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N Engl J Med.* 2016;375(23):2255-62.
6. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 2016;17(1):174.
7. Rooney MS, Shukla SA, Wu CJ, Getz G, and Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 2015;160(1-2):48-61.
8. Chevrier S, Levine JH, Zanotelli VRT, Silina K, Schulz D, Bacac M, et al. An Immune Atlas of Clear Cell Renal Cell Carcinoma. *Cell.* 2017;169(4):736-49 e18.
9. Lavin Y, Kobayashi S, Leader A, Amir ED, Elefant N, Bigenwald C, et al. Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. *Cell.* 2017;169(4):750-65 e17.

10. Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell.* 2017;169(7):1342-56 e16.
11. Schumacher TN, and Schreiber RD. Neoantigens in cancer immunotherapy. *Science.* 2015;348(6230):69-74.
12. Nishino M, Ramaiya NH, Hatabu H, and Hodi FS. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol.* 2017;14(11):655-68.
13. Suzuki A, Matsushima K, Makinoshima H, Sugano S, Kohno T, Tsuchihara K, et al. Single-cell analysis of lung adenocarcinoma cell lines reveals diverse expression patterns of individual cells invoked by a molecular target drug treatment. *Genome Biology.* 2015;16(1):66.
14. Kim KT, Lee HW, Lee HO, Kim SC, Seo YJ, Chung W, et al. Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. *Genome Biol.* 2015;16:127.
15. Min JW, Kim WJ, Han JA, Jung YJ, Kim KT, Park WY, et al. Identification of Distinct Tumor Subpopulations in Lung Adenocarcinoma via Single-Cell RNA-seq. *PLoS One.* 2015;10(8):e0135817.
16. Steimle V, Siegrist CA, Mottet A, Lisowska-Grospierre B, and Mach B. Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. *Science.* 1994;265(5168):106-9.
17. Forero A, Li Y, Chen D, Grizzle WE, Updike KL, Merz ND, et al. Expression of the MHC Class II Pathway in Triple-Negative Breast Cancer Tumor Cells Is Associated with a Good Prognosis and Infiltrating Lymphocytes. *Cancer Immunol Res.* 2016;4(5):390-9.

18. Johnson DB, Estrada MV, Salgado R, Sanchez V, Doxie DB, Opalenik SR, et al. Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nat Commun.* 2016;7:10582.

19. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* 2016;17:29.

20. Loffler-Wirth H, Kalcher M, and Binder H. oposSOM: R-package for high-dimensional portraying of genome-wide expression landscapes on bioconductor. *Bioinformatics.* 2015;31(19):3225-7.

21. Benci JL, Xu B, Qiu Y, Wu TJ, Dada H, Twyman-Saint Victor C, et al. Tumor Interferon Signaling Regulates a Multigenic Resistance Program to Immune Checkpoint Blockade. *Cell.* 2016;167(6):1540-54 e12.

22. Castro F, Cardoso AP, Gonçalves RM, Serre K, and Oliveira MJ. Interferon-Gamma at the Crossroads of Tumor Immune Surveillance or Evasion. *Frontiers in Immunology.* 2018;9:847.

23. Scanlan MJ, Gure AO, Jungbluth AA, Old LJ, and Chen YT. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. *Immunol Rev.* 2002;188:22-32.

24. Zhao M, Flynt FL, Hong M, Chen H, Gilbert CA, Briley NT, et al. MHC class II transactivator (CIITA) expression is upregulated in multiple myeloma cells by IFN-γ. *Molecular Immunology.* 2007;44(11):2923-32.

25. Tsujisaki M, Igarashi M, Sakaguchi K, Eisinger M, Herlyn M, and Ferrone S. Immunochemical and functional analysis of HLA class II antigens induced by recombinant immune interferon on normal epidermal melanocytes. *J Immunol.* 1987;138(4):1310-6.

26. Maruhashi T, Okazaki I-m, Sugiura D, Takahashi S, Maeda TK, Shimizu K, et al. LAG-3 inhibits the activation of CD4+ T cells that recognize stable pMHCII through its

conformation-dependent recognition of pMHCII. *Nature Immunology.* 2018;19(12):1415-26.

27. Goel S, DeCristo MJ, Watt AC, BrinJones H, Sceneay J, Li BB, et al. CDK4/6 inhibition triggers anti-tumour immunity. *Nature.* 2017;548(7668):471-5.

28. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, et al. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell.* 2018;175(4):984-97 e24.

29. Ott PA, Hodi FS, Kaufman HL, Wigginton JM, and Wolchok JD. Combination immunotherapy: a road map. *J Immunother Cancer.* 2017;5:16.

30. Vansteenkiste JF, Cho BC, Vanakesa T, De Pas T, Zielinski M, Kim MS, et al. Efficacy of the MAGE-A3 cancer immunotherapeutic as adjuvant therapy in patients with resected MAGE-A3-positive non-small-cell lung cancer (MAGRIT): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* 2016;17(6):822-35.

31. Li B, and Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.

32. Bacher R, Chu LF, Leng N, Gasch AP, Thomson JA, Stewart RM, et al. SCnorm: robust normalization of single-cell RNA-seq data. *Nat Methods.* 2017;14(6):584-6.

33. Efron B, and Tibshirani R. On Testing the Significance of Sets of Genes. *Ann Appl Stat.* 2007;1(1):107-29.

34. Hanzelmann S, Castelo R, and Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.

35. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, and Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics.* 2014;30(23):3310-6.

FIGURES AND FIGURE LEGENDS

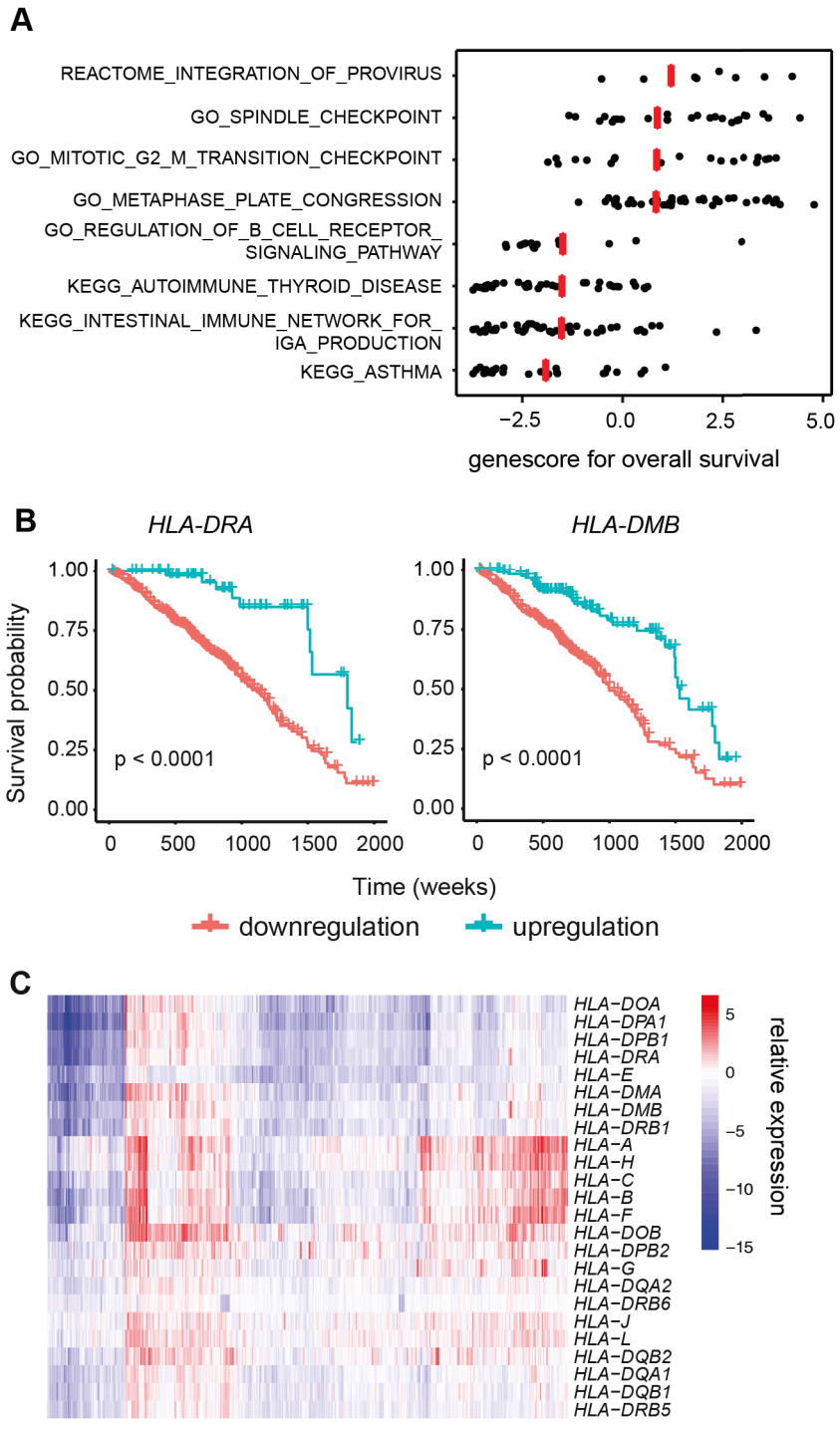

Figure 1. Cell cycle genes and *MHCII* genes are potential prognostic predictors of LUAD.

A. Gene set analysis of TCGA LUAD data to determine the significance of curated canonical

pathways with respect to patient overall survival. Each dot represents the individual gene score within corresponding pathway and red line is the score for the gene set calculated from R package of GSA. B. Kaplan-Meier plot showing the 5-year overall survival with respect to *HLA-DRA* and *HLA-DMB* expressions for patients in TCGA LUAD cohorts. Log-rank test was performed for significance. C. Heatmap of the relative expression of MHC genes in tumor tissues compared to matched normal tissues for TCGA LUAD data.

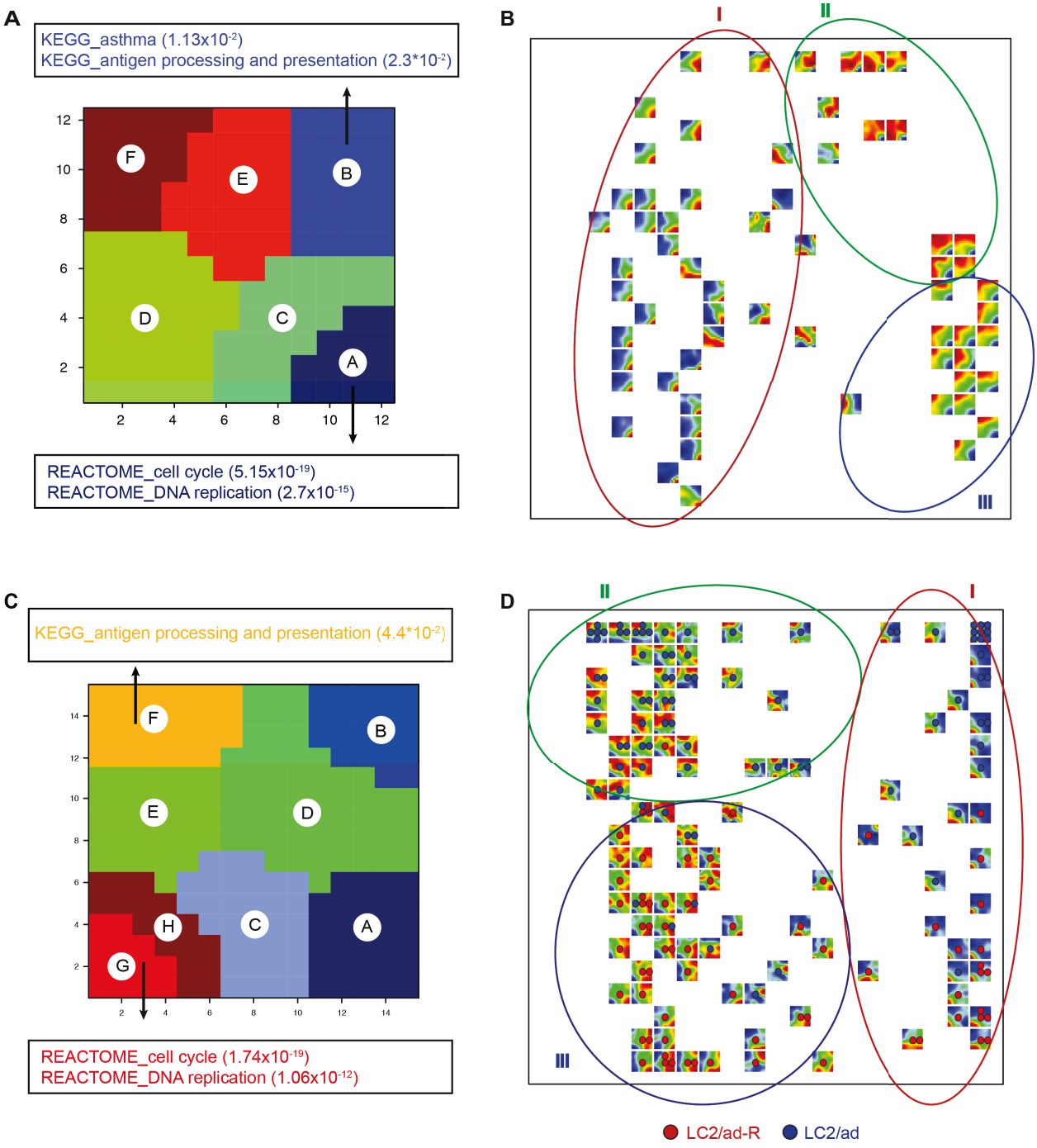

Figure 2. Expression of genes in cell cycle pathway and antigen presentation pathway among single cancer cells are co-regulated. A and C, Gene pathway clusters from metagene analysis of single cells from PDX LC-PT-45 (A) or from LC2/ad-R and LC2/ad cell lines (C). K-means clustering of metagenes on SOMs into 6 (A) and 8 clusters (C), respectively. Hypergeometric test was then performed on each cluster of metagenes to determine enrichment

of canonical pathways. For PDX LC-PT-45 (A), Cluster A is enriched for cell cycle pathways, while Cluster B is enriched for antigen presentation pathways. For LC2/ad-R and LC2/ad cell lines (C), Cluster G/H is enriched for cell cycle pathways, while Cluster F is enriched for antigen presentation pathways. P-values in the bracket are false discovery rate corrected. B and D, Second level clustering of SOMs for 77 single cells from PDX LC-PT-45 (B) or 159 single cells from LC2/ad-R and LC2/ad cell lines (D). Each square is a unique SOM pattern with heatmap indicating the gene expression level of metagenes. Cells with the same SOM are collapsed with only one representative SOM. SOMs are arranged by mapping all cells together into a two dimensional mosaic pattern based on metagene expression. Co-expression of cell cycle pathway and antigen presentation pathway further separate cells into three major groups: cell cycle pathway high (I), antigen presentation pathway high (II), and both pathways low (III).

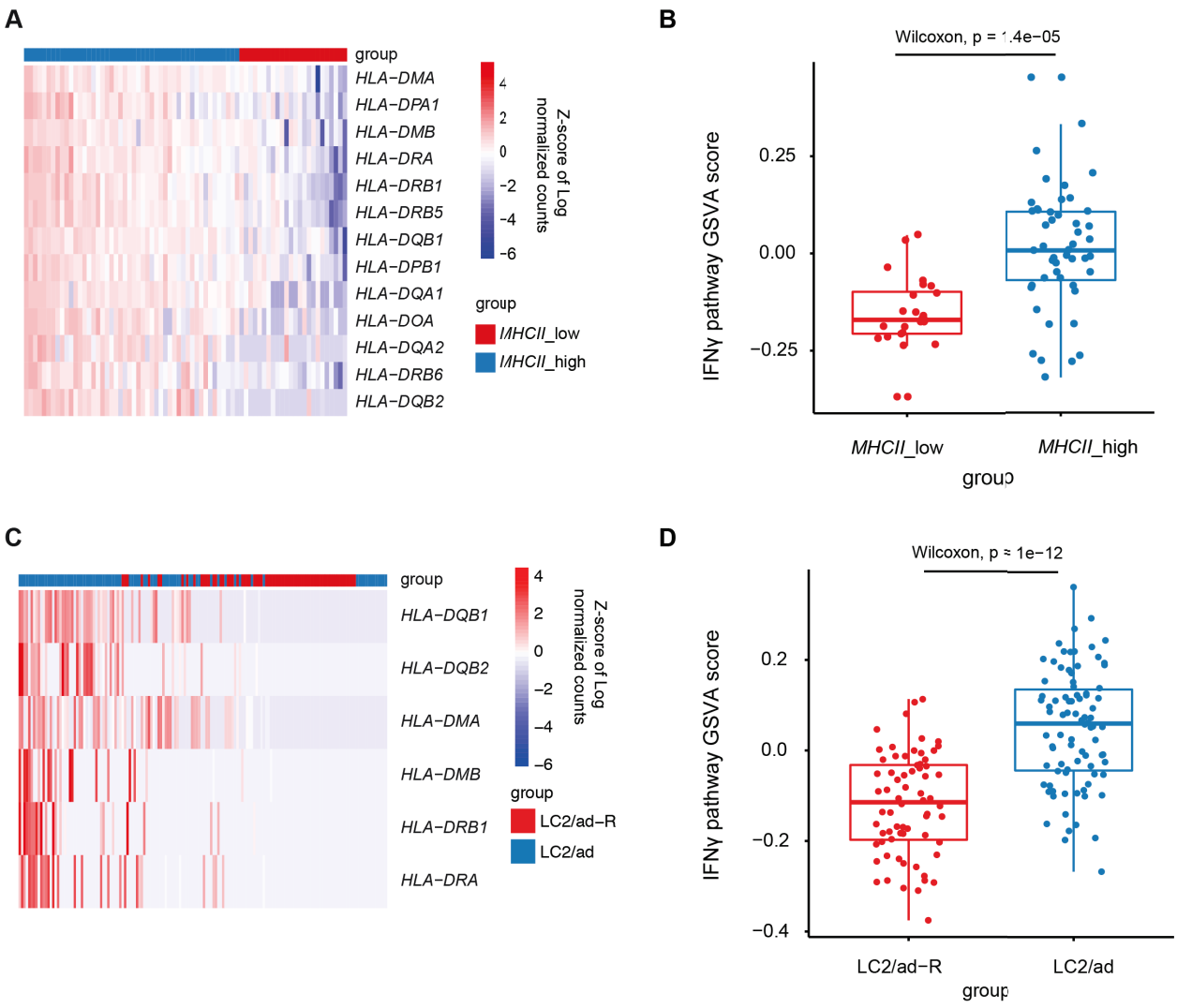

Figure 3. Co-expression of IFNγ signaling pathway genes in LC-PT-45 and cell lines. A and C, Heatmaps of *MHCII* gene expression levels for single cell in PDX LC-PT-45 (A) or in LC2/ad-R and LC2/ad cell lines (C). K-means clustering was performed to group all single cells into two clusters, *MHCII* low and *MHCII* high, in (A); while in (C), LC2/ad-R or LC2/ad were used to label cells. B and D, Comparison of GSVA scores of IFNγ signaling pathway genes between *MHCII*_low and *MHCII*_high groups (B) and LC2/ad-R and LC2/ad groups (D). GSVA scores of IFNγ signaling pathway were calculated for each individual cell analyzed in (A) and (D) respectively. Non-parametric Wilcoxon test was performed between different groups.

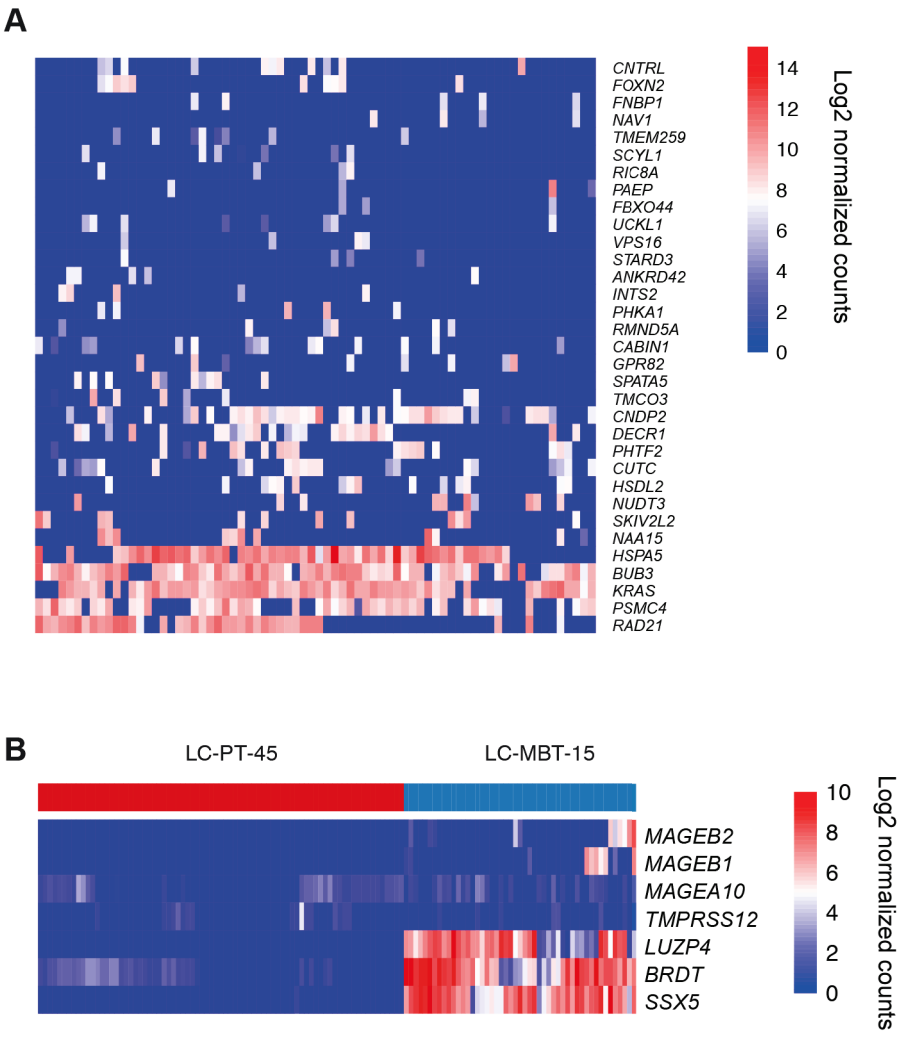

Figure 4. Heterogeneous expression of neoantigens and CTAs in single cancer cells. A. Heatmap showing the expression of neoantigens in single cells from LC-PT-45. Log2 normalized counts being 0 represents either no expression or no somatic mutation detected. Cells were included if a neoantigen is detected regardless of its corresponding wildtype antigen detection status. Only neoantigens detected in more than three cells were selected. B. Heatmap showing the expression of CTAs in single cells from LC-PT-45 and LC-MBT-15. CTAs whose expressions are transcriptionally silent in normal non-germline tissues based on Genotype-Tissue Expression (GTEx) data and with normalized counts > 0 in more than 2 cells were selected.