DNA barcoding reveals ongoing immunoediting of clonal cancer populations during metastatic progression and in response to immunotherapy

Louise A. Baldwin[1,2], Nenad Bartonicek[1,2], Jessica Yang[1], Sunny Z. Wu[1,2], Niantao Deng[1,2], Daniel L. Roden[1,2], Chia-Ling Chan[1], Ghamdan Al-Eryani[1,2], Alexander Swarbrick[1,2], and Simon Junankar[1,2*].

Affiliations

(1) The Kinghorn Cancer Centre and Cancer Research Theme, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia

(2) St Vincent's Clinical School, Faculty of Medicine, UNSW Sydney, NSW 2052, Australia

*Corresponding author: s.junankar@garvan.org.au

Key words: Immunoediting, immunotherapy resistance, breast cancer

Abstract

Cancers evade the immune system in order to grow or metastasise through the process of cancer immunoediting. While checkpoint inhibitor therapy has been effective for reactivating tumour immunity in some cancers, many solid cancers, including breast cancer, remain largely non-responsive. Understanding the way non-responsive cancers evolve to evade immunity, what resistance pathways are activated and whether this occurs at the clonal level will improve immunotherapeutic design. We tracked cancer cell clones during the immunoediting process and determined clonal transcriptional profiles that allow immune evasion in murine mammary tumour growth in response to immunotherapy with anti-PD1 and anti-CTLA4. Clonal diversity was significantly restricted by immunotherapy treatment at both the primary and metastatic sites. These findings demonstrate that immunoediting selects for pre-existing breast cancer cell populations, that immunoediting is not a static process and is ongoing during metastasis and immunotherapy treatment. Isolation of immunotherapy resistant clones revealed unique and overlapping transcriptional signatures. The overlapping gene signature was predictive of poor survival in basal-like breast cancer patient cohorts. Some of these overlapping genes have existing small molecules which can be used to potentially improve immunotherapy response.

Introduction

All cancers must find ways to evade the immune system so that they can continue to grow (*1*). Previous studies have established that this occurs through a process called immunoediting (*2*). During immunoediting, the more immunogenic cancer cells are selectively eliminated by the immune system thus leaving behind less immunogenic cancer cells that are then free to expand. Immunoediting can occur through multiple mechanisms, these include the elimination of cells with strong immunogenic mutations, leading to the loss of neo-antigens (*3*) or the selection of cells with elevated expression of various immunosuppressive programs (*4*).

Immunotherapies look to overcome some of the immune evasion pathways established by the cancer cells to avoid recognition and elimination by the immune system during immunoediting. The prominent clinically approved immunotherapies for solid tumours target T cell checkpoint molecules (eg. anti-

CTLA-4 and anti-PD1) to overcome T cell exhaustion (*5, 6*). In select cancer types such as melanoma, these drugs have dramatic effects in a large proportion of patients (*7*). Unfortunately for metastatic breast cancer early clinical trials have seen few patients experiencing durable responses even in the most sensitive basal-like subtype of breast cancer (*8*). This indicates that in metastatic breast cancer, resistance can rapidly develop to anti-PD1/PDL1 therapy and suggests that alternate immune drug targets are needed for breast cancer.

While the immune system is known to play a role in breast cancer outcome (*9*) and recent evidence has indicated that immunoediting can occur in a transgenic mouse model of breast cancer (*10*), very little is known about the specifics of immune evasion by breast cancer cells. The majority of studies examining this phenomenon were performed in the highly mutated MCA carcinogen driven sarcoma model, and could not track the response at a clonal level (*11*). Of interest, a recent study suggested that immunoediting by T cells can occur at the clonal level, by demonstrating the selection of clones that contain less-immunogenic fluorophores (*12*). This leaves an important gap in our collective knowledge as to the mechanisms employed in less immunogenic tumours such as breast.

NK cells and T cells have both been demonstrated to play a role in immunoediting (*13, 14*). However, the majority of recent research has focussed on pathways relevant to T cell recognition (*15-19*). Downregulation of MHC is one mechanism by which cancer clones become impervious to T cells (*20*), but this inherently makes them targets of NK activity. In breast cancer NK dysfunction is noted and this is regulated by microenvironmental factors (*21*). Data on resistance pathways that allow for immune evasion from both T cells and NK cells are currently more limited.

To understand the process of immunoediting in breast cancer we conducted DNA barcoding (*22, 23*) of murine EMT6 and 4T1 mammary carcinoma cells, which were introduced into both immune competent and immunocompromised mice (Fig 1). DNA barcoding allows tracking of individual clones and clonal expansion while avoiding introduction of potentially immunogenic proteins (*24*). This system allowed us to analyse immunoediting *in vivo* at primary and metastatic sites and to study whether resistance to checkpoint immunotherapy developed from pre-existing or *de novo* generated cell populations. We show that immunoediting is initiated by the endogenous immune system in primary tumours and that cancer cells undergo a second round of immunoediting during metastasis. We further observed that immunoediting of specific clones is enhanced at both sites by checkpoint immunotherapy. We also identified cancer cell clones highly resistant to immunotherapy. RNA-sequencing (RNA-Seq) analysis of these resistant clones demonstrated that each clone had activated unique immune evasion pathways, with one downregulating MHC-I expression and another upregulating PD-L1. However, these clones also contained a common gene expression signature that is highly predictive of poor survival in both the METABRIC and TCGA basal-like breast cancer cohorts. This study has thus determined the patterns of immunoediting at a clonal level, that metastatic cells undergo a second round of immunoediting, and that immunotherapy significantly restricts clonal diversity. We go on to determine unique pathways activated in immune evasive cancer cells that are targetable and could improve immunotherapy response in breast cancer patients.

Results:

Immunoediting of breast cancer cells in the primary tumour

To understand the role of the immune system and immunotherapy in shaping the clonal dynamics of cancer cells within primary tumours we used the immunotherapy-sensitive syngeneic mammary carcinoma model EMT6 (*25*). The ClonTracer DNA barcode library (*22*) was introduced into the EMT6 cells resulting in ~41 000 unique barcodes identified by DNA sequencing. Following inoculation of 250 000 cells (~6 fold over representation of each barcode) into the mammary fat pad we compared the number of clones able to engraft and grow in immune-competent, syngeneic wild-type (WT) Balb/c mice or severely immunocompromised NOD SCID Gamma (NSG) mice (Fig 2A).

Tumour growth was much more rapid in the NSG mice with tumours reaching ethical endpoint on day 14 post-transplant, whereas wild-type mice all reached ethical endpoint by day 23 (Fig 2B). This led to NSG mice having significantly shorter overall survival (median survival 14 days) when compared to wild-type mice (median survival 22 days, Mantel-Cox p=0.009), demonstrating that the immune system plays an important role in controlling primary tumour growth in the EMT6 model (Fig 2C).

To examine the influence of immunotherapy on tumour growth & clonal dynamics, we compared wild-type mice treated with combination immunotherapy (anti-PD1 + anti-CTLA4) or control antibodies starting from day 10 when tumours were approximately 200mm^3 (Fig 2A). All control mice reached ethical endpoint by day 23 (Fig 2D). In contrast, all treated tumours regressed following treatment, with 50% relapsing and reaching ethical endpoint between days 46 and 54 (Fig 2D). The remaining immunotherapy treated mice remained tumour free when the experiment was terminated on day 60. Kaplan-Meier analysis demonstrated that immunotherapy significantly increased survival with median survival increasing from 22 days to 57 days (Mantel-Cox p=0.0006) (Fig 2E).

To determine if immune control of tumour growth was driven at a clonal level, we examined the number and distribution of barcodes present in primary tumours. We found that at ethical endpoint the tumours grown in NSG mice had over 50 times the number of unique barcodes as tumours grown in control WT mice (p=0.0002, unpaired t-test), which in turn had more than 20 times the number of unique barcodes found in immunotherapy treated WT mice (p=0.0019, unpaired t-test) (Fig 2F). We applied Shannon diversity analysis to these samples to understand how the immune system influenced the diversity of barcodes. Shannon diversity index is determined by how evenly distributed the barcodes are within a population and is only moderately influenced by barcode number. Analysis of barcode diversity revealed a trend to a reduction in barcode diversity in the tumours from the control Balb/c mice compared to the NSG mice, whereas there was a dramatic reduction in barcode diversity following immunotherapy treatment (p<0.001 unpaired t-test) when compared to the control treated mice (Fig 2G). This data indicates that a subset of EMT6 cells are more resistant to the endogenous immune system but that this selection does not skew the evenness of the barcode distribution dramatically. This suggests that the clones that are resistant to the immune system all have a similar level of resistance. In contrast immunotherapy applies a much more stringent bottleneck that only a limited

number of clones can readily overcome and with a high variability in the levels of resistance. Further analysis identified specific EMT6 clones that were reproducibly enriched across multiple replicate mice following immunotherapy treatment indicating that they had a pre-existing resistance phenotype that was being positively selected for (Fig 2H).

Immunoediting of breast cancer cells during metastasis

To determine whether immunoediting continued during metastatic dissemination and whether specific metastatic clones were enriched or depleted, we turned to the highly metastatic 4T1 mammary carcinoma model, as the EMT6 cell line is poorly metastatic (*26*). The barcode library was introduced into the 4T1 cells leading to a cell line with ~5000 unique barcodes. To see how the immune system regulated clonal cancer growth during the metastatic cascade, 50,000 cells (a ~ 10-fold over- representation of each barcode) were inoculated into the mammary fat pad of WT and NGS mice. Primary tumours were resected 15 days following inoculation to allow metastases to develop (Fig 3A). Following primary tumour resection all mice developed lethal lung metastases. NSG mice succumbed to metastatic disease earlier than WT mice with a median survival of 25.5 days verses 35 days ($p = 0.0002$, Mantel-Cox Log-rank test) (Fig 3B). Primary tumour size at resection was equivalent between the groups (Sup Fig 1A). Adjuvant immunotherapy with combination anti-PD1 + anti-CTLA4 led to a modest but significant increase in survival (37.5 days) versus control treated mice (33 days; $p=0.0121$, Mantel-Cox Log-rank test) (Fig 3C).

We then examined whether the endogenous immune system shaped metastatic clonal dynamics. While primary tumours contained similar numbers of clones and barcode diversity in NSG and WT hosts (Fig. 3D and Sup Fig 1B, C), metastatic lungs of NSG mice contained ~3 times as many barcode clones as WT controls (Fig 3D). We next sought to determine if the increase in survival following immunotherapy (Fig. 3C) was associated with alterations in clonal dynamics. As the treatment was only given after excision of the primary tumour, the immunotherapy would only affect the outgrowth of cancer cells that had already metastasised to the lung. Despite only observing a modest increase in survival following combination immunotherapy (Fig 3C), we observed a 70% reduction in the number of clones able to form metastases (Fig 3E).

The increase in barcode number in the lungs of NSG mice was associated with a significant increase in diversity as measured using the Shannon diversity index (Fig 3F). This shows that the endogenous immune system restricts the number and skews the diversity of metastatic clones that can reach and outgrow in the lungs. In addition to the reduction in barcode number following immunotherapy treatment, we also saw a significant reduction in barcode diversity as measured using the Shannon diversity index (Fig 3F). This indicates that immunotherapy is leading to the immunoediting of specific clonal cell populations over others.

To further understand the key immune cell types that control clonal outgrowth in the metastatic lungs we depleted either CD8 T cells (anti-CD8) or NK cells (anti-asialo-GM1) in wildtype mice starting one day prior to tumour resection. Neither treatment led to a significant change in overall survival (Sup Fig 2A). However, depletion of either cell type led to an increase in the number of clones detected within the lungs, reaching statistical significance with T-cell depletion (Fig 3G).

This indicates that CD8 T cells and perhaps NK cells play a role in eliminating metastatic clones within the lungs.

These results were reproducible, as demonstrated by a second pool of barcoded 4T1 cells containing a larger barcode library (300 000 barcodes). Following the injection of 50 000 barcoded cells we recovered approximately 10 000 - 12 000 barcode sequences from each primary tumour and this was relatively unchanged in the NSG mice (Sup Fig 3A). This suggests that roughly a fifth of the injected cells are able to engraft and grow in the mammary gland. As clone diversity was unchanged in NSG mice this confirms that the immune system does not play a major role in restricting growth of 4T1 carcinoma cells in the primary tumour setting. In contrast when we examined the number of clones that had spread to the lungs of NSG mice we again found approximately 3 times as many when compared to the wild-type mice (Sup Fig 3A). In addition, we similarly saw a ~3-fold reduction in barcode diversity in response to immunotherapy (Sup Fig 3B). Due to the high complexity of the 300 000 barcode library each mouse received only a partially overlapping complement of barcode clones.

Patterns of enrichment and depletion of specific clones

To better understand how specific clonal cell populations responded to the immune system and immunotherapy we combined the barcode frequencies from the two datasets utilising the 5000 barcode library (WT vs NSG and Control vs Immunotherapy). We performed unsupervised hierarchical clustering of these samples and selected barcodes that were observed at greater than 5% frequency in any one sample. We found that the primary tumours from the two experiments cluster together irrespective of the immune status of the mouse (Balb/c or NSG), further suggesting that 4T1 cells do not undergo immunoediting at the primary site (Fig 4A). In contrast, lung tumours formed in the NSG hosts did not cluster with lung tumours formed in Balb/c lungs, with the immunotherapy treated samples mostly clustering alone or with metastases formed in WT mice. A number of specific barcodes were enriched in the metastatic lungs of all the NSG mice indicating these clones were highly metastatic (Fig 4A). This agrees with the findings of Wagenblast and colleagues (*23*).

We identified a number of barcodes that had striking patterns of enrichment or depletion in response to the immune system and immunotherapy, we replotted these using a dot plot (Fig 4B). As these barcodes were enriched or depleted in a reproducible manner across replicate mice this suggests these are due to inherent features of these clones. Firstly, there are three barcodes that were enriched within the NSG lungs (NSG1-3), observed at lower abundance in the untreated wild-type mice, and that were completely eliminated following immunotherapy treatment. This suggest that while these clones are highly metastatic in the absence of an immune system, they are immunogenic and are thus subjected to immunoediting in WT mice particularly following immunotherapy. Another group of metastatic clones that were present in the lungs from NSG and WT mice were further enriched following immunotherapy (IE1-2). These immunotherapy enriched clones were detected in the lungs of all six replicate mice. With the dramatic reduction in the number of barcodes present following immunotherapy, the odds ratio of this happening by chance is 0.0034 (95% confidence interval: 0.0010-0.0079; chi square p value: 3.67×10^{-251}). This suggests that these clones

have a pre-existing resistance phenotype and are positively selected for following immunotherapy.

To further analyse how specific barcodes were enriched in lung metastases following immunotherapy we visualised the top nine clonal populations (based on average barcode proportions in the metastatic lungs) and generated fish-plots. These showed that different clones were preferentially enriched in the lungs of NSG mice when compared to WT mice (Fig 4C). Furthermore, we observed that a small subset of clones were highly enriched in the lungs of immunotherapy treated mice (Fig 4D).

Analysis of immunotherapy resistant clones

To understand more about the phenotype of these immunotherapy resistant clones we established clonal cell populations from two of them (designated IE1 and IE2), and two independent control clones (NT1 and NT2) that were not enriched following immunotherapy (3-4 independent clonal cell lines were generated per barcode). These clonal cell lines were isolated from the parental barcoded 4T1 cell population purely *in vitro* with no additional selective manipulation. The barcode within each of these clonal cell lines was confirmed to be correct using Sanger sequencing. All four clonal cell lines had similar growth kinetics *in vitro* indicating no proliferative advantage of the immune evasive clones *in vitro* (Sup Fig 4A).

Genomic analysis for barcode integration site and copy number variation (CNV)

To identify the barcode integration sites and determine whether the clones contained large scale genomic alterations we performed whole genome sequencing (WGS) at around 30x coverage of the clones. The WGS analysis determined the precise genomic location at which the barcodes integrated (Sup Table 1). IE1 integrated in the intergenic region between Kpna2 and Smurf2 and IE2 within an intron of Nrf1, neither integration site changed the coding sequence of these genes. Copy number analysis determined that no clone demonstrated dramatic copy number changes when compared to the other clones. Each clone only contained a small number of single copy number gains and losses (IE1 only 6 CNVs and IE2 only 5 CNVs), with clone NT2 showing the greatest number of CNVs at 41 (summarised in Sup Table 2). We found one locus with a single copy number gain in both IE1 and IE2 that led to 3 copies of the genes *Nc3r1* (the Glucocorticoid receptor) and *Arhgap26* a Rho GTPase that associates with focal adhesion kinase (FAK), however, this gain was also present in the NT2 clone. These results demonstrate that large scale genomic changes likely do not play a large role in the various phenotypes of the different clones but suggest that copy number changes may be selected against during immunoediting.

Transcriptomic analysis of the clonal cell lines

To investigate the mechanism of immune evasion by these clones we performed RNAseq analysis and compared the two immunotherapy resistant clones to the bulk 4T1 population. The IE1 clone had 1553 differentially expressed genes (Log fold change >2 and FDR $p<0.05$) with 478 significantly upregulated and 1075 significantly downregulated (Fig 5A). The IE2 clone had 1099 differentially expressed genes with 375 significantly upregulated and 724 significantly

downregulated (Fig 5B). The non-target clones had fewer gene expression changes compared to the bulk 4T1 population with NT1 having 621 and NT2 having only 262 differentially expressed genes. We examined the top differentially expressed genes between each of IE1 and IE2 with the parental 4T1 cells, however, we did not find any with an obvious role in immune evasion (Sup Tables 3 & 4). Gene set enrichment analysis revealed that two of the top ten gene sets were upregulated in both IE1 and IE2 (CHEN_HOXA5_TARGETS_9HR_UP, BLUM_RESPONSE_TO_SALIRASIB_UP), while most were unique to either cell line (Sup Tables 5 & 6). *Hoxa5* is a known tumour suppressor gene in breast cancer (*27*) and although we see an enrichment of its target genes, the expression of *Hoxa5* itself was significantly reduced in the IE1 clone and there was a trend to reduced expression in the IE2 clone. There was no overlap in the top ten down regulated gene-sets between IE1 and IE2. The top downregulated gene-set for IE1 was the REACTOME_UB_SPECIFIC_PROCESSING_PROTEASES gene-set, that contained two genes involved in antigen processing for display by MHC-I (*Psmb8* and *Psmb9*). As down regulation of the MHC-I pathway is a common mechanism of immune evasion we investigated this in more detail.

Through this analysis we found that the IE1 clone had significantly reduced expression of many genes related to antigen presentation including MHC-I (*H2-k1*), *Tap2*, *Psmb8*, *Psmb9* and *Psmb10* (Fig 5C). *H2-k1* is the main MHC molecule expressed by the Balb/c strain of mouse that the 4T1 carcinoma cell line was derived from. We validated the reduction in MHC-I expression levels seen in the RNAseq data at the protein level using flow cytometry (Fig 5D). This analysis showed that the IE1 clone had significantly reduced cell surface MHC-I protein expression compared to the bulk 4T1 population. We thus examined the WGS data and this showed that the loss of MHC-I expression in IE1 was not due to genomic loss at the MHC locus on chromosome 17 (Sup Fig 5A. In contrast the IE2 clone had elevated levels of a number of these MHC related genes (Fig 5C), in addition to a non-classical MHC molecule *H2-t23* (Fig 5E) that is known to negatively regulate NK cells through their inhibitory receptor *Nkg2a* (*28*). Interestingly IE2 cells also demonstrated a significantly increased expression of the T cell inhibitory molecule *Cd274*/PD-L1 (Fig 5E). This again was validated at the protein level using flow cytometry (Fig 5F). These results demonstrate that each of the two immunotherapy resistant clones are phenotypically unique.

We next examined whether the copy number changes or barcode integration sites identified above impacted gene expression. In IE1 and IE2 the copy number changes in *Nc3r1* (the Glucocorticoid receptor) and *Arhgap26* were associated with significantly increased expression of these genes but NT2 demonstrated no change in expression (Sup Fig 5B). Elevated Nc3r1 expression has been associated with poor prognosis and metastasis in TNBC although whether it plays a role in immune evasion is not known (*29, 30*). As stated above the barcode for IE1 integrated in the intergenic region between Kpna2 and Smurf2 and IE2 within an intron of Nrf1. The expression of Smurf2 was the only of these genes that was significantly altered with a modest log fold increase of 0.59 in IE1.

Demethylating drugs do not fully restore MHC expression

Demethylating agents such as 5-aza-2'-deoxycytidine (5-aza) are known to upregulate MHC-I expression in cancer cells (*31*), thus we treated our clonal cell

lines utilising 5-aza for 72 hours to determine whether DNA methylation was a mechanisms suppressing MHC expression in the IE1 clone. Using flow cytometry, we observed that MHC-I expression was elevated in a dose dependent manner following 5-aza treatment in all clones. However, MHC-I expression in the IE1 clone was consistently lower than the parental 4T1 cell line at all doses of 5-aza (Sup Fig 6A). This indicates that gene hyper-methylation is not the mechanism of MHC-I suppression in the IE1 clone.

IFN-gamma stimulation is another mechanism by which MHC expression can be increased on cancer cells. The IE1 clone responded to IFN-gamma treatment by upregulating MHC-I expression but again it remained suppressed compared to the parental 4T1 cells (Sup Fig 6B). This suggests these cells broadly retain the transcriptional regulatory machinery that is required to upregulate MHC-I in response to IFN-gamma stimulation. These results indicate that MHC downregulation is likely regulated by epigenetic factors other than DNA methylation and that the majority of MHC expression in this clone can be restored by IFN-gamma treatment.

Overlapping gene signature predicts poor survival in breast cancer patients

We had noted that the GSEA analysis showed some overlap in enriched gene-sets between the two immune evasive clones (IE1 and IE2), we thus reasoned that as well as having unique immune evasion features these clones may have some pathways in common. To identify the common immune evasion pathways being initiated by both the immunotherapy resistant clones we generated Venn diagrams to identify overlapping gene expression changes (Sup Table 7). This analysis demonstrated that the immunotherapy resistant clones had more gene expression changes in common with each other and less in common with the control non-target clones (Fig 6A). We generated a heatmap of the top 50 upregulated and downregulated genes across all the samples (Fig 6B) and performed GSEA analysis using C2 on the longer list (Sup Fig 7A, Sup Table 8). Only two gene sets had significant p values when you consider multiple testing, these were the HOXA5 gene set mentioned previously and a COVID19 related gene set. Although not significant there were several additional COVID19 related gene sets from the same recent publication identified in the overlapping upregulated gene list suggesting an immune related role of these genes (*32*).

Therefore we wanted to understand the role of these genes in patients, so we generated signatures from the top 25 upregulated and downregulated genes that had human orthologs and were detectable in both the METABRIC (*33*) and TCGA datasets (*34*). We then analysed the role of these signatures in survival using the basal-like breast cancer patients from these cohorts. While the patients in these cohorts have not been treated with immunotherapy it has previously been demonstrated that immune features such as the number of tumour infiltrating lymphocytes or regulatory T cells influence prognosis in basal-like breast cancer patients (*35*). When we analysed overall survival in these cohorts we observed that the upregulated signature associated with substantially poorer outcome in both cohorts (METABRIC: p=0.0043, HR=2.0, Fig 6C; TCGA: p=0.042, HR=4.3, Fig 6D). In contrast the downregulated gene signature did not show any significance in either cohort (data not shown). We generated heatmaps with unsupervised clustering to determine whether specific genes or groups of genes from the

signature were specifically driving the association with survival (Sup Fig 7B). In the TCGA data we observed a number of clusters that seemed to associate more with survival, one of these included the genes *FAM71F2, MASP2, HLF, PPP1R15A, MMP23B* and *LIMS2*. Interestingly one of these genes *Ppp1r15a* also known as GADD34 has previously been demonstrated to be critical in blocking immunogenic cell death following chemotherapy, and when it was inhibited the chemotherapy response was improved in immunocompetent mice but not immunocompromised mice (*36*). A second group of genes included *SEZ6L2* and this gene has been associated with survival in a number of cancers but not through an immune related mechanism (*37, 38*). There was no enrichment of proliferation or invasion gene sets in our GSEA analysis suggesting that these processes were not behind the poor outcome of patients that highly express genes in the common upregulated signature.

Previous studies have shown that CTL infiltration correlates with survival in basal-like breast cancer so it is possible our signature is a surrogate measure of T cell infiltration. To test this, we performed TIDE analysis (*39*) on the TCGA and METABRIC cohorts followed by correlation analysis between the CTL signature score and our upregulated immune evasion signature score. This showed no correlation in the METABRIC cohort and only a weak negative correlation in the TCGA cohort (Sup Fig 7C) suggesting little overlap between these two predictors of patient survival. Future studies will be necessary to determine how the genes in this signature regulate survival, influence immune evasion and immunotherapy response.

Discussion:

Immunotherapy is revolutionising cancer therapy, with dramatic long-term responses seen in certain patient groups. Unfortunately, other patients see limited response to the current immunotherapies, and in particular some cancer types such as breast cancer are less responsive in general. Thus, our understanding of immune evasion in non-responsive cancer types needs significant improvement. We looked to address this by examining the fundamental pattern of immune evasion at a clonal level, and furthermore to use this information to identify pathways that could be targeted to overcome immunotherapy resistance. Using cellular DNA barcoding, we have demonstrated that the immune system shapes clonal evolution of primary and metastatic cancer cells in syngeneic models of basal-like breast cancer. Additionally, we have demonstrated that a second round of immunoediting occurs during metastasis and thus immune evasion at the primary site does not guarantee immune evasion during metastasis. DNA barcoding allows for tracking of specific clones with exquisite sensitivity. We find that immunotherapy enhances immunoediting at both the primary tumour and metastatic sites. Finally, we identify a common immune evasion gene signature in metastatic clones that is predictive of survival in basal-like breast cancer patients.

Here we show, for the first time in a solid tumour, that clonal immunoediting occurs in response to immunotherapy. Using the more immunotherapy sensitive EMT6 model we demonstrate that immunoediting occurs at the primary tumour site and that immunotherapy leads to strong clonal selection. The more immunotherapy resistant 4T1 model demonstrated that while most 4T1 cells are able to evade the immune system at the primary site, only a subset of these cells

are able to evade the immune system during metastasis. This indicates immune evasion is not a static process but requires ongoing regulation through tumour progression, even in a highly aggressive allograft model. These findings agree broadly agree with recent comprehensive genomic analysis of patient samples assessed across metastatic sites and over time (*3, 40*). These studies tracked clonal populations in metastatic lesions using whole genome sequencing, examined a number of immune correlates and could identify immunoediting that was associated with the immune response. However, they were unable to examine clonal heterogeneity driven by epigenetic or transcriptomic changes and were limited in identification of rare clones by sequencing depth. Future clinical studies utilising single cell approaches to analyse multiple biopsies from individual patients over a time-course of treatment will be necessary to confirm the key findings of this study in patients. While extremely challenging these studies are becoming more feasible with recent technological improvements.

Our study and that of others have indicated that immunoediting can be mediated by both CD8 T cells and NK cells (*13, 14*). In addition, immunoediting can select for clones with more immune evasive phenotypes irrespective of specific neoantigens (*2, 4, 6*). The only previous study examining immunoediting at a clonal level, used a fluorescent barcoding approach in a B cell leukemia model (*12*). While our findings broadly agree with the findings of this study, the study by Milo and colleagues was limited to five unique fluorescent clones that could be tracked, and were confounded by the variable immunogenicity of each of the different fluorescent proteins. Our DNA barcoding approach in contrast allowed for the labelling of thousands of clones and a much more precise identification of immune evasive clones. We could then isolate these specific clones and identify both common and variable features of immune evasion in individual clones. Furthermore, this technique unlike a fluorescent barcode approach allowed us to determine that pre-existing clones existed in both the EMT6 and 4T1 models that had greater immunotherapy resistance, as these exact clones were enriched from the same starting pool of cells across replicate mice.

We identified and isolated two of the immunotherapy resistant clones from the 4T1 model, based on their unique barcodes. In depth analysis of these resistant clones demonstrated they had differential gene expression of key immune evasive pathways (MHC-I and PD-L1). Intratumoural heterogeneity (ITH) has previously been associated with immune evasion and resistance to immunotherapy in melanoma and lung cancer. Tumours with higher ITH were shown to be more resistant to immunotherapy (*41, 42*). McGranaham and colleagues postulated that this was due to improved T cell killing of tumours with clonal neo-antigens. Another non-mutually exclusive explanation is that clonal tumours are less likely to contain cancer cells with a pre-existing resistance mechanism to immunotherapy. These findings refine the concept of cancer immunoediting, demonstrating that there are clonal populations of cancer cells with variable resistance to the immune system. Based on their phenotype these clones are either enriched or depleted by an active immune system and immunotherapy.

We identified a core overlapping gene expression profile between the immunotherapy resistant clones. The common upregulated gene signature showed a dramatic ability to stratify basal-like breast cancer patients into good and poor prognosis. This gene signature appeared to represent a novel immune

evasion pathway associated with poor prognosis. Aside from *Ppp1r15a* it did not have genes known to be associated with immune evasion, and in addition did not contain genes associated with other poor prognostic signatures such as proliferation or invasion.

Importantly as both T cells and NK cells are present during immunoediting in our model system this common signature likely enables the cancer cell clones to evade both T cells and NK cells. Interestingly our common signature did not strongly correlate with CTL infiltration as determined by TIDE analysis (*39*), indicating that this signature was not a surrogate for the lack of T cell infiltration. This also suggests these genes likely do not regulate immune evasion by influencing immune cell recruitment. These common genes may offer new insights into developing therapeutic approaches to improve immunotherapy response in breast cancer. One of the common immunotherapy resistance genes we identified was *PPP1R15A* that may be important in immunotherapy response. Our identification of *PPP1R15A* as elevated in immuno-resistant clones is consistent with its known role in immunogenic cell death in the response to chemotherapy (*36*). Some recent studies in another disease setting have demonstrated that this pathway can be targeted using a small molecule (Sephin-1), where it was used to treat a mouse model of multiple sclerosis (*45*). Future studies are necessary to examine whether this compound or others targeting this pathway could synergise with immunotherapy, or immunogenic chemotherapy to treat breast or other cancer types.

As breast cancers have a lower mutational burden than lung cancer and melanoma it is likely that epigenetic factors may play a greater role than mutational events in driving ITH in breast cancer. One current approach to improve immunotherapy response under investigation has been to combine immunotherapy with epigenetic targeting drugs such as decitabine and HDAC inhibitors (*31, 43*). This combination has been shown to lead to an upregulation of MHC protein expression and improved response to immunotherapy. However, the use of epigenetic drugs might have an additional effect of reducing ITH in certain immune evasion pathways. Thus, epigenetic drugs may reduce the diversity of clones and overcome other epigenetically driven immune evasion strategies as well as enhancing MHC expression (*44*). Further research is needed to test this hypothesis more fully. Our results, however, suggest that while demethylating agents could increase MHC-I expression in the MHC-I low immunotherapy resistant clone, they did not increase it above the baseline seen in the parental 4T1 cell line. This is corroborated by a recent study that demonstrated that while treating breast cancer patients with demethylating agents could increase MHC-I expression in most patients, a subset appeared resistant to this therapy (*31*). This suggests that while epigenetic treatments may improve the proportion of patients who respond to immunotherapy that in some cases pre-existing clones could still mediate resistance to this combination.

Some limitations of this study include the reliance on mouse cell line models, as these do not recapitulate the early stages of tumorigenesis and do not represent the full diversity of human breast cancer. However, syngeneic allograft models have delivered central insights about the immune response to cancer and demonstrated the utility of immunotherapies (*46*). A further limitation is that the integration of the barcode and selection markers into the genome and the

potential immunogenicity of the RFP could affect the phenotype of the cancer cells. In previous studies we and others have found some fluorophores and luciferase to be immunogenic and negatively affect tumour growth and metastasis in the 4T1 model (*24, 47*). However, in this study we found that tumour growth and metastasis were unaffected by RFP expression. While the introduction of the DNA barcodes could have influenced the phenotype of the specific clones we feel that this is unlikely, with no dramatic impact on the expression of the genes closest to the integration site. Furthermore, none of the genes associated with each integration site came out as being significantly involved in cancer cell evasion of CD8 T cell responses in a recent CRISPR screen (*15*).

Overall this study has demonstrated that immunoediting occurs at the clonal level in primary tumours and that a second round of immunoediting occurs during metastasis. Immunotherapies dramatically enhanced immunoediting, however, pre-existing resistant populations were still responsible for relapse. The large reduction in clonal diversity following immunotherapy in the 4T1 model, a model that is known to be poorly responsive to immunotherapy, suggests that slight improvements through combination therapy could eliminate the remaining clones and lead to dramatic improvements in survival. By isolating immunotherapy resistant clones and phenotyping them, we identified common and distinct immune evasion pathways. We anticipate through the targeting of the pathways identified in this study, in particular the common pathways, it will be possible to further reduce the numbers of resistant clones and improve the efficacy of immunotherapies.

Acknowledgments

This work is supported by a research grants from The National Breast Cancer Foundation (NBCF) of Australia (IN-17-02; IIRS-19-106) and supported by the generosity of John McMurtrie (AM) and Deborah McMurtrie and The Petre Foundation. LAB is the recipient of an Australian Government research training (RTP) scholarship. AS is the recipient of a Senior Research Fellowship from the National Health and Medical Research Council of Australia. SJ is supported by a research fellowship from the NBCF. We acknowledge the contribution of Isabelle Shapiro and Cheryl Grant as consumer advocates.

Methods:

Cells

4T1 cells were obtained from ATCC. 4T1 cells were grown in RPMI (Gibco) supplemented with 10% FCS, D-Glucose, Sodium Pyruvate, 2mM HEPES, and Penicillin/Streptomycin. EMT6 cells were obtained from ATCC. EMT6 cells were grown in Waymouth's MB 752/1 Medium supplemented with 15% FCS and 2mM L-glutamine.

Cellular DNA barcoding

The ClonTracer library was a gift from Dr Frank Stegmeier (Addgene #67267). Lentiviral particles containing the high-complexity barcode library were produced by transfecting 293T cells. 4T1 and EMT6 cancer cell lines were barcoded by lentiviral infection using 0.8 µg/ml polybrene. Cells from each line,

were infected with a target MOI of 0.1, corresponding to 10% infectivity to ensure single lentiviral integration. Cells that received a barcode were then sorted based on the RFP reporter protein using a BD FACSAriaII, these cells were then expanded and frozen into a number of aliquots for the subsequent experiments. 4T1 cells were generated with two different barcode complexities, one with ~5000 barcodes (4T1 BC5000) and one with ~300 000 barcodes.

Mice

All animal experiments were approved by the Garvan Institute of Medical Research/St. Vincent's Hospital Animal Experimentation Ethics Committee.

Immunocompetent BALB/c mice and immunocompromised NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) mice aged 6-to-8 weeks were obtained from Australian BioResources (Moss Vale, Australia) and housed at the Garvan Institute of Medical Research.

In vivo tumour growth

For tumour transplantation, barcoded EMT6 cells (ATCC, USA) were resuspended in Matrigel 2.5x10^5 cells in 100ul volume were injected into the 4th inguinal mammary gland; barcoded 4T1 cells (ATCC, USA) were resuspended in PBS and 5x10^4 cells in a 10ul volume were injected into the 4th inguinal mammary fat pad. For studies with the 4T1 model primary tumours were surgically resected at day 15. At resection or ethical endpoint tumours and metastatic lung tissue were removed, minced and snap frozen in liquid nitrogen for barcode analysis.

Immunotherapy treatment

Mice were treated with four 200ug doses of either combination immunotherapy antibodies via intraperitoneal injection: anti-CTLA4 (BE0032), anti-PD1 (BE0146), or isotype control antibodies Armenian hamster IgG (BE0091), Rat IgG (BE0089) all from BioXCell (Lebanon, NH, USA). Antibodies were given every 2-3 days from day 10 post tumour implantation for the EMT6 model and following resection on day 15 for the 4T1 model.

CD8 T cell and NK cell depletion

Starting one day prior to primary tumour resection mice were given 100ug of depleting antibodies for CD8 T cells (anti-CD8; BEO223; BioXCell), or NK cells (anti-Asialo-GM1; 986-10001; Novachem), or isotype control antibodies. Antibodies were then given every 2-3 days for a total of 4 doses.

DNA extraction

Frozen tissue samples were lysed in 5 ml QIAGEN buffer P1 (with RNaseA) and 0.5% SDS within a Miltenyi M-Tube (# 130-096-335). Samples were processed on the gentleMACS or gentleMACS Octo using the RNA_02 program. DNA was then extracted using a standard phenol/chloroform process.

Targeted Barcode PCR and Sequencing

All samples underwent targeted barcode PCR amplification according to the updated version of the original protocol (*22*) available on the Addgene website (https://www.addgene.org/pooled-library/clontracer/). Specific PCR products (180 bp) were gel purified, quantified by Qubit 2.0 fluorometer (ThermoFisher

Scientific, Waltham, MA USA) and pooled into a library. Prior to sequencing, an equal combination of additional PCR products containing two inverse barcodes (GACTCAGTGTCAGACTGAGTGTCTGACTGT and CTGAGTCACAGTCTGACTCACAGACTGACA) plus the PhiX Control V3 (Cat. FC-110-3001, Illumina, CA, USA) were spiked in to balance the nucleotide distribution within the library. Samples were sequenced using a custom sequencing primer (GCGACCACCGAGATCTACACACTGACTGCAGTCTGAGTCTGACAG) and the NextSeq® 500/550 Mid Output Kit v2 - 150 cycles (FC-404-2002, Illumina, CA, USA) on the Illumina NextSeq® platform.

Barcode Analysis

Barcode composition analysis and calculation of barcode overlap between samples was performed as indicated in the original protocol (*22*) and updated Python scripts available from the Addgene website (https://www.addgene.org/pooled-library/clontracer/).
Further analysis was performed using R statistical framework and packages EntropyExplorer for analysis of differential Shannon Entropy (*48*), DEBRA for differential barcode expression (*49*), and libraries fishplot and UpSetR for visualisation purposes.

Generating clonal cell lines

Cells of interest were isolated from the barcoded parental population using a sub-pooling approach.
The barcoded 4T1 BC5000 cells were seeded into a 96 well plate at a density of 150 cells per well. At approximately 80% confluence, cells were trypsinised and split identically into two plates. One plate was viably frozen in freezing media (10% DMSO, 40% FCS, 50% 4T1 media). DNA was extracted from one plate using the Promega SV Wizard Genomic DNA kit. Target barcodes of each sample were PCR amplified and sequenced using the method described above.
After sequencing, wells containing cells with the target barcodes were thawed, pooled and seeded at 40 cells/well in a 96 well plate. Media was changed every three days for 8 days before cells were split into two identical plates as above. One plate was viably frozen in freezing media, while DNA was extracted and prepared for targeted sequencing as above.
Wells with the highest proportion of target barcodes were revived into a 6 well plate and grown for 4 days before being single-cell sorted by BD FACSAria II into 96 well plate. Sorted single cells were grown in conditioned media for 5 days before being changed to 4T1 media and grown until 80% confluence. As previously, cells were lifted and split identically into two plates – one for freezing and one for targeted sequencing.
Once wells containing single cells clones of the cells of interest were identified, target wells were revived. Cells were expanded before being aliquoted and viably frozen for future experiments.
Barcoded sequences of isolated cells were confirmed by targeted Sanger sequencing of barcode regions.

Bulk RNA sequencing

RNA was extracted from established subclonal cell lines using the QIAGEN RNeasy Mini Kit. 3-4 unique clonal cell populations were sequenced for each barcode.

Libraries were prepared using the KAPA RNA HyperPrep Kit with RiboErase, and sequenced on the NextSeq500 platform using a High Output V2.5 300 cycle kit.

Transcriptome analysis

FastQ files from sequencing libraries were first trimmed with FASTQC v0.11 *Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc*. Raw reads were subsequently mapped to the mouse transcriptome (Gencode release M9, GRCm38.p4), to the mouse genome (mm10 assembly), with STAR aligner v.2.4.1d, allowing for multimapping reads (*50*). The reads were counted over gene models with RSEM, v.1.2.18 (*51*). Differentially expressed genes and repeat elements were defined with EdgeR with FDR<0.01 (*52*).

Survival analysis

To assess the clinical relevance of our isolated immune evasion clones, we assessed the association between the gene signatures derived from our bulk RNA-Sequencing studies with the overall survival of basal (PAM50) breast cancer patients from the METABRIC and The Cancer Genome Atlas (TCGA; https://www.cancer.gov/tcga) cohorts. Mouse gene signatures were first converted to human orthologs using the biomaRt package (*53*). Shared up-regulated genes across both immune evasion clones IE1 and IE2 were then filtered, and only genes detected in each expression cohort were considered. For each tumour from the bulk cohort, signature scores were computed based on the average expression of the top 25 genes ranked by log fold change. Patients were then stratified based on the signatures scores into the top 30%, middle 40% and bottom 30% groups. Survival curves were generated using the Kaplan Meier method with the 'survival' package in R (https://cran.r-project.org/package=survival). The Cox proportional hazards model was used to compute Hazard Ratios. We assessed the significance between groups using the log-rank test statistics.

Whole genome sequencing

DNA was extracted from established subclonal cell lines using the QIAGEN DNeasy blood & tissue kit. Libraries were prepared using the Roche KAPA PCR-free library preparation kit and genomes were sequenced on the HiSeq X platform to a depth of ~30x.

Whole genome analysis

Fastq files from the WGS were firstly aligned to mouse genome reference mm10. The output bam files were subsequently used for copy number analysis. Copy number analysis was performed using a R package cn.mops (*54*) in "paired mode" with a window length of 10kb. Reads were aligned to the BALB/c reference genome using BWA before being indexed and sorted with Novosort. Reads that mapped incompletely to the reference genome were then mapped to the barcode plasmid sequence with BWA and sorted and indexed with Novosort. Read pairs where only one pair mapped to the barcode plasmid sequence were blasted against mm10 to establish the barcode plasmid insertion site.

Flow cytometry for MHC1 and PD-L1

The 4T1 subclones (IE1, IE2, NT1, NT2), as well as the parental 4T1 bulk population were revived and passaged three times before being seeded into a 6 well plate at a density of 200 000 cells per well. At approximately 80% confluence, cells were collected into FACS buffer (DPBS supplemented with 2% FCS and 2% HEPEs) for flow cytometry. Cells were stained with a mastermix of APC conjugated anti-mouse CD274 (Biolegend) and Alexa Fluor488 conjugated anti-mouse H2-kD (Biolegend) at a final concentration of 1:200 for 20 minutes. Cells were washed three times with FACs buffer before being stained with DAPI and run on the BD FACSCanto II flow cytometer, utilising BD FACSDIVA software. Data was analysed in FlowJo (version 10.6.1) and median fluorescence intensity of live, single cells was calculated.

Treating cells with 5-Aza-2'-deoxycytidine and flow cytometry for MHC-I

5-Aza-2'-deoxycytidine (5aza) was sourced from Sigma and reconstituted in DMSO according to manufacturer's instructions. Subclones (IE1, IE2) and the parental 4T1 cell line were seeded into a 24 well plate at a density of 8000 cells per well in 4T1 media (previously described). Cells were allowed to settle overnight before being treated with 5aza at 200nM, 100nM or 50nM for 72 hours. 5aza was removed and cells were cultured in media only for 24 hours before being collected for flow analysis. Cells were stained with Alexa Fluor488 conjugated anti-mouse H2-kD (Biolegend) at a concentration of 1:200 in FACS buffer for 20 minutes. Cells were washed three times after staining before being stained with DAPI. Data was collected using the BC FACSCanto II flow cytometer with BD FACSDIVA software. The resulting data was analysed using FlowJo (version 10.6.1) and media fluorescence intensity of live, single cells was calculated.

Treating cells with interferon gamma and flow cytometry for MHC-I

Active mouse interferon gamma (IFNγ) was sourced from Abcam and reconstituted in sterile water, as per manufacturer's instructions. Subclones (IE1, IE2) and the parental 4T1 cell line were grown in a 24 well plate until approximately 70% confluence was achieved. Cells were then treated with IFNγ (100ng/ml) for 24 hours. Cells were stained with Alexa Fluor488 conjugated anti-mouse H2-kD (Biolegend) at a concentration of 1:200 in FACS buffer for 20 minutes. Cells were washed three times before being stained with DAPI. Data was generated using the BC FCSCanto II flow cytometer with BD FACSDIVA software. Analysis was carried out using FlowJo (version 10.6.1) and median fluorescence intensity of live single cells was calculated.

References:

1. D. Hanahan, Robert A. Weinberg, Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646-674 (2011).
2. R. D. Schreiber, L. J. Old, M. J. Smyth, Cancer Immunoediting: Integrating Immunity's Roles in Cancer Suppression and Promotion. *Science* **331**, 1565-1570 (2011).
3. M. Angelova *et al.*, Evolution of Metastases in Space and Time under Immune Selection. *Cell* **175**, 751-765.e716 (2018).

4. S. C. Casey *et al.*, MYC regulates the antitumor immune response through CD47 and PD-L1. *Science* **352**, 227-231 (2016).
5. D. M. Pardoll, The blockade of immune checkpoints in cancer immunotherapy. *Nature Reviews Cancer* **12**, 252-264 (2012).
6. P. Sharma, S. Hu-Lieskovan, J. A. Wargo, A. Ribas, Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* **168**, 707-723 (2017).
7. L. Galluzzi, T. A. Chan, G. Kroemer, J. D. Wolchok, A. López-Soto, The hallmarks of successful anticancer immunotherapy. *Science Translational Medicine* **10**, (2018).
8. P. Schmid *et al.*, Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *New England Journal of Medicine* **379**, 2108-2121 (2018).
9. H. R. Ali, L. Chlon, P. D. P. Pharoah, F. Markowetz, C. Caldas, Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. *PLOS Medicine* **13**, e1002194 (2016).
10. E. T. E. Gross *et al.*, Immunosurveillance and immunoediting in MMTV-PyMT-induced mammary oncogenesis. *Oncoimmunology* **6**, e1268310 (2017).
11. C. M. Koebel *et al.*, Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* **450**, 903-907 (2007).
12. I. Milo *et al.*, The immune system profoundly restricts intratumor genetic heterogeneity. *Science Immunology* **3**, (2018).
13. H. Matsushita *et al.*, Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* **482**, 400-404 (2012).
14. K.-J. Malmberg *et al.*, Natural killer cell-mediated immunosurveillance of human cancer. *Seminars in Immunology* **31**, 20-29 (2017).
15. K. A. Lawson *et al.*, Functional genomic landscape of cancer-intrinsic evasion of killing by T cells. *Nature*, (2020).
16. R. T. Manguso *et al.*, In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature* **547**, 413 (2017).
17. C.-Q. Song *et al.*, Genome-Wide CRISPR Screen Identifies Regulators of Mitogen-Activated Protein Kinase as Suppressors of Liver Tumors in Mice. *Gastroenterology* **152**, 1161-1173.e1161 (2017).
18. G. Wang *et al.*, CRISPR-GEMM Pooled Mutagenic Screening Identifies KMT2D as a Major Modulator of Immune Checkpoint Blockade. *Cancer Discovery* **10**, 1912-1933 (2020).
19. S. J. Patel *et al.*, Identification of essential genes for cancer immunotherapy. *Nature* **548**, 537-542 (2017).
20. N. Anfossi *et al.*, Human NK Cell Education by Inhibitory Receptors for MHC Class I. *Immunity* **25**, 331-342 (2006).
21. E. Mamessier *et al.*, Human breast cancer cells enhance self tolerance by promoting evasion from NK cell antitumor immunity. *The Journal of Clinical Investigation* **121**, 3609-3622 (2011).
22. H.-e. C. Bhang *et al.*, Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med* **21**, 440-448 (2015).
23. E. Wagenblast *et al.*, A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. *Nature* **520**, 358-362 (2015).

24. A. Gambotto *et al.*, Immunogenicity of enhanced green fluorescent protein (EGFP) in BALB/c mice: identification of an H2-Kd-restricted CTL epitope. *Gene Therapy* **7**, 2036-2040 (2000).
25. Y. Lan *et al.*, Enhanced preclinical antitumor activity of M7824, a bifunctional fusion protein simultaneously targeting PD-L1 and TGF-β. *Science Translational Medicine* **10**, (2018).
26. M. Ouzounova *et al.*, Monocytic and granulocytic myeloid derived suppressor cells differentially regulate spatiotemporal tumour plasticity during metastatic cascade. *Nature Communications* **8**, 14979 (2017).
27. X. Peng, L. Zha, A. Chen, Z. Wang, HOXA5 is a tumor suppressor gene that is decreased in gastric cancer. *Oncol Rep* **40**, 1317-1329 (2018).
28. R. E. Vance, J. R. Kraft, J. D. Altman, P. E. Jensen, D. H. Raulet, Mouse CD94/NKG2A Is a Natural Killer Cell Receptor for the Nonclassical Major Histocompatibility Complex (MHC) Class I Molecule Qa-1b. *Journal of Experimental Medicine* **188**, 1841-1848 (1998).
29. M. M. S. Obradović *et al.*, Glucocorticoids promote breast cancer metastasis. *Nature* **567**, 540-544 (2019).
30. D. Pan, M. Kocherginsky, S. D. Conzen, Activation of the Glucocorticoid Receptor Is Associated with Poor Prognosis in Estrogen Receptor-Negative Breast Cancer. *Cancer Research* **71**, 6360-6370 (2011).
31. N. Luo *et al.*, DNA methyltransferase inhibition upregulates MHC-I to potentiate cytotoxic T lymphocyte responses in breast cancer. *Nature Communications* **9**, 248 (2018).
32. D. Blanco-Melo *et al.*, Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045.e1039 (2020).
33. C. Curtis *et al.*, The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-352 (2012).
34. N. The Cancer Genome Atlas, Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61 (2012).
35. S. Dushyanthen *et al.*, Relevance of tumor-infiltrating lymphocytes in breast cancer. *BMC Medicine* **13**, 202 (2015).
36. M. Obeid *et al.*, Calreticulin exposure dictates the immunogenicity of cancer cell death. *Nature Medicine* **13**, 54-61 (2007).
37. N. Ishikawa *et al.*, Characterization of SEZ6L2 cell-surface protein as a novel prognostic marker for lung cancer. *Cancer Science* **97**, 737-745 (2006).
38. L. Wang *et al.*, Upregulated Seizure-Related 6 Homolog-Like 2 Is a Prognostic Predictor of Hepatocellular Carcinoma. *Disease Markers* **2020**, 7318703 (2020).
39. P. Jiang *et al.*, Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nature Medicine* **24**, 1550-1558 (2018).
40. A. W. Zhang *et al.*, Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer. *Cell*, (2018).
41. N. McGranahan *et al.*, Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463-1469 (2016).
42. Y. Wolf *et al.*, UVB-Induced Tumor Heterogeneity Diminishes Immune Response in Melanoma. *Cell* **179**, 219-235.e221 (2019).

43. K. B. Chiappinelli, C. A. Zahnow, N. Ahuja, S. B. Baylin, Combining Epigenetic and Immunotherapy to Combat Cancer. *Cancer Research* **76**, 1683-1689 (2016).
44. M. Guo, Y. Peng, A. Gao, C. Du, J. G. Herman, Epigenetic heterogeneity in cancer. *Biomarker Research* **7**, 23 (2019).
45. Y. Chen *et al.*, Sephin1, which prolongs the integrated stress response, is a promising therapeutic for multiple sclerosis. *Brain* **142**, 344-361 (2019).
46. D. R. Leach, M. F. Krummel, J. P. Allison, Enhancement of Antitumor Immunity by CTLA-4 Blockade. *Science* **271**, 1734-1736 (1996).
47. V. P. Baklaushev *et al.*, Luciferase Expression Allows Bioluminescence Imaging But Imposes Limitations on the Orthotopic Mouse (4T1) Model of Breast Cancer. *Scientific Reports* **7**, 7715 (2017).
48. K. Wang, C. A. Phillips, A. M. Saxton, M. A. Langston, EntropyExplorer: an R package for computing and comparing differential Shannon entropy, differential coefficient of variation and differential expression. *BMC Research Notes* **8**, 832 (2015).
49. Y. Akimov, D. Bulanova, S. Timonen, K. Wennerberg, T. Aittokallio, Improved detection of differentially represented DNA barcodes for high-throughput clonal phenomics. *Molecular Systems Biology* **16**, e9195 (2020).
50. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2012).
51. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
52. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2009).
53. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols* **4**, 1184-1191 (2009).
54. G. Klambauer *et al.*, cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Research* **40**, e69-e69 (2012).

Figure legends:

Fig 1: Experimental workflow schematic. Barcode library is introduced into mammary carcinoma cell lines *in vitro* at a low multiplicity of infection (MOI). Cells are sorted based on RFP expression to select for those that have incorporated a barcode. Barcoded cells are then transplanted into the mammary fat pad. Following immunoselection either with the endogenous immune system or immunotherapy then barcode abundance and diversity are analysed within the primary tumours and metastatic lungs.

Fig 2: Immune selection and clonal immunoediting of EMT6 primary tumours. A. Outline of experimental design. B. EMT6 primary tumour growth plotted as tumour volume comparing wild-type Balb/c mice with immunocompromised NSG mice. Average volume +/- SEM; n=5-6 mice per group. C. Kaplan-Meier survival analysis comparing Balb/c and NSG EMT6 tumour bearing mice (Mantel-Cox p=0.009). D. EMT6 primary tumour growth plotted as tumour volume comparing

wild-type Balb/c mice with or without immunotherapy (anti-PD1+anti-CTLA4) IP on days 10, 12, 14, and 17 indicated with red arrows. Average volume +/- SEM; n=5-6 mice. E. Kaplan-Meier survival analysis comparing immunotherapy treated and isotype control treated Balb/c EMT6 tumour bearing mice (Mantel-Cox p=0.0006). F. Number of unique barcodes identified in EMT6 primary tumours grown in NSG mice, and Balb/c mice treated with isotype control antibodies or anti-PD1 + anti-CTLA4. Unpaired t-test. G. Shannon diversity index analysis comparing NSG mice with Balb/c mice treated with isotype control antibodies or anti-PD1 + anti-CTLA4. Unpaired t-test. H. Dot plot of a subset of barcodes with interesting patterns of enrichment.

Fig 3: Immune selection and clonal immunoediting in the metastatic setting utilising the 4T1 model.
A. Outline of experimental design. B. Kaplan-Meier survival analysis comparing Balb/c and NSG 4T1 tumour bearing mice; n=5 mice/group (p = 0.0002, Mantel-Cox Log-rank test). C. Kaplan-Meier survival analysis comparing immunotherapy treated and isotype control treated Balb/c 4T1 tumour bearing mice. IP on days 15, 17, 19, and 21 indicated with red arrows; n=5-6 mice/group (p = 0.0141, Mantel-Cox Log-rank test). D. Number of unique barcodes identified in 4T1 primary tumours and lung metastases grown in NSG mice or Balb/c mice. Unpaired t-test. E. Number of unique barcodes identified in 4T1 primary tumours and lung metastases grown in Balb/c mice treated with isotype control antibodies or anti-PD1 + anti-CTLA4. Unpaired t-test. F. Shannon diversity index analysis comparing primary tumours from NSG mice with Balb/c mice and lung metastases from NSG mice, Balb/c mice, isotype control treated Balb/c mice, and anti-PD1 + anti-CTLA4 treated Balb/c mice. Unpaired t-test. G. Analysis of number of unique barcodes in the metastatic lungs of mice treated with control antibodies, or anti-CD8a, or anti-asialo GM1. CD8 T cells depleted with anti-CD8a, NK cells depleted with anti-asialo GM1. Unpaired t-test, 5 mice/group.

Fig 4: Analysis of specific barcodes enriched or depleted by the immune system and immunotherapy in the 4T1 model.
A. Unsupervised hierarchical clustering heatmap of barcodes with an abundance of above 5% in at least one 4T1 primary tumour or lung metastatic sample; * indicate barcodes detected at a frequency above 0.1% in a particular sample. B. Dot plot of a subset of specific barcodes. C. Fishplot of the nine most abundant barcodes detected in lungs of WT and NSG mice each barcode is given a unique colour with the remaining barcodes being combined and being represented in grey. D. Fishplot of the nine most abundant barcodes detected in the lungs of WT mice treated with combination immunotherapy or control antibodies, a bottleneck has been introduced between primary tumour and lung metastases to depict transition through the blood stream.

Fig 5: Gene expression analysis of immunotherapy resistant clones.
A. Volcano plot showing differentially expressed genes between parental 4T1 cells and the immunotherapy enriched 1 (IE1) clone. B. Volcano plot showing differentially expressed genes between parental 4T1 cells and the immunotherapy enriched 2 (IE2) clone. C. Expression of MHC related genes co-ordinately down-regulated in clone IE1, measured as transcripts per million (TPM) in the parental

4T1 population and in the specific cell clones; unpaired t-test. D. MHC1 protein expression as quantified by flow cytometry in the specific clones and the parental 4T1 population measured as mean fluorescence intensity (MFI); representative histogram showing each immune enriched clone and the parental cells on the right; unpaired t-test. E. Expression of immune related genes up-regulated in clone IE2, measured as transcripts per million (TPM) in the parental 4T1 population and in the specific cell clones; unpaired t-test. F. PD-L1 protein expression as determined by flow cytometry in the specific clones and the parental 4T1 population measured as mean fluorescence intensity (MFI), representative plot of three independent experiments; unpaired t-test.

Fig 6. Overlapping gene signatures of the immunotherapy resistant clones show prognostic significance in basal-like breast cancer patients.
A. Upset plots and Venn diagrams showing the overlap in significantly up-regulated (right) and down-regulated (left) genes between the two immunotherapy enriched clones and two other control clones. B. Heatmap of the top 50 overlapping up regulated and down regulated genes between the IE1 and IE2 clones across all the clonal cell lines. C. Kaplan-Meier survival analysis of basal-like breast cancer patients from the METABRIC cohort, examining overall survival with patients split based on the top 30% and bottom 30% expression of the overlapping upregulated 25 gene signature. D. Kaplan-Meier survival analysis of basal-like breast cancer patients from the TCGA cohort, examining overall survival with patients split based on the top 30% and bottom 30% expression of the overlapping upregulated 25 gene signature. The Cox proportional hazards model was used to compute Hazard Ratios. Significance between stratification groups were computed using log-rank test statistics.

Supplementary Figure Legends:

Sup Fig1. 4T1 primary tumour growth or barcode diversity unchanged in immunocompromised mice. A. Primary tumour mass was slightly increased in immunocompromised NSG mice. B. Total number of unique barcodes identified in the primary tumours of Balb/c or NSG mice. C. Shannon diversity analysis comparing primary tumours from Balb/c and NSG mice.

Sup Fig2. Depletion of CD8+ T cells or NK cells does not significantly affect survival in 4T1 tumour bearing mice. CD8 T cells depleted with anti-CD8a, NK cells depleted with anti-asialo GM1. A. Kaplan-Meier survival analysis of Balb/c mice transplanted with 4T1 cells, primary tumour was resected on day 15. Cell depletion was initiated 1 day prior to resection on day 14.

Sup Fig3. Analysis of higher complexity barcode library in the 4T1 cells replicates changes in barcode proportions that were seen in the ~5000 barcode library. A. Number of unique barcodes identified in 4T1 primary tumours and lung metastases grown in NSG mice or Balb/c mice. Unpaired t-test. 5 mice/group. B. Number of unique barcodes identified in 4T1 primary tumours and lung metastases grown in Balb/c mice treated with isotype control antibodies or anti-PD1 + anti-CTLA4. Unpaired t-test. 5 mice/group.

Sup Fig 4. 4T1 clonal cell lines growth kinetics *in vitro*. Growth kinetics as measured by percentage confluence over time.

Sup Fig 5. A. No genomic loss was observed in any of the clones at the MHC-I locus on chromosome 17. B. Gene expression changes in genes that have a single copy number increase were observed in IE1 and IE2 but not NT2.

Sup Fig 6. MHC regulation by 5-aza and IFN-γ in the clonal cell lines. A. 5-aza treatment of clonal cell lines increases MHC-I protein expression in a dose dependent manner in all cell lines as measured by flow cytometry. Unpaired t-test. B. IFN-γ treatment of clonal cell lines increases MHC-I protein expression in all cell lines.

Sup Fig 7. Gene set enrichment analysis of the overlapping genes between the IE1 and IE2 clones. Red indicates gene sets with a significant

Supplementary Table Legends:

Sup Table 1: DNA barcode insertion sites of IE1 and IE2. By searching the whole genome sequencing (WGS) data and identifying read pairs where only one read mapped to the barcode sequence, the matching mate was blasted against mm10 to identify the barcode plasmid insertion site. Both barcodes were found to reside in introns of cancer-related genes but there was no difference in expression of these genes detected.

Sup Table 2: Copy number variation (CNV) locations found within all subclones. By analysing whole genome sequencing data in R using the cn.mops pacakges, copy number variations could be determined. No major copy number aberrations were detected across the clones, although a single copy number gain was detected in IE1, IE2 and NT2.

Sup Table 3. List of genes differentially expressed in IE1 compared to parental 4T1 bulk. Differentially expressed genes (DEGs) were generated by analyzing bulk RNA sequencing data using R and the EdgeR package. DEGs were filtered for significance based on a FDR <0.05.

Sup Table 4. List of genes differentially expressed in IE2 compared to parental 4T1 bulk. Differentially expressed genes (DEGs) were generated by analyzing bulk RNA sequencing data using R and the EdgeR package. DEGs were filtered for significance based on a FDR <0.05. A greater number of differentially expressed genes were detected in IE2 than IE1.

Sup Table 5. All significantly enriched gene sets found in IE1. Differentially expressed genes generated from comparing IE1 to bulk were preranked by fold change before searching for gene set enrichment using the Molecular Signature Database (MSigDB) across all available collections. Gene sets were filtered for significance based on a FDR<0.05.

Sup Table 6. All significantly enriched gene sets found in IE2. Differentially expressed genes generated from comparing IE2 to bulk were preranked by fold change. Gene set enrichment was carried out using Molecular Signature Database (MSigDB) across all available collections. Gene sets were filtered for significance based on a FDR<0.05

Sup Table 7: List of common differentially expressed genes found in IE1 and IE2. Commonly differentially expressed genes (DEGs) were determined by overlapping significant DEGs in IE1 and IE2 by gene name. An average fold change was calculated across the two samples.

Sup Table 8: Top significantly enriched gene sets found from common differentially expressed genes in IE1 and IE2. The fold change of the common differentially expressed genes in IE1 and IE2 were averaged together to generate an average fold change across both IE1 and IE2. The gene list was then preranked before searching for gene set enrichment using Molecular Signature Database (MSigDB) across the C2 All collection. Gene sets were filtered for significance based on a FDR<0.05. The majority of significant gene sets that were returned were negatively enriched in IE1 and IE2.

Figure 1

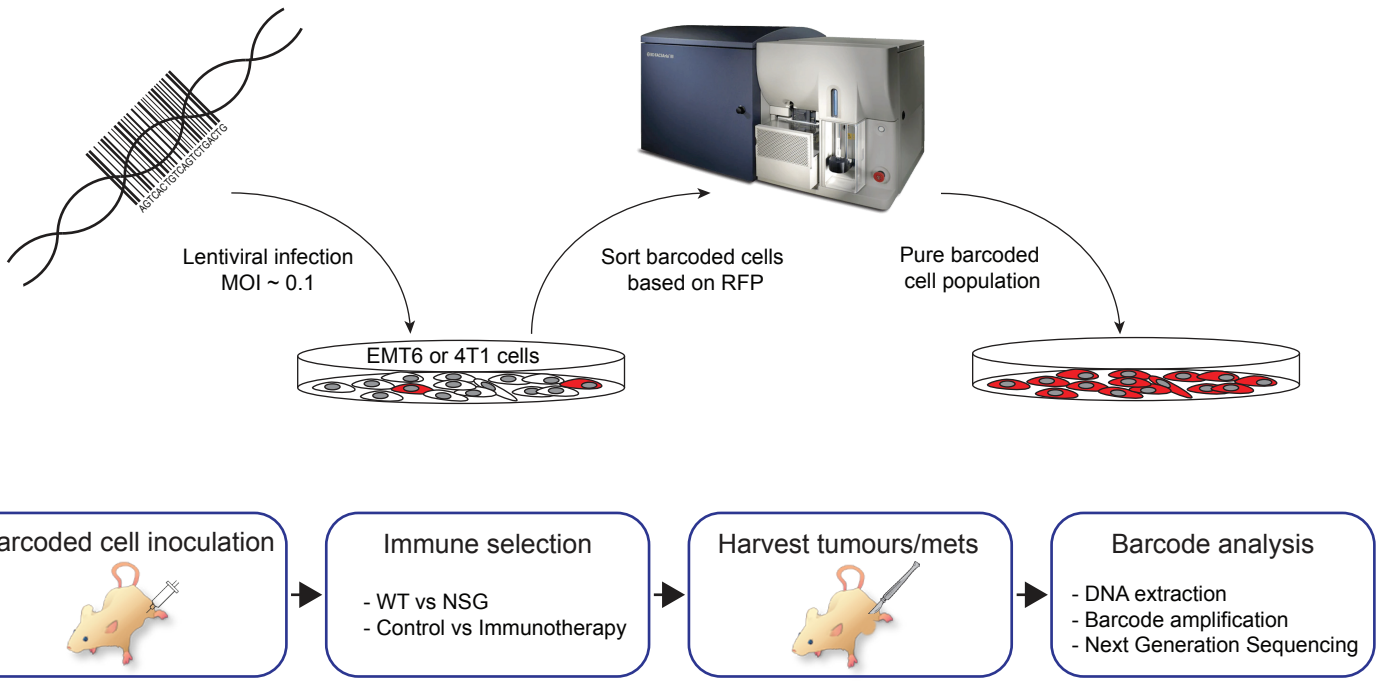


Figure 2

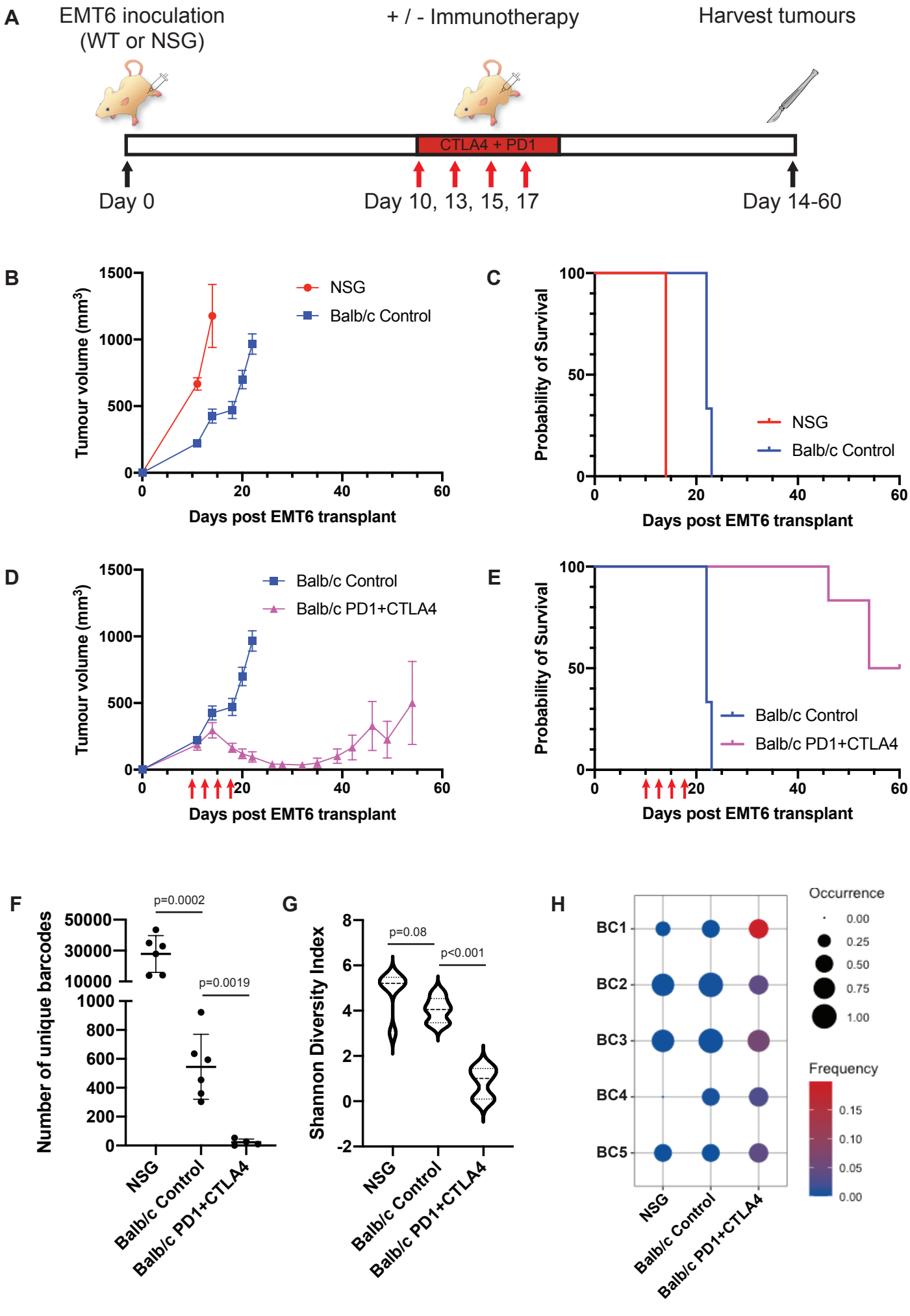
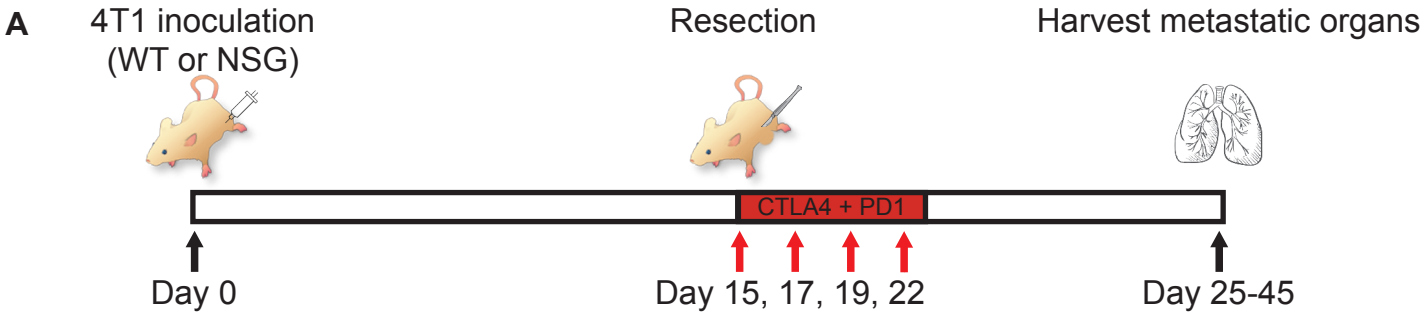
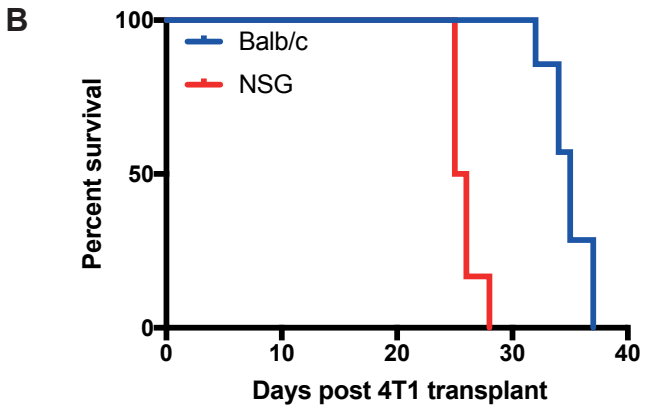
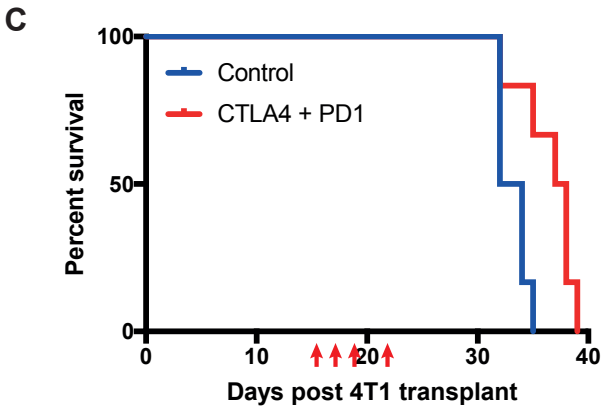
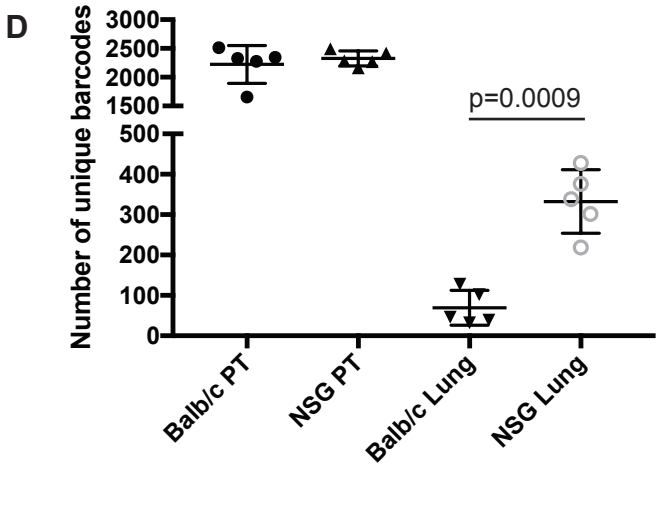


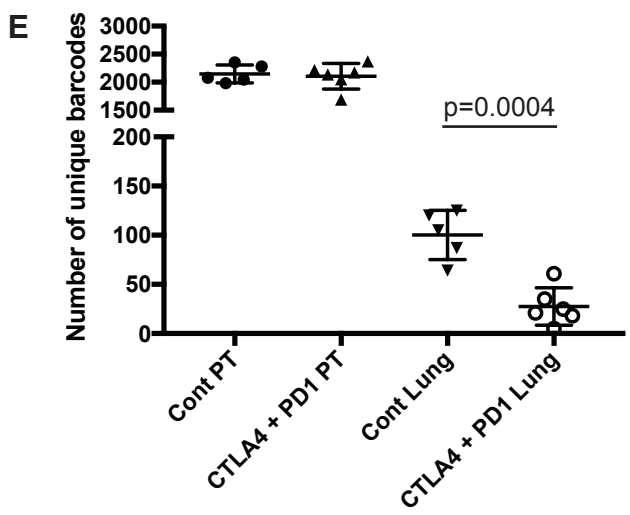
Figure 3

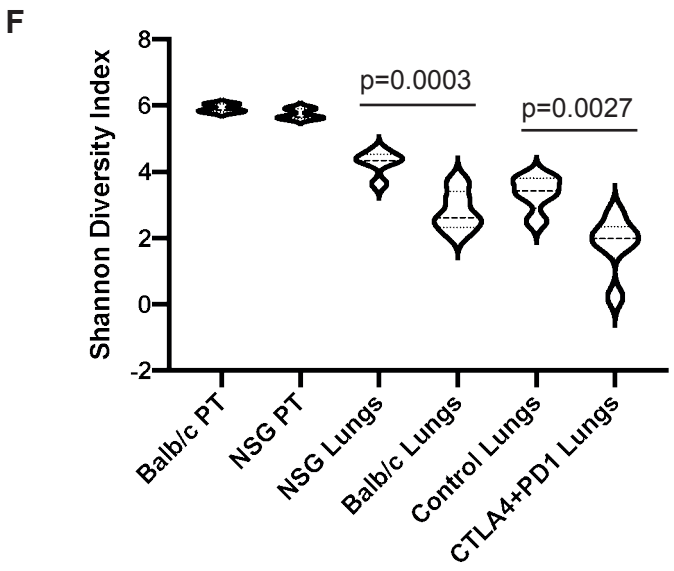












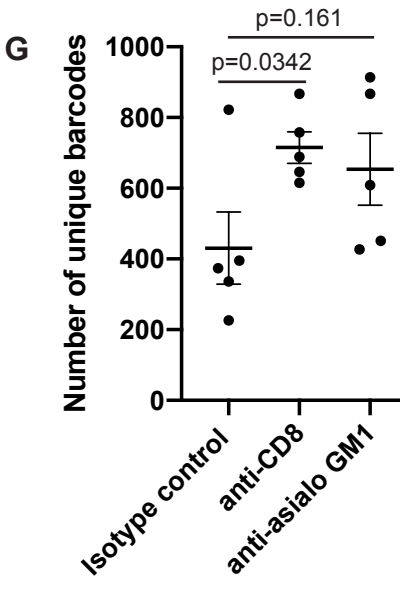


Figure 4

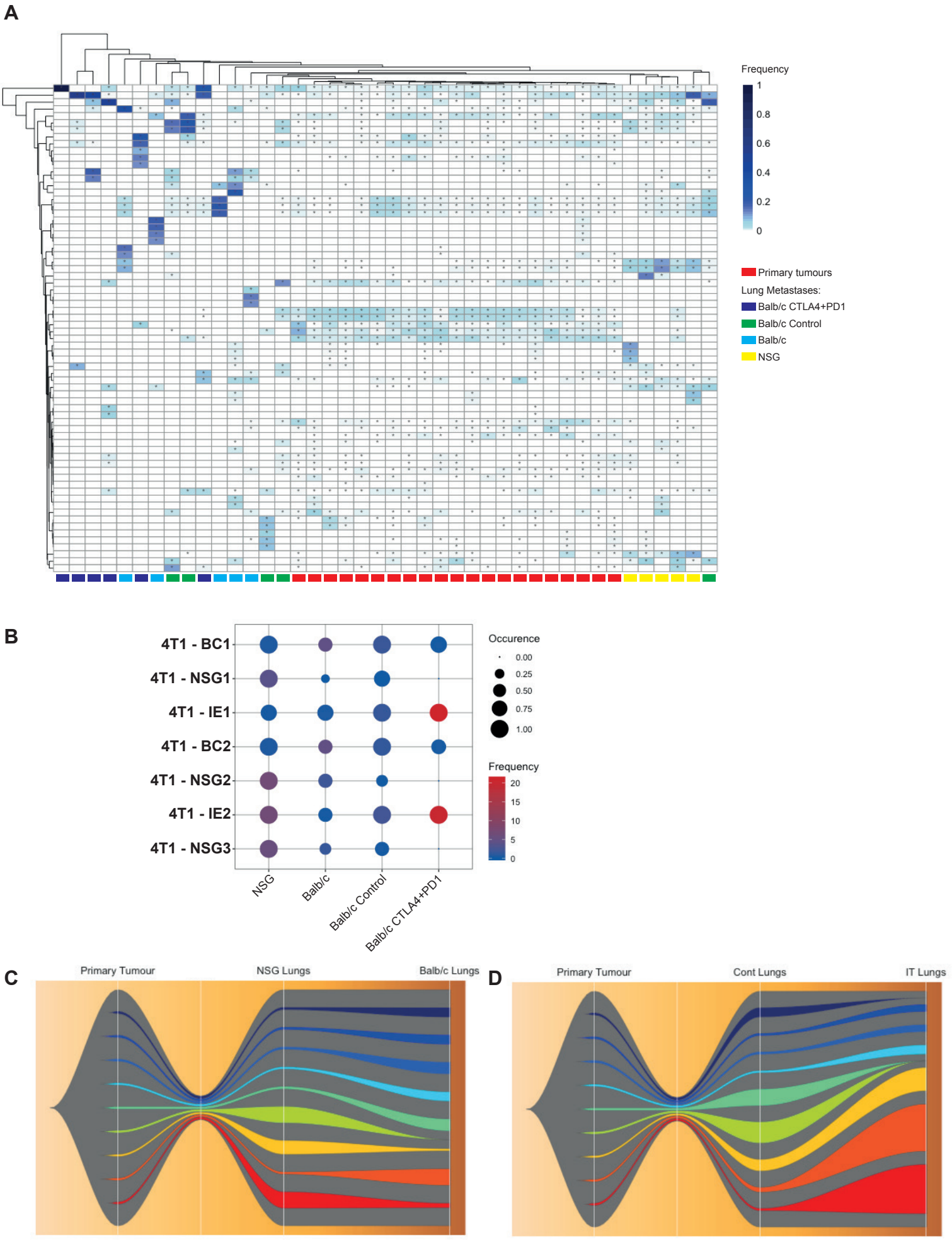


Figure 5

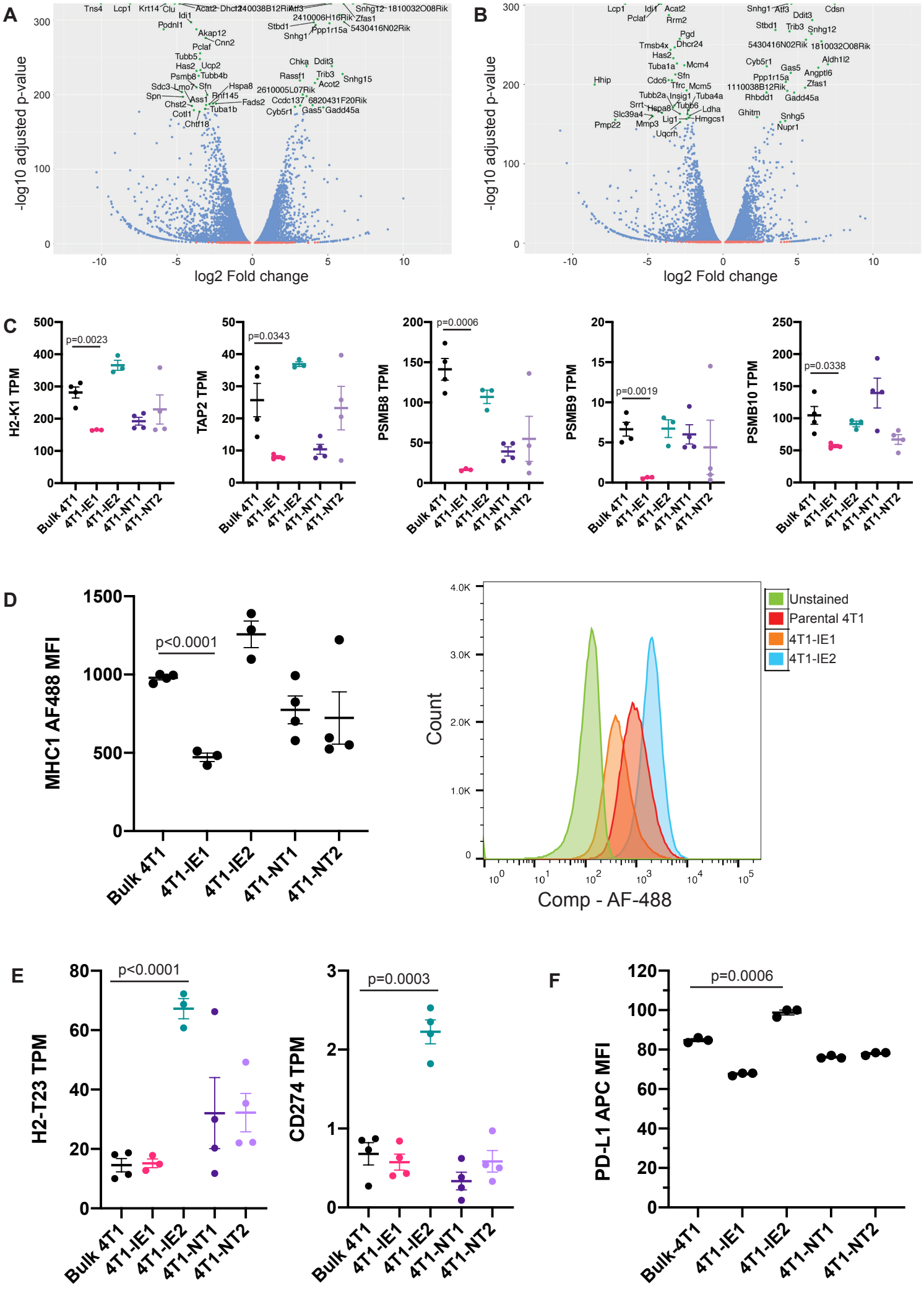
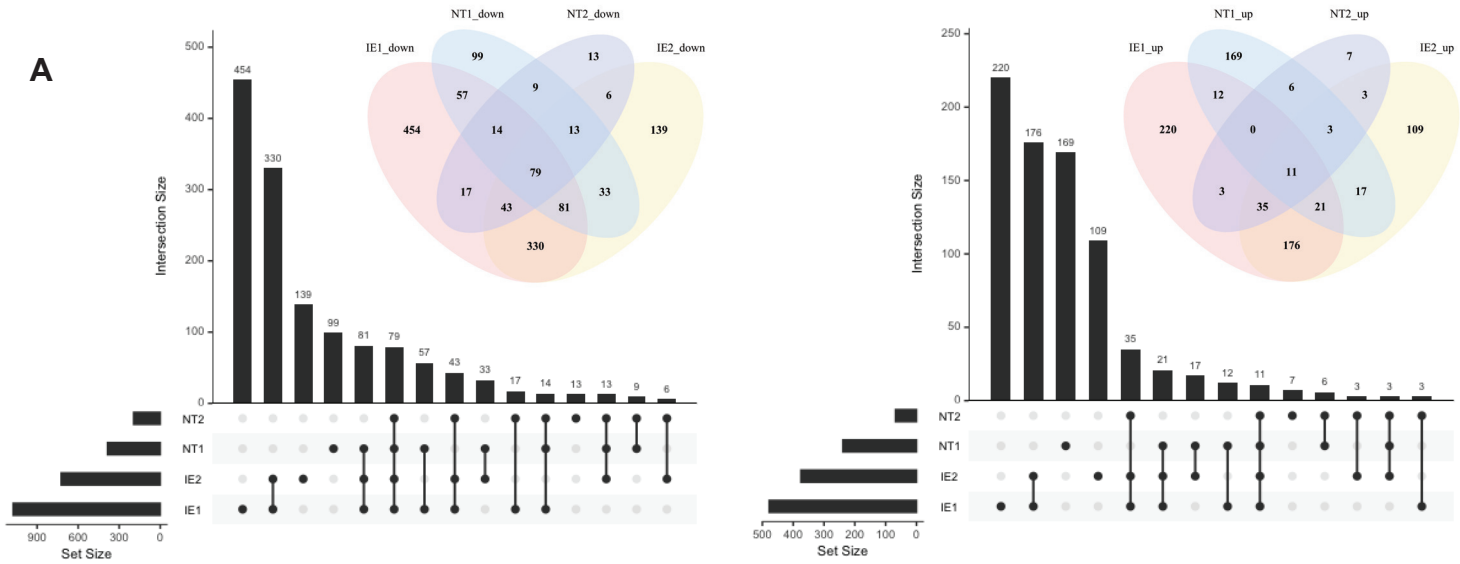
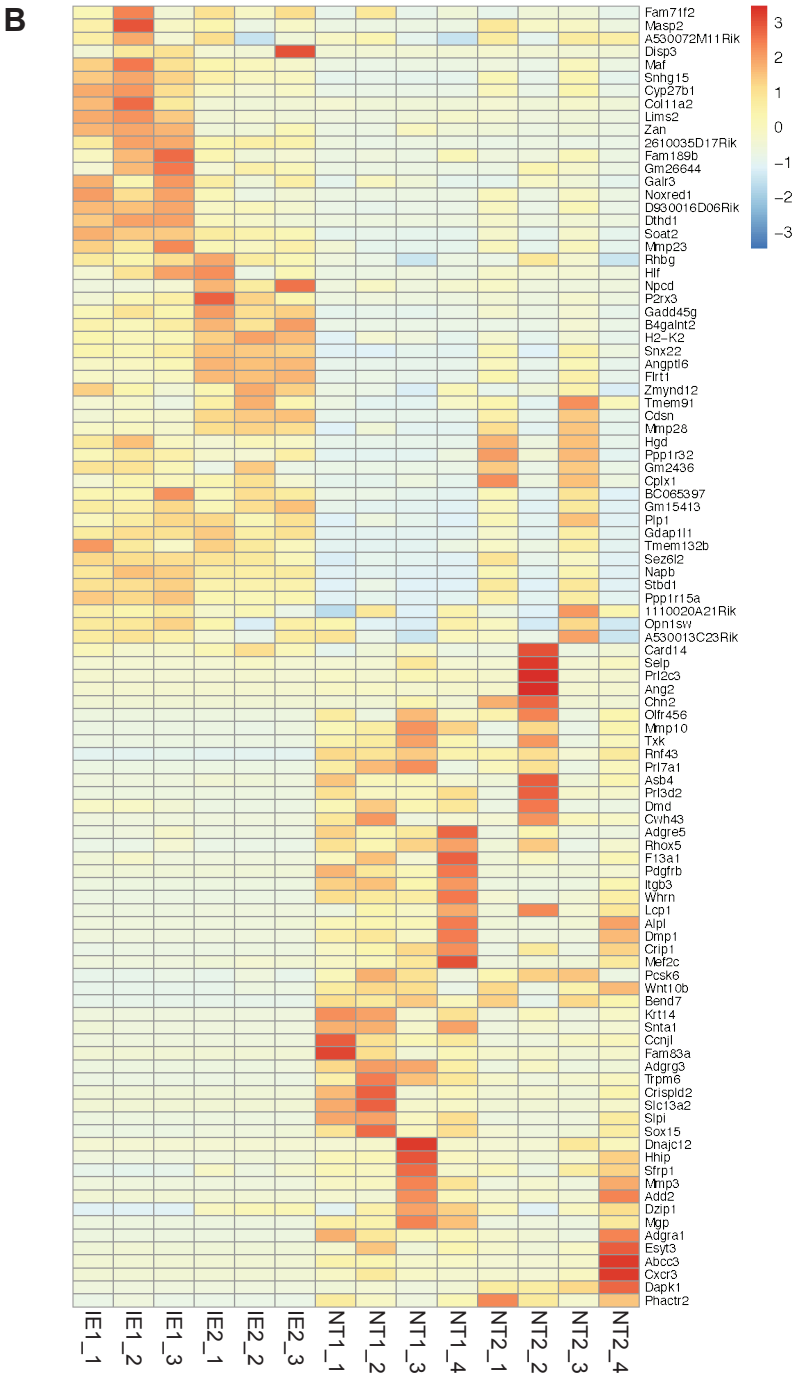
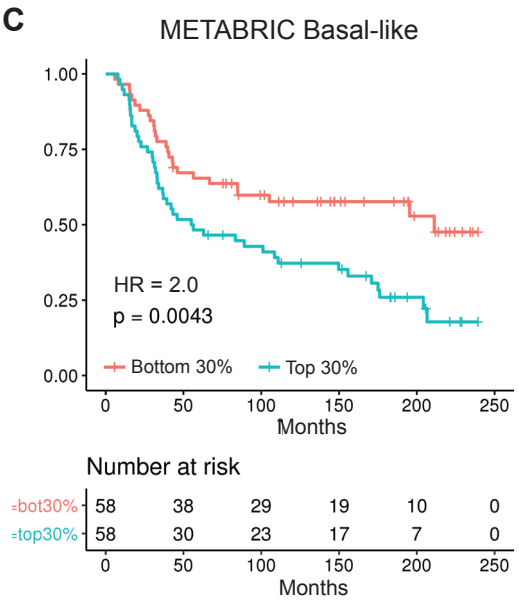
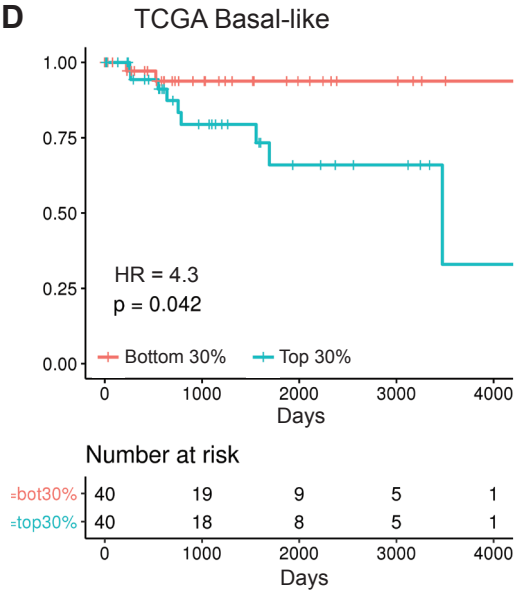


Figure 6

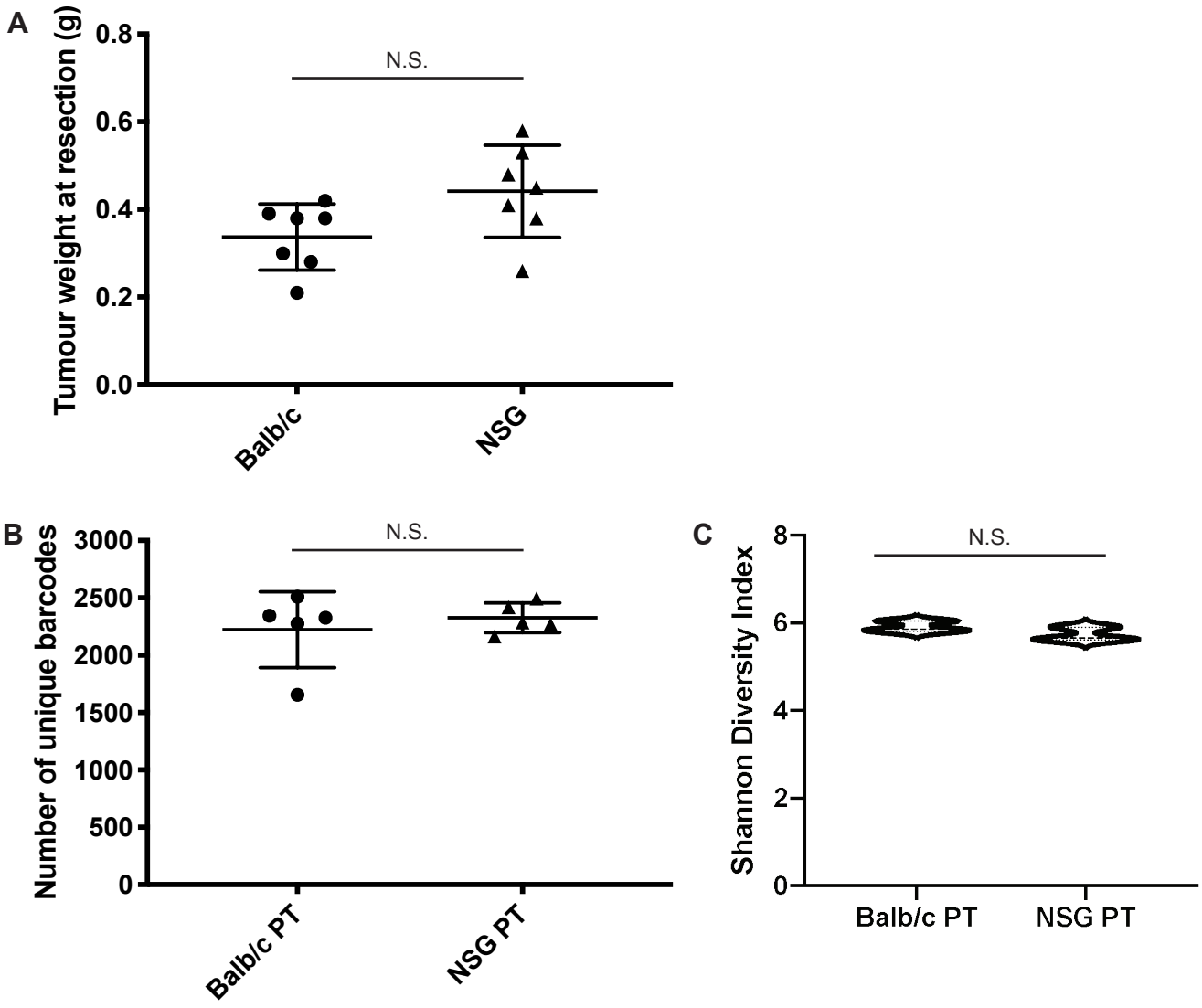




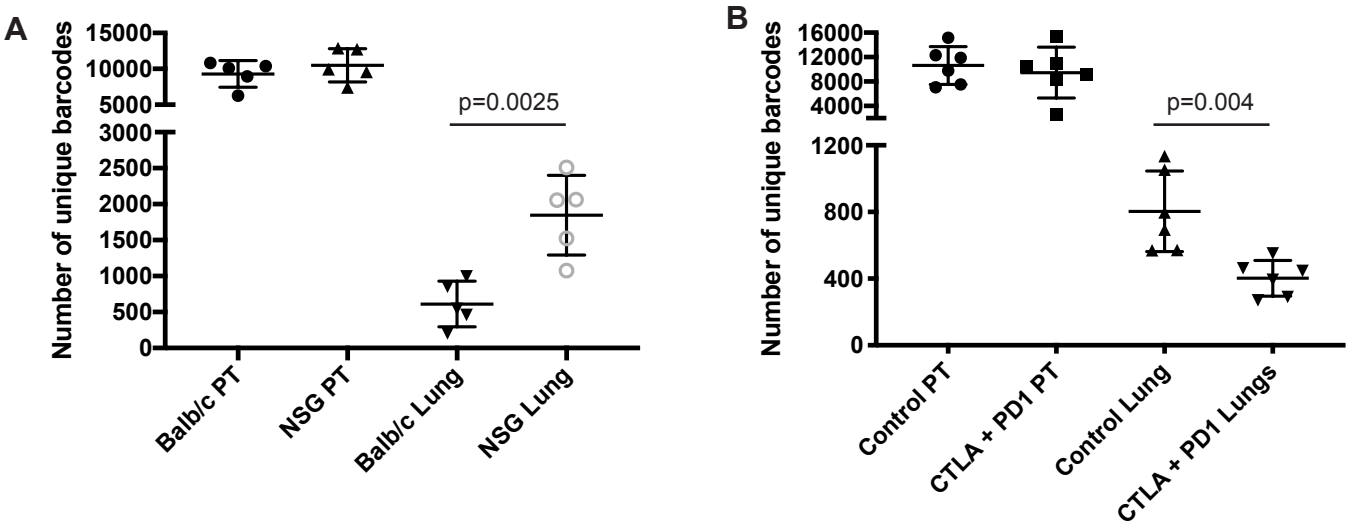




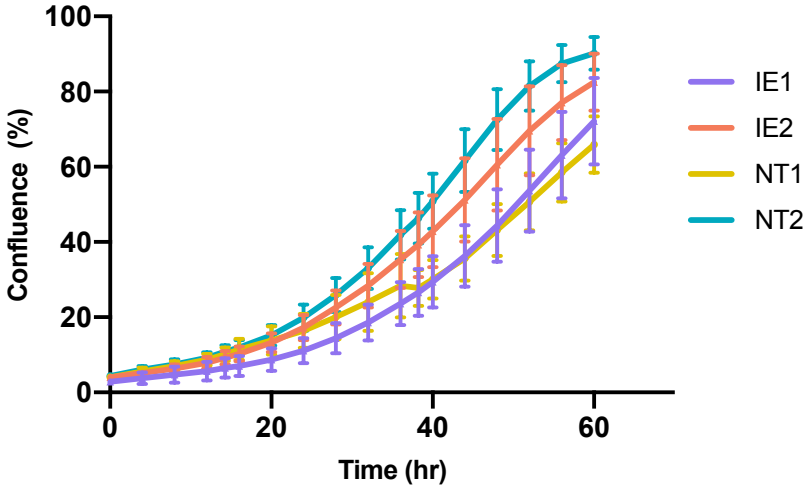
Sup Figure 1



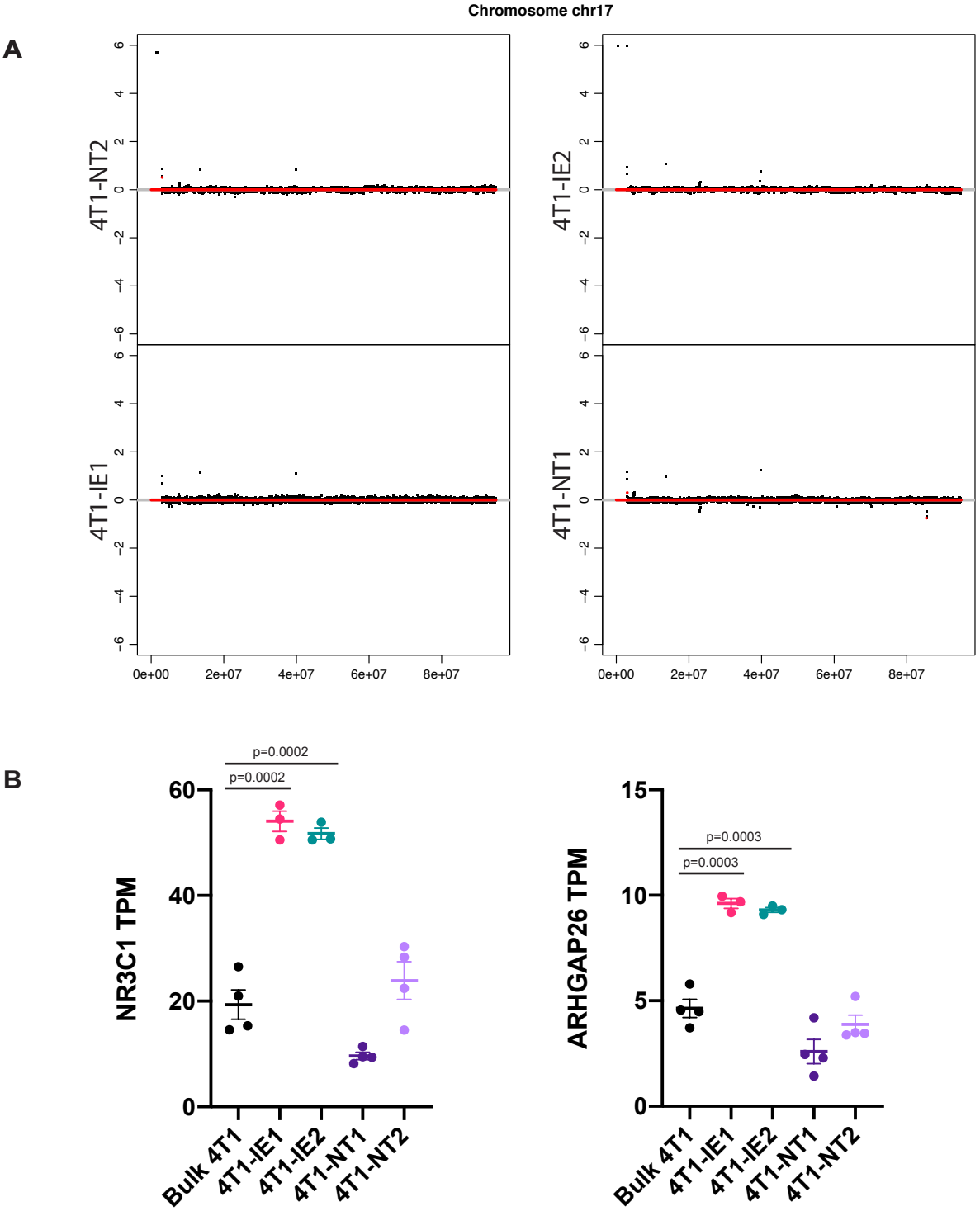
Sup Figure 3



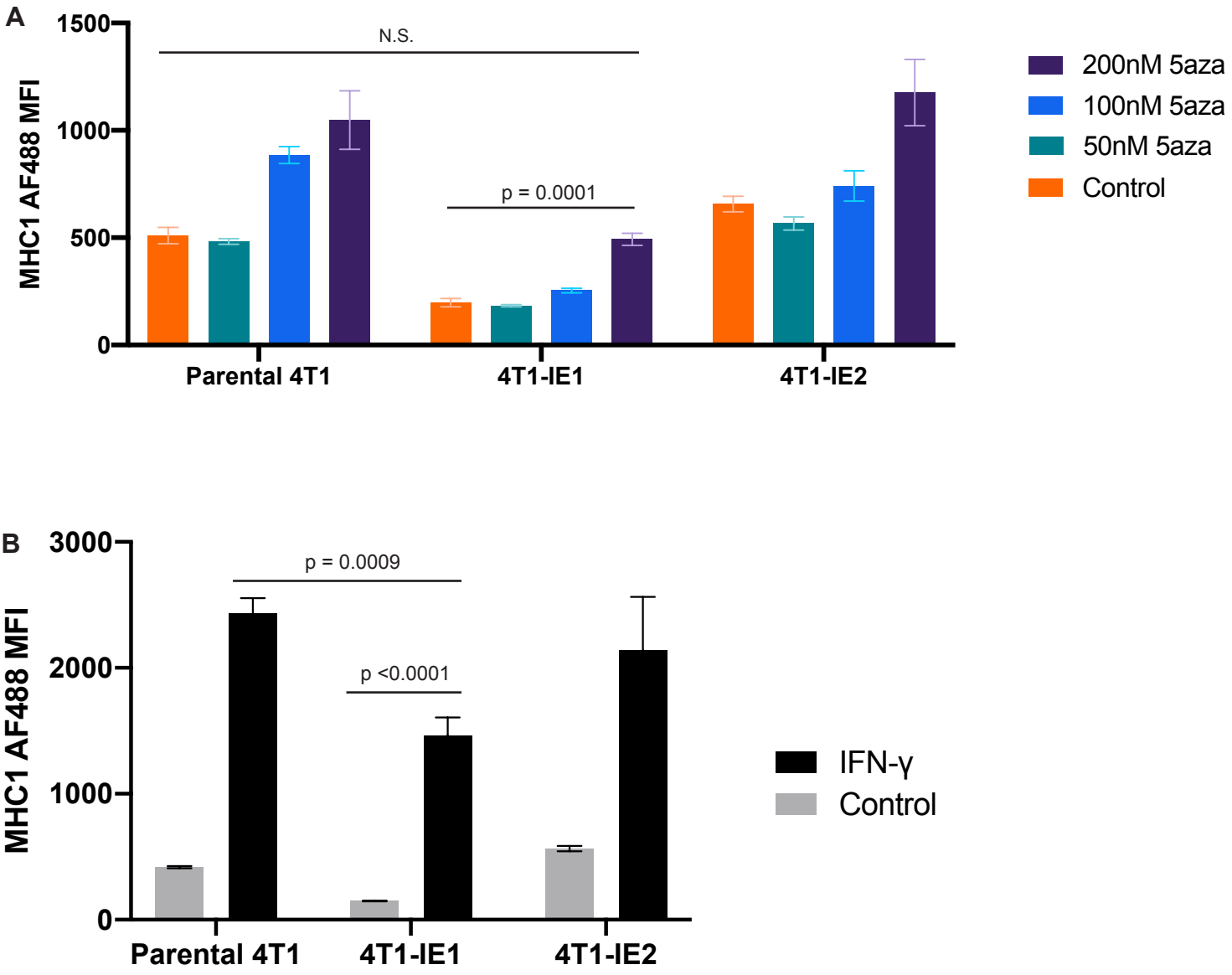
Sup Figure 4



Sup Figure 5



Sup Figure 6



Sup Figure 7