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], Damien J. Zanker[3,4], Belinda S. Parker[3,4], 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) School of Clinical Medicine, Faculty of Medicine, UNSW Sydney, NSW 2052, Australia

(3) Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville 3010, Australia

(4) Cancer Immunology and Therapeutics Programs, Peter MacCallum Cancer Centre, Melbourne 3000, Australia

*Corresponding authors: s.junankar@garvan.org.au, a.swarbrick@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 immune checkpoint inhibitors have been effective for reactivating tumour immunity in some types of cancer, many other 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 in both primary tumours and metastases. These findings demonstrate that immunoediting selects for pre-existing breast cancer cell populations and that immunoediting is not static, it is ongoing during metastasis and immunotherapy treatment. Isolation of immunotherapy resistant clones revealed unique and overlapping transcriptional signatures. The overlapping gene signature was associated with poor survival of basal-like breast cancer patients in two cohorts. At least one of these overlapping genes has an existing small molecule that can potentially be used to 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, 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, which 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 cancer cells. The prominent clinically approved immunotherapies for solid tumours target T cell checkpoint molecules (eg. anti-CTLA4 and anti-PD1) to overcome T cell

50 exhaustion *(5, 6)*. In select cancer types, such as melanoma, immune checkpoint
51 inhibitors have dramatic effects in a large proportion of patients *(7)*. Unfortunately for
52 metastatic breast cancer, few patients, even those having the most sensitive basal-like
53 breast cancer, had durable responses in clinical trials *(8)*. This indicates that in
54 metastatic breast cancer, resistance is either pre-existing or can rapidly develop to anti-
55 PD1/PDL1 therapy and suggests that alternate immune drug targets are needed for
56 breast cancer.

57 Although the immune system is known to play a role in breast cancer outcome *(9)* and
58 immunoediting can occur in a transgenic mouse model of breast cancer *(10)*, very little
59 is known about immune evasion by breast cancer cells. The majority of studies
60 examining immune evasion by cancers were performed using the highly mutated,
61 methylcholanthrene-driven sarcoma model, the response of which cannot be tracked at
62 the clonal level (*11*), or colon cancer (*12*). Of interest, a recent study suggests that
63 immunoediting by T cells can occur at the clonal level by demonstrating the selection
64 of clones that contain less-immunogenic fluorophores *(13)*. This leaves an important
65 gap in our collective knowledge as to the mechanisms employed in less immunogenic
66 tumours such as breast.

67 Both natural killer (NK) cells and T cells have been demonstrated to play a role in
68 immunoediting *(14, 15)*). However, the majority of recent research has focussed on
69 pathways relevant to T cell recognition *(16-20)*. Downregulation of MHC is one
70 mechanism by which cancer clones become impervious to T cells *(21)*, but this
71 inherently makes them targets of NK activity. In breast cancer dysfunction of NK cells
72 is noted and this is regulated by microenvironmental factors *(22)*. Data on resistance
73 pathways that allow for immune evasion from both T cells and NK cells are currently
74 more limited.

75 Intratumoural heterogeneity (ITH) has been identified as a major contributor to
76 treatment response. Prior work in non-small cell lung cancer (NSCLC) demonstrated
77 that neoantigen heterogeneity, and tumour mutation burden more broadly, is strongly
78 associated with T cell anticancer immune responses (*23, 24*). Chemotherapy(*25*) or loss
79 of HLA (*26*) have also been shown to increase ITH, which in turn was associated with
80 treatment relapse in NSCLC. Similar results have been described in cell line models of
81 melanoma (*27*), with Williams and colleagues providing recent evidence of ITH
82 enabling clonal cooperatively and immune escape in melanoma-derived cell lines (*28*).
83 While these previous studies have been crucial to our understanding of ITH and anti-
84 tumour immune responses and have provided undeniable evidence associating
85 increased ITH with therapy resistance, studies of this nature have not yet been carried
86 out in breast cancer. Furthermore, direct evidence of ITH enabling immune escape and
87 immunotherapy resistance in vivo has not been reported.

88 DNA barcoding allows tracking of individual clones and clonal expansion while
89 avoiding introducing potentially immunogenic proteins *(29)*. We used DNA barcoding
90 to analyse immunoediting *in vivo* at primary tumours and metastases and to study
91 whether resistance to immune checkpoint inhibition develops from pre-existing or *de*
92 *novo* generated cell populations.

93 **Results**

94 **Immunoediting of breast cancer cells in the primary tumour**

95 To understand the role of the immune system and immunotherapy in shaping the clonal
96 dynamics of cancer cells within primary tumours, we used the DNA barcoding
97 approach (Fig 1). We introduced the ClonTracer DNA barcode library that contains
98 approximately 7 million unique barcodes *(30)* into the immunotherapy-sensitive mouse

99 mammary carcinoma EMT6 cells *(31)* (Sup Fig 1), resulting in ~41 000 unique
100 barcodes identified by DNA sequencing (Sup Fig 2A) . We then inoculated 250 000
101 cells (~6 fold over representation of each barcode) into the mammary fat pad of
102 syngeneic immune-competent wild-type (WT) Balb/c mice or severely
103 immunocompromised NOD SCID Gamma (NSG) mice that lack T cells, B cells and
104 functional NK cells, and compared the number of clones that were able to engraft and
105 grow (Fig 2A, Sup Fig 2).

106 Tumour growth was faster in NSG mice than in WT mice, with tumours reaching ethical
107 endpoint in NSG mice on day 14 post-transplant and in WT mice by day 23 (Fig 2B),
108 leading to NSG mice having significantly shorter median overall survival (14 days) than
109 WT mice (22 days, Mantel-Cox p=0.009). These results suggest that the immune
110 system plays an important role in controlling the growth of the EMT6 primary tumours
111 (Fig 2C).

112 To examine the influence of immunotherapy on tumour growth and clonal dynamics,
113 we compared WT mice treated with immunotherapy (anti-PD1 + anti-CTLA4) or
114 control antibodies starting from day 10 when tumours were approximately 200mm^3 (Fig
115 2A). Tumours in all mice treated with control antibodies reached ethical endpoint by
116 day 23 (Fig 2D). In contrast, tumours in all mice treated with immunotherapy regressed
117 following treatment, with 50% relapsing and reaching ethical endpoint between days
118 46 and 54 (Fig 2D). The remaining mice treated with immunotherapy remained tumour
119 free until the experiment was terminated on day 60. During harvest a small residual
120 lesion was observed in two of these mice, no metastatic lesions were observed in any
121 of the mice irrespective of treatment. Kaplan-Meier analysis demonstrated that
122 immunotherapy significantly increased the median survival from 22 days to 57 days
123 (Mantel-Cox p=0.0006) (Fig 2E).

124 To determine if immune control of tumour growth was driven at a clonal level, we
125 examined the number and distribution of barcodes present in primary tumours collected
126 at ethical endpoint in the experiment described above. We found that at ethical
127 endpoint, tumours grown in NSG mice had over 50 times the number of unique
128 barcodes as tumours grown in control WT mice ($p<0.0001$, generalised linear model
129 (GLM) with Tukey's correction), which in turn had more than 20 times the number of
130 unique barcodes found in WT mice treated with immunotherapy ($p<0.0001$, GLM with
131 Tukey's correction) (Fig 2F, Sup Fig 2A). We applied Shannon diversity analysis to
132 understand how the immune system influenced the diversity of barcodes in these
133 samples. Shannon diversity index determines how evenly distributed the barcodes are
134 within a population and is only moderately influenced by barcode number. Analysis of
135 barcode diversity revealed a trend to a lower barcode diversity in tumours from control
136 (WT Balb/c) mice than that in tumours from NSG mice whereas the barcode diversity
137 was significantly lower in immunotherapy treated than in control treated Balb/c mice
138 (Fig 2G, $p<0.001$ one way ANOVA with Tukey's correction). These data suggests that
139 a subset of EMT6 cells are more resistant to the endogenous immune system but this
140 selection does not skew the evenness of the barcode distribution significantly, which
141 suggests that all clones resistant to the immune system have similar levels of resistance.
142 In contrast, immunotherapy applies a more stringent bottleneck that only a limited
143 number of clones can overcome, and with a high degree of variability in doing so.
144 Further analysis identified EMT6 clones that were reproducibly enriched across
145 multiple mice following immunotherapy treatment, indicating that they had a pre-
146 existing resistance phenotype that was being positively selected for (Fig 2H).

147 **Immunoediting of breast cancer cells during metastasis**

148 To determine whether immunoediting continued during metastatic dissemination and
149 whether specific metastatic clones were enriched or depleted, we turned to the highly
150 metastatic 4T1 mammary carcinoma model, as the EMT6 cell line is poorly metastatic
151 *(32)*. We introduced the barcode library into 4T1 cells, leading to a cell pool with ~5000
152 unique barcodes (Sup Fig 3, Sup Fig 4), then inoculated 50,000 of these cells (a ~ 10-
153 fold over- representation of each barcode) into the mammary fat pad of WT and
154 NSGmice (Sup Fig 4). We resected primary tumours 15 days following inoculation to
155 allow metastases to develop (Fig 3A). All mice developed lethal lung metastases, with
156 NSG mice succumbing to metastatic disease earlier than WT mice (median survival of
157 25.5 days verses 35 days, p = 0.0002, Mantel-Cox Log-rank test) (Fig 3B). Primary
158 tumour sizes at the time of resection were similar between the groups (Sup Fig 5A).
159 Adjuvant immunotherapy with anti-PD1 + anti-CTLA4 led to a modest but significant
160 increase in survival (37.5 days) versus control-treated mice (33 days; p=0.0121,
161 Mantel-Cox Log-rank test) (Fig 3C).

162 We then examined whether the endogenous immune system shaped metastatic clonal
163 dynamics. While primary tumours contained similar numbers of clones and barcode
164 diversity in NSG and WT hosts (Fig. 3D and Sup Fig 5B, C), the lung metastases of
165 NSG mice contained ~3 times as many barcode clones as those of WT controls (Fig
166 3D). We next determined if the increase in survival following immunotherapy was
167 associated with alterations in clonal dynamics. As the treatment was only given after
168 excision of the primary tumour, the immunotherapy would only affect the outgrowth of
169 cancer cells that had already metastasised to the lung. Despite only a modest increase
170 in survival following combination immunotherapy (Fig 3C), we observed a 70%
171 reduction in the number of clones in metastases (Fig 3E).

172 The higher barcode number in the lung metastases of NSG mice than in the lung of WT
173 mice was associated with a higher diversity of barcode as measured using the Shannon
174 diversity index (Fig 3F). This shows that the endogenous immune system restricts the
175 number and skews the diversity of metastatic clones that can reach and outgrow in the
176 lungs. In addition to the reduction in barcode number following immunotherapy
177 treatment, we also saw a significant reduction in barcode diversity (Fig 3F). This
178 suggests that immunotherapy is leading to the immunoediting of specific clonal cell
179 populations over others.

180 To further understand the key immune cell types that control clonal outgrowth in the
181 lung, we depleted CD4 T cells using anti-CD4, CD8 T cells using anti-CD8 or NK cells
182 using anti-asialo-GM1 in wildtype mice starting one day prior to tumour resection (Sup
183 Figs 6B, 6C, 6D). None of these treatments led to a significant change in overall
184 survival (Sup Fig 6A). Initial experiments indicated a small change in the number of
185 clones detected within the lung between treatment groups (Sup Fig 6E). However, this
186 was not recapitulated in repeat experiments (Sup Fig 6F). This may suggest that none
187 of these cell types alone are sufficient to restrict clonal diversity after seeding of
188 pulmonary metastases.

189 We were surprised with the lack of immunoediting in the primary tumours of the 4T1
190 model so we repeated these experiments with a second pool of barcoded 4T1 cells
191 containing a larger barcode library (300 000 barcodes). Following the injection of 50
192 000 barcoded cells, we recovered approximately 10 000 - 12 000 barcode sequences
193 from each primary tumour grown in WT mice and in NSG mice (Sup Fig 8A). This
194 suggests that roughly a fifth of the injected cells are able to engraft and grow in the
195 mammary gland. As clone diversity was again similar between NSG mice and WT
196 mice, this confirms that the immune system does not play a major role in restricting

197 growth of 4T1 cells in the primary tumour setting. In contrast, when we examined the
198 number of clones that had spread to the lung we again found approximately 3 times as
199 many in NSG mice as in wild-type mice (Sup Fig 8A). In addition, we similarly saw a
200 ~3-fold reduction in barcode diversity in response to immunotherapy (Sup Fig 8B).
201 These results confirm that the 4T1 cell line is already highly immunoedited for growing
202 at the primary site yet is subjected to a second round of immunoediting during
203 metastasis.

204 **Patterns of enrichment and depletion of specific clones**

205 To better understand how specific clonal cell populations responded to the immune
206 system and immunotherapy, we combined the barcode frequencies from the two
207 datasets utilising the 5000 barcodes library (WT vs NSG and Control vs
208 Immunotherapy). We performed unsupervised hierarchical clustering of these samples
209 and selected barcodes that were observed at greater than 5% frequency in any one
210 sample. We found that the primary tumours from the two experiments cluster together
211 irrespective of the immune status of the mouse (Balb/c or NSG), further suggesting that
212 4T1 primary tumours, in contrast to EMT6 tumours, do not undergo further
213 immunoediting (Fig 4A). In contrast, lung tumours formed in the NSG hosts did not
214 cluster with lung tumours formed in Balb/c mice, with the immunotherapy treated
215 samples mostly clustering alone or with metastases formed in WT mice. A number of
216 specific barcodes were enriched in lung metastases of all NSG mice, indicating that
217 these clones were highly metastatic (Fig 4A), this is similar to the findings of
218 Wagenblast and colleagues who examined 4T1 clonal diversity in metastases in NSG
219 (but not wild type mice) and found tissue-specific enrichment of unique barcode
220 clones*(33)*.

221 We identified from the heatmap (Fig 4A) a number of barcodes that had striking
222 patterns of enrichment or depletion in response to the immune system and
223 immunotherapy, we replotted these using a dot plot (Fig 4B). These barcodes were
224 enriched or depleted in a reproducible manner across replicate mice, suggesting this is
225 due to inherent features of these clones. There are three barcodes that were enriched in
226 NSG lung metastases (NSG1-3), present at lower abundance in untreated wild-type
227 mice, and were completely eliminated following immunotherapy treatment. This
228 suggests that these clones are highly metastatic in the absence of an immune system but
229 they are immunogenic and are thus subjected to immunoediting in WT mice,
230 particularly following immunotherapy. Another group of metastatic clones present in
231 the lung of NSG and WT mice were further enriched following immunotherapy (IE1-
232 2). These immunotherapy enriched clones were detected in the lungs of all six replicate
233 mice. With the significant reduction in the number of barcodes present following
234 immunotherapy, the odds ratio of this reproducible enrichment happening by chance is
235 0.0034 (95% confidence interval: 0.0010-0.0079; chi square p value: 3.67×10^{-251}). This
236 suggests that these clones have a pre-existing resistance phenotype and are positively
237 selected for following immunotherapy.

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

244 **Analysis of immunotherapy resistant clones**

245 To understand more about the phenotype of these immunotherapy resistant clones, we
246 established clonal cell populations from two of them (designated IE1 and IE2) and from
247 two independent control clones (NT1 and NT2) that were not enriched following
248 immunotherapy. We generated 3-4 independent clonal cell lines per barcode clone.
249 These clonal cell lines were isolated from the parental barcoded 4T1 cell population *in*
250 *vitro* with no additional selective manipulation. The barcode within each of these clonal
251 cell lines was confirmed to be correct using Sanger sequencing. All four clonal cell
252 lines had similar growth kinetics *in vitro* indicating no proliferative advantage of the
253 immune evasive clones *in vitro* (Sup Fig 9A). All clonal cell lines were able to form
254 tumours in the primary setting (Sup Fig 9B). IE2 demonstrated considerable metastatic
255 potential with 66% of mice forming extensive lung metastases and regularly
256 metastasised to the lungs (Sup Fig 9C). IE1 also demonstrated some metastatic capacity
257 with 30% of mice forming lung metastases. Neither NT1 nor NT2 successfully
258 metastasised to the lungs (Sup Fig 9C).

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

260 To identify the barcode integration sites and determine whether the clones contained
261 large scale genomic alterations, we performed whole genome sequencing (WGS) at
262 around 30x coverage of the clones. The WGS analysis determined the precise genomic
263 location where barcodes integrated (Sup Table 2). The integration site in IE1 was in the
264 intergenic region between *Kpna*2 and *Smurf*2 and the integration site in IE2 was within
265 an intron of *Nrf*1, neither integration site changed the coding sequence of these genes.
266 Copy number analysis determined that no clone had dramatic copy number changes
267 when compared to other clones. Each clone only contained a small number of single
268 copy number gains and losses (IE1 only 6 CNVs and IE2 only 5 CNVs), with clone
269 NT2 showing the greatest number of CNVs at 41 (summarised in Sup Table 3). We

270 found one locus on chromosome 18 with a single copy number gain in both IE1 and
271 IE2 that led to 3 copies of *Nc3r1*, encoding the Glucocorticoid receptor, and *Arhgap26*,
272 encoding a Rho GTPase that associates with focal adhesion kinase. However, this copy
273 number gain on chromosome 18 was also present in the NT2 clone. These results
274 demonstrate that large scale genomic changes unlikely play a major role in determining
275 various phenotypes of different clones but instead suggest that copy number changes
276 may be selected against during immunoediting.

277 **Transcriptomic analysis of the clonal cell lines**

278 To investigate the mechanism of immune evasion by these clones, we performed
279 RNAseq analysis and compared the two immunotherapy resistant clones to the bulk
280 4T1 population. Differential gene expression analysis was carried out using EdgeR. The
281 IE1 clone had 1553 differentially expressed genes (Log fold change >2 and FDR
282 $p<0.05$) with 478 significantly upregulated and 1075 significantly downregulated (Fig
283 5A). The IE2 clone had 1099 differentially expressed genes with 375 significantly
284 upregulated and 724 significantly downregulated (Fig 5B). The non-target clones had
285 fewer gene expression changes compared to the bulk 4T1 population with NT1 having
286 621 and NT2 having only 262. We examined the top differentially expressed genes
287 between each of IE1 and IE2 with the parental 4T1 cells, however, we did not find any
288 with an obvious role in immune evasion (Sup Tables 4 & 5). Gene set enrichment
289 analysis revealed that among the top ten gene sets upregulated in IE1 and in IE2, only
290 two (CHEN_HOXA5_TARGETS_9HR_UP,
291 BLUM_RESPONSE_TO_SALIRASIB_UP) were common between them (Sup Tables
292 6 & 7). *Hoxa5* is a known tumour suppressor gene in breast cancer *(34)*, although we
293 see an enrichment of its target genes, the expression of *Hoxa5* itself was significantly
294 reduced in the IE1 clone and trended to be reduced in the IE2 clone. There was no

295 overlap in the top ten downregulated gene-sets between IE1 and IE2. The top
296 downregulated gene-set for IE1 was the
297 REACTOME_UB_SPECIFIC_PROCESSING_PROTEASES gene-set, which
298 contained two genes involved in antigen processing for display by MHC-I (*Psmb8* and
299 *Psmb9*). As down regulation of the MHC-I pathway is a common mechanism of
300 immune evasion, we investigated this in more detail.

301 Through this analysis we found that the IE1 clone had significantly reduced expression
302 of many genes related to antigen presentation, including MHC-I (*H2-k1*), *Tap2*, *Psmb8*,
303 *Psmb9* and *Psmb10* (Fig 5C). *H2-k1* encodes the main MHC molecule that is expressed
304 by the Balb/c mouse from which the 4T1 carcinoma cell line was derived from. We
305 validated the reduction in MHC-I expression levels seen in the RNAseq data at the
306 protein level using flow cytometry (Fig 5D). This analysis showed that the IE1 clone
307 had significantly reduced cell surface MHC-I protein compared to the bulk 4T1
308 population. We thus examined the WGS data and found that the loss of MHC-I
309 expression in IE1 was not due to genomic loss at the MHC locus on chromosome 17
310 (Sup Fig 10A). In contrast, the IE2 clone had elevated levels of a number of these MHC
311 related genes (Fig 5C) in addition to *H2-t23* that encodes a non-classical MHC molecule
312 (Fig 5E) known to negatively regulate NK cells through their inhibitory receptor
313 NKG2A (*35*). Interestingly, IE2 cells also demonstrated a significantly increased
314 expression of *Cd274* that encodes the T cell inhibitory molecule PD-L1 (Fig 5E). This
315 again was validated at the protein level using flow cytometry (Fig 5F). These results
316 demonstrate that these two immunotherapy resistant clones are phenotypically unique.
317 To determine if these two mechanisms are simultaneously maintained *in vivo*, we
318 carried out flow cytometry for MHC-I and PD-L1 on advanced 4T1 lung metastases
319 from immunotherapy treated BALB/c mice. By isolating RFP+ cancer epithelial cells

320 from lung tissue, we observed MHC-I-high and MHC-I-low neoplastic populations,
321 both with varying expression of PD-L1 (Sup Fig 11). This provides convincing
322 evidence of these two mechanisms of immune evasion being maintained
323 simultaneously *in vivo*.

324 We next examined whether copy number changes or barcode integration sites identified
325 above impacted gene expression. The copy number changes of *Nc3r1* and *Arhgap26*
326 were associated with significantly increase of their expression in IE1 and IE2 cells but
327 not in NT2 cells (Sup Fig 10B). Elevated *NC3R1* expression has been associated with
328 poor prognosis and metastasis in triple-negative breast cancer, although whether it plays
329 a role in immune evasion is not known (*29, 30*). As stated above, the barcode integrated
330 in the intergenic region between *Kpna2* and *Smurf2* in IE1 and within an intron of *Nrf1*
331 in IE2. Among these three genes, only the expression of *Smurf2* was significantly
332 altered with a modest log fold increase of 0.59 in IE1.

333 **Demethylating drugs do not fully restore MHC expression**

334 Demethylating agents such as 5-aza-2'-deoxycytidine (5-aza) are known to upregulate
335 MHC-I expression in cancer cells (*36*), thus we treated our clonal cell lines with 5-aza
336 for 72 hours to determine whether DNA methylation was a mechanisms suppressing
337 MHC expression in the IE1 clone. Using flow cytometry, we observed that MHC-I
338 expression was elevated in a dose dependent manner following 5-aza treatment in all
339 clones. However, MHC-I expression in the IE1 clone was consistently lower than the
340 parental 4T1 cell line at all doses of 5-aza (Sup Fig 12A). This indicates that gene hyper-
341 methylation is not the mechanism of MHC-I suppression in the IE1 clone.

342 Interferon (IFN)-gamma(g) stimulation is another mechanism by which MHC
343 expression can be increased on cancer cells. The IE1 clone responded to IFN-gamma

344 treatment by upregulating MHC-I expression but again it remained suppressed
345 compared to the parental 4T1 cells (Sup Fig 12B). This suggests these cells broadly
346 retain the transcriptional regulatory machinery that is required to upregulate MHC-I in
347 response to IFN-gamma stimulation. These results indicate that MHC-I downregulation
348 is likely regulated by epigenetic factors other than DNA methylation and that the
349 majority of MHC-I expression in this clone can be restored by IFN-g treatment.

350 **The 4T1-IE2 clone can directly suppress anti-cancer CD8 T-cell responses**

351 To further characterise the immunotherapy resistant clones (IE1 and IE2) in comparison
352 to the control (NT1 and NT2) clones we performed an *in vitro* CD8 T cell activation
353 assay (*37, 38*). This assay utilises a pool of *de novo* generated 4T1 specific CD8+ T
354 cells, with intracellular IFN-□ production utilised as a functional measure of T cell
355 activation. We then determined how T cell activation changed following co-culture with
356 the clonal 4T1 populations. In the absence of 4T1 cells 1% of CD8 T cells had
357 detectable intracellular IFN-□ production (Fig 6A) while in the presence of the parental
358 4T1 cells 23% of the T cells had detectable intracellular IFN-g production. When the
359 activation of T cells by cloal cell lines was examined, only the IE2 clone significantly
360 reduced CD8 T cell activation (Fig 6A ; $p = 0.0229$). As we had demonstrated that the
361 immune evasive clones were also more immunotherapy resistant, we examined the
362 effect of adding anti-PD1 immunotherapy to the T-cells *in vitro*. The addition of anti-
363 PD1 significantly increased anti-cancer T-cell responses to the control clones and the
364 IE1 clones, while T cell response to the IE2 clone was unchanged with the addition of
365 anti-PD1 ($p=0.36$; Fig 6B).
366 This assay demonstrated that the IE2 clone uniquely evades activation of anti-cancer
367 CD8 T-cells, and that the addition of anti-PD1 could not overcome the T cell
368 suppression in response to this 4T1 clone. In contrast the IE1 clone induced a similar

369 response to the control clones suggesting that it does not directly suppress CD8 T-cell
370 activation.
371

372 **Overlapping gene signature is associated with poor survival in breast cancer**
373 **patients**

374 We had noted that the GSEA analysis showed some overlap in enriched gene-sets
375 between IE1 and IE2. We thus reasoned that, in addition to having their unique immune
376 evasion features, these clones may have some pathways in common. To identify
377 immune evasion pathways common to both immunotherapy resistant clones, we
378 identified overlapping gene expression changes (Sup Table 8). This analysis
379 demonstrated that immunotherapy resistant clones had more genes with expression
380 changes in common than either one had with either non-target clones (Fig 7A). We
381 generated a heatmap of the top 50 upregulated and downregulated genes across all
382 samples (Fig 7B) and performed GSEA analysis using C2 on the longer list (Sup Fig
383 13A, Sup Table 9). Only two gene sets had significant p values when multiple testing
384 was considered, these were the HOXA5 gene set mentioned previously and a COVID-
385 19 related gene set. Although not significant, there were several additional COVID-19
386 related gene sets from the same recent publication identified in the overlapping
387 upregulated gene list suggesting an immune related role of these genes (*39*). We also
388 identified three gene sets that were closely related and had clear associations with
389 epigenetic regulation, "BENPORATH_ES_WITH_H2K27ME3",
390 "BENPORATH_PRC2_TARGETS" and "BENPORATH_SUZ_12_TARGETS". In
391 patient data, Ben-Porath and colleagues found negative enrichment of these genesets to
392 be correlated with a stem-like phenotype and to be associated with poor prognosis (*40*).
393 Further investigation showed the top genes driving these signatures were also

394 negatively enriched in our derived gene signature, namely HHIP, CWH43, WNT10B,
395 ABC3, CHN2 and CRIP1. This suggests a possible role for PRC2-mediated MHC-I
396 suppression in our subclones.

397 To understand the role of these genes in patients, we generated signatures from the top
398 25 upregulated and downregulated genes that had human orthologs and were detectable
399 in both the METABRIC (*41*) and TCGA (*42*)datasets. We then analysed the association
400 of these signatures with the survival of patients with basal-like breast cancer from these
401 cohorts. We limited our analysis to basal-like breast cancer as we considered this the
402 most relevant patient subgroup as the gene signature was derived from the 4T1 mouse
403 model of basal-like breast. Although patients in these cohorts were not treated with
404 immunotherapy, it has previously been demonstrated that immune features such as the
405 number of tumour infiltrating lymphocytes or regulatory T cells influence prognosis in
406 basal-like breast cancer patients (*43*). When we analysed overall survival, we observed
407 that the upregulated gene signature associated with significant poorer outcome in both
408 cohorts (METABRIC: p=0.0043, HR=2.0, Fig 7C; TCGA: p=0.042, HR=4.3, Fig 7D).
409 In contrast, the downregulated gene signature did not show any significance in either
410 cohort (data not shown). We generated heatmaps with unsupervised clustering to
411 determine whether specific individual genes or groups of genes from the signature were
412 driving the association with survival (Sup Fig 13B). In the TCGA dataset, we observed
413 a number of clusters that seemed to associate more with survival, one of these included
414 *FAM71F2, MASP2, HLF, PPP1R15A, MMP23B* and *LIMS2*. Interestingly, GADD34,
415 encoded by *PPP1R15A*, had previously been demonstrated to be critical in blocking
416 immunogenic cell death following chemotherapy, when it was inhibited, the
417 chemotherapy response was improved in immunocompetent mice but not in
418 immunocompromised mice (*44*). A second group of genes included *SEZ6L2*, which had

419 been associated with survival in a number of cancers but not through an immune related
420 mechanism (*45, 46*) There was no enrichment of proliferation or invasion gene sets in
421 our GSEA analysis, suggesting that these processes were not behind the poor outcome
422 of patients whose breast cancer highly expressed genes in the common upregulated
423 signature.

424 Previous studies have shown that cytotoxic T lymphocytes (CTL) infiltration correlates
425 with survival in basal-like breast cancer, so it is possible that our signature was a
426 surrogate measure of T cell infiltration. To test this, we performed TIDE analysis (*47*)
427 on TCGA and METABRIC cohorts followed by correlation analysis between the CTL
428 signature score and our upregulated immune evasion signature score. This showed no
429 correlation in the METABRIC cohort (Sup Fig 13C) and only a weak negative
430 correlation in the TCGA cohort (Sup Fig 13D), suggesting little overlap between these
431 two predictors of patient survival. Thus, it suggests we have discovered a novel immune
432 related signature associated with patient survival in basal-like breast cancer. Future
433 studies will be necessary to determine how the genes in this signature regulate survival,
434 influence immune evasion and immunotherapy response.

435 **Discussion**

436 Immunotherapy has revolutionised cancer therapy, with long-term responses seen in
437 certain patient groups with a few types of cancer. Unfortunately, other patients having
438 these same types of cancer and most patients having other types of malignancies,
439 including breast cancer, have limited to no response to the current immunotherapies.
440 Thus, our understanding of immune evasion in non-responsive cancer types needs
441 significant improvement. We have addressed this by examining immune evasion at a
442 clonal level and used this information to identify pathways that could be targeted to
443 overcome immunotherapy resistance.

444 Here we show, for the first time in a solid tumour, that clonal immunoediting occurs
445 and is enhanced by immunotherapy. Using the more immunotherapy sensitive EMT6
446 model we demonstrate that immunoediting occurs during the development of primary
447 tumours and that immunotherapy leads to strong clonal selection. Most of the more
448 immunotherapy resistant 4T1 cells are able to evade the immune system during the
449 development of primary tumours but only a subset of them are able to evade the immune
450 system during metastasis. This indicates that immune evasion is not a static process but
451 requires ongoing regulation through tumour progression, even in a highly aggressive
452 allograft model. These findings broadly agree with findings from a recent
453 comprehensive genomic analysis of patient samples assessed across metastatic sites and
454 over time (*3, 48*). These studies tracked clonal populations in metastatic lesions using
455 whole genome sequencing, examined a number of immune correlates and identified
456 immunoediting that was associated with the immune response. However, they were
457 unable to examine clonal heterogeneity driven by epigenetic or transcriptomic changes
458 and were limited in identification of rare clones by sequencing depth. Future clinical
459 studies utilising single cell approaches to analyse multiple biopsies from individual
460 patients over a time-course of treatment will be necessary to confirm the key findings
461 of this study in patients. While extremely challenging, these studies are becoming more
462 feasible with recent technological improvements.

463 Intriguingly the 4T1 model, unlike the EMT6 model, showed little immunoediting in
464 the primary tumour. This suggests that either the vast majority of 4T1 cells are
465 inherently resistant to immune control at the orthotopic site, or that 4T1 cells very
466 rapidly set up a suppressive immune microenvironment that protects the majority of
467 clones from immune mediated killing. The ability of 4T1 cells to induce myeloid
468 derived suppressor cells could well contribute to a suppressive immune

469 microenvironment, however, further studies would be necessary to further clarify this
470 (*49*).

471 As we wanted to understand the role of immunotherapy in controlling metastatic disease
472 we examined lung metastases from the 4T1 model following resection of the primary
473 tumour. Lung metastasis occurs early in the 4T1 model with our studies indicating
474 metastases can form following resection as early as day 13 (data not shown), and others
475 showing the related 4T1.2 model robustly metastasizes by day 10 (*50*). This indicates
476 that when adjuvant immunotherapy was given, these therapies were activating immune
477 cells to target micrometastases that had already formed within the lungs. We postulate
478 that while possible it is unlikely that circulating cancer cells were a major target of
479 adjuvant immunotherapy as previous studies have indicated that circulating breast
480 cancer cells only have a short half-life of 1-2.4 hours in circulation (*51*).

481 A number of previous barcoding studies in breast cancer had focused on metastasis and
482 response to chemotherapy. However, these were performed in immunocompromised
483 mice so that the role of the immune system in these processes was not addressed (*33,*
484 *52-54*). In the absence of a fully intact immune system it was demonstrated that specific
485 clones have greater metastatic ability (*33, 52, 54*). Our results suggest that in the context
486 of an intact immune system, a subset of these highly metastatic clones identified in
487 these studies may have been recognised and removed via immunoediting. Similar to
488 our results, these studies found that the dominant clone within the primary tumour
489 generally was not the dominant clone in the metastases (*33, 52, 54*). Some of these
490 studies also showed that chemotherapy treatment of PDX models led to a decrease in
491 clonal abundance and diversity in relapsed disease, which is similar to what we found
492 with immunotherapy (*53, 54*). Future studies combining immunotherapy with
493 chemotherapy utilising a similar regimen to the atezolizumab plus nab-paclictaxel of

494 the Impassion130 trial would give important insights into how combining these two
495 treatment modalities affected clonal diversity following relapse (*55*).

496 Our and others' studies have indicated that immunoediting can select for clones with
497 immune evasive phenotypes irrespective of specific neoantigens (*2, 6, 56*). One
498 previous study examining immunoediting at a clonal level used a fluorescent barcoding
499 approach in a B cell leukemia model (*13*). While our findings broadly agree with the
500 findings of that study, the study by Milo and colleagues was limited to five unique
501 fluorescent clones that could be tracked and was confounded by the variable
502 immunogenicity of these fluorescent proteins. DNA barcoding , in contrast, allowed for
503 the labelling of thousands of clones and a much more precise identification of immune
504 evasive clones. We could then isolate these clones and identify both common and
505 variable features of immune evasion in them. Furthermore, this technique, unlike a
506 fluorescent barcode approach, allowed us to demonstrate that clones that had greater
507 immunotherapy resistance pre-existed in both EMT6 and 4T1 models, as these clones
508 were enriched from the same starting pool of cells across replicate mice.

509 We identified and isolated two immunotherapy resistant clones from the 4T1 model. In
510 depth analysis of these resistant clones demonstrated that they expressed genes of key
511 immune evasive pathways (MHC-I and PD-L1) differently. Intratumoural
512 heterogeneity (ITH) has previously been associated with resistance to immunotherapy
513 in melanoma and lung cancer, with higher ITH being associated with resistance to
514 immunotherapy (*26, 27, 57*). McGranahan and colleagues postulated that this was due
515 to improved T cell killing of tumours with clonal neo-antigens. A non-mutually
516 exclusive explanation is that clonal tumours are less likely to contain cancer cells with
517 a pre-existing resistance mechanism to immunotherapy. These findings refine the
518 concept of cancer immunoediting, demonstrating that there are clonal populations of

519 cancer cells with variable resistance to the immune system. Based on their phenotype,
520 these clones are either enriched or depleted by an active immune system and
521 immunotherapy.

522 We identified a core overlapping gene expression profile between the two
523 immunotherapy resistant clones. The common upregulated gene signature was able to
524 stratify basal-like breast cancer patients into good and poor prognosis. This gene
525 signature appeared to represent a novel immune evasion pathway associated with poor
526 prognosis. Aside from *Ppp1r15a*, this signature did not contain genes known to be
527 associated with immune evasion and it did not contain genes associated with other poor
528 prognostic signatures, such as proliferation or invasion.

529 Because both T cells and NK cells are present during immunoediting in our models,
530 this common signature likely enables cancer cells to evade both T cells and NK cells.
531 Interestingly, our common signature does not strongly correlate with CTL infiltration,
532 which indicates that this signature is not a surrogate for the lack of T cell infiltration
533 and suggests that these genes likely do not regulate immune evasion by influencing
534 immune cell recruitment. These common genes may offer new insights into developing
535 therapeutic approaches to improve immunotherapy response in breast cancer. One of
536 the common immunotherapy resistance genes we identified was *PPP1R15A*, which is
537 consistent with its known role in immunogenic cell death in the response to
538 chemotherapy (*44*). Recent studies in a mouse model of multiple sclerosis demonstrated
539 that this pathway can be targeted using Sephin-1, a small molecule (*58*). Future studies
540 are necessary to examine whether this compound or others targeting this pathway could
541 synergise with immunotherapy or immunogenic chemotherapy to treat breast or other
542 types of cancer. As breast cancers have relatively low mutational burden, it is likely
543 that epigenetic factors may play a greater role than mutational events in driving ITH in

544 breast cancer. One current approach to improve immunotherapy response under
545 investigation is combining immunotherapy with epigenetic targeting drugs such as
546 decitabine and HDAC inhibitors (*36, 59*). This combination has been shown to increase
547 MHC protein expression and improve response to immunotherapy. However,
548 epigenetic drugs may reduce the diversity of clones and overcome other epigenetically
549 driven immune evasion mechanisms in addition to enhancing MHC expression (*60*).
550 Further research is needed to test this hypothesis more fully. Our results, however,
551 suggest that while demethylating agents could increase MHC-I expression in the MHC-
552 I low immunotherapy resistant clone, they did not increase it above the baseline seen in
553 the parental 4T1 cell line. This is corroborated by a recent study demonstrating that
554 while treating breast cancer patients with demethylating agents could increase MHC-I
555 expression in most patients' tumors, a subset appeared resistant to this therapy (*36*).
556 This suggests that while epigenetic treatments may improve the proportion of patients
557 responding to immunotherapy, in some cases pre-existing clones could still mediate
558 resistance to this combination.

559 A limitation of this study is the reliance on mouse cell line models, which do not
560 recapitulate early stages of tumorigenesis and do not represent the full diversity of
561 human breast cancer. However, syngeneic allograft models have delivered central
562 insights about the immune response to cancer and demonstrated the utility of
563 immunotherapies (*61*). Another limitation is that the integration of the barcode and
564 selection markers into the genome and the potential immunogenicity of red fluorescent
565 protein (RFP) could affect the phenotype of these cancer cells. We and others have
566 found in previous studies that some fluorophores and luciferase were immunogenic and
567 negatively affect tumour growth and metastasis in the 4T1 model (*29, 62*). However,
568 we found that tumour growth and metastasis were unaffected by RFP expression in this

569 study. While the introduction of DNA barcodes could have influenced the phenotype
570 of specific clones, we feel that this is unlikely given that no dramatic impact on the
571 expression of the genes closest to the integration site. Furthermore, none of the genes
572 associated with integration sites was identified to be significantly involved in cancer
573 cell evasion of CD8 T cell responses in a recent CRISPR screen .

574 To survive in any given system, cancer cells must utilise a number of mechanisms to
575 avoid more than just immune destruction, as extensively reviewed in the recent work
576 of Hanahan (*64*). Not only are IE1 and IE2 immune evasive, they are also highly
577 metastatic and by their very nature must be able to grow independent of anchorage, as
578 well as possessing abilities to engraft in both the mammary fat pad and lungs, shed from
579 the primary tumour prior to resection and resist all other mechanisms of host anti-cancer
580 response. We attribute the decrease in unique clones present in the lungs of control
581 treated 4T1 tumour bearing mice compared to the primary tumour to precisely this; only
582 some clones within the engrafted tumour possess the required abilities to be able to
583 successfully metastasise. By making comparisons between controlled conditions, e.g.
584 immunotherapy treated lungs vs control treated lungs, we believe we have effectively
585 demonstrated immunoediting occurring *in vivo*.

586

587 Overall, this study has demonstrated that immunoediting occurs at the clonal level in
588 primary tumours and that a second round of immunoediting occurs during metastasis.
589 Immunotherapies dramatically enhanced immunoediting, but pre-existing resistant
590 populations were still responsible for relapse. The large reduction in clonal diversity
591 following immunotherapy in the 4T1 model, which is known to be poorly responsive
592 to immunotherapy, suggests that slight improvements through combination therapy
593 could eliminate the remaining clones and lead to dramatic improvements in survival.

By isolating immunotherapy resistant clones and characterizing them, we identified common and distinct immune evasion pathways. We anticipate that through targeting pathways identified in this study, in particular 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 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. The authors would like to acknowledge Dr Tim Peters and thank him for his valuable statistical advice. We acknowledge the contribution of Isabelle Shapiro and Cheryl Grant as consumer advocates. This manuscript was edited at Life Science Editors.

Author Contributions

L.A.B. performed experiments, analysed and interpreted the data, and drafted the manuscript. N.B. performed bioinformatic analysis of the barcoding and RNAseq data. J.Y. performed experiments. S.Z.W. performed bioinformatic analysis of patient cohorts. N.D. performed bioinformatic analysis of the DNAseq data. D.L.R. performed bioinformatic analysis of the barcoding data. C-L.C. performed experiments. G.A. analysed and interpreted the data. D.J.Z and B.S.P performed experiments and analysed data. A.S. helped conceive the study, analysed and interpreted the data, and helped draft

619 the manuscript. S.J. conceived the study, performed experiments, analysed and
620 interpreted the data, and drafted the manuscript.

621 **Methods**

622 Cells

623 4T1 cells were obtained from ATCC. 4T1 cells were grown in RPMI (Gibco)
624 supplemented with 10% FCS, D-Glucose, Sodium Pyruvate, 2mM HEPES, and
625 Penicillin/Streptomycin. EMT6 cells were obtained from ATCC. EMT6 cells were
626 grown in Waymouth's MB 752/1 Medium supplemented with 15% FCS and 2mM L-
627 glutamine. Cells were tested for mycoplasma contamination using the MycoAlert
628 Mycoplasma Detection Kit, 100 Test Kit (Catalog# LT07-318).

629

630 Cellular DNA barcoding

631 The ClonTracer library was a gift from Dr Frank Stegmeier (Addgene #67267).
632 Lentiviral particles containing the high-complexity barcode library were produced by
633 transfecting 293T cells. 4T1 and EMT6 cancer cell lines were barcoded by lentiviral
634 infection using 0.8 µg/ml polybrene. Cells from each line were infected with a target
635 MOI of 0.1, corresponding to 10% infectivity to ensure single lentiviral integration.
636 Cells that received a barcode were then sorted based on the RFP reporter protein using
637 a BD FACSAriaII, these cells were then expanded and frozen into a number of aliquots
638 for the subsequent experiments. 4T1 cells were generated with two different barcode
639 complexities, one with ~5000 barcodes (4T1 BC5000) and one with ~300 000 barcodes.
640 The EMT6 cells and the high complexity 4T1 cells were passaged twice following cell
641 sorting, frozen and cells from these aliquots were thawed and passaged once more prior
642 to transplantation. The low complexity 4T1 cells (4T1 BC5000) were a subpool derived

643 from the higher complexity 4T1 cell line, these were passaged an additional 3 times to
644 expand and freeze and were then thawed and used as described above.
645
646 *In vivo* experiments
647 All animal experiments were approved by the Garvan Institute of Medical Research/St.
648 Vincent's Hospital Animal Experimentation Ethics Committee.
649 Immunocompetent BALB/c mice and immunocompromised NOD.Cg-*Prkdc*^{scid}
650 *Il2rg*^{tm1Wjl}/SzJ (NSG) mice aged 6-to-8 weeks were obtained from Australian
651 BioResources (Moss Vale, Australia) and housed at the Garvan Institute of Medical
652 Research.
653
654 *In vivo* tumour growth
655 For tumour transplantation, barcoded EMT6 cells (ATCC, USA) were resuspended in
656 Matrigel 2.5x10^5 cells in 100 ml volume were injected into the 4th inguinal mammary
657 gland; barcoded 4T1 cells (ATCC, USA) were resuspended in PBS and 5x10^4 cells in
658 a 10 ml volume were injected into the 4th inguinal mammary fat pad. For studies with
659 the 4T1 model primary tumours were surgically resected at day 15. At resection or
660 ethical endpoint tumours and whole lung tissue were removed, minced and snap frozen
661 in liquid nitrogen for barcode analysis.
662
663 Immunotherapy treatment
664 Mice were treated with four 200 µg doses of either combination immunotherapy
665 antibodies via intraperitoneal injection: anti-CTLA4 (BE0032), anti-PD1 (BE0146), or
666 isotype control antibodies Armenian hamster IgG (BE0091), Rat IgG (BE0089) all
667 from BioXCell (Lebanon, NH, USA). Antibodies were given every 2-3 days from day

668 10 after tumour implantation for the EMT6 model and following resection of primary
669 tumour on day 15 for the 4T1 model.
670
671 CD8 T cell, CD4 T cell and NK cell depletion
672 Starting one day prior to primary tumour resection mice were given 100 µg of depleting
673 antibodies for CD8 T cells (anti-CD8; BEO223; BioXCell), or NK cells (anti-Asialo-
674 GM1; 986-10001; Novachem), or CD4 T cells (anti-CD4; BE0003-1; BioXCell) or
675 isotype control antibodies. Antibodies were then given every 2-3 days for a total of 4
676 doses.
677
678 Isolation of buffy coat and flow cytometry to confirm depletion of key cell types
679 Whole blood was collected into K2EDTA coated tubes (BD, cat# 365973) from mice
680 two days after the last intraperitoneal injection of depleting or control antibodies. Buffy
681 coat was isolated by spinning blood at 800g at room temperature for 10 minutes and
682 removing the small layer of the buffy coat. Any red blood cells were lysed using
683 Pharmlyse (BD, 5 min) and quenching with FACS buffer (DPBS supplemented with
684 2% FCS and 2% HEPES). Cells were washed and stained with BV711 conjugated anti-
685 mouse CD8a (Biolegend, 1:200, cat# 563046), PE conjugated anti-mouse CD4
686 (Biolegend, 1:200, cat# 116006), and APC conjugated anti-mouse NKp46 (Biolegend,
687 1:200, cat# 137607). Cells were washed 3 times before staining with DAPI. Data was
688 collected using the BD LSR Fortessa (FACSDiva 8.0.1) .
689
690 DNA extraction
691 Frozen tissue samples were lysed in 5 ml QIAGEN buffer P1 (with RNaseA) and
692 0.5% SDS within a Miltenyi M-Tube (# 130-096-335). Samples were processed on

693 the gentleMACS or gentleMACS Octo using the RNA_02 program. DNA was then
694 extracted using a standard phenol/chloroform process.
695
696 Targeted Barcode PCR and Sequencing
697 All samples underwent targeted barcode PCR amplification according to the updated
698 version of the original protocol (*30*) available on the Addgene website
699 (https://www.addgene.org/pooled-library/clontracer/). Specific PCR products (180 bp)
700 were gel purified, quantified by Qubit 2.0 fluorometer (ThermoFisher Scientific,
701 Waltham, MA USA) and pooled into a library. Prior to sequencing, an equal
702 combination of additional PCR products containing two inverse barcodes
703 (GACTCAGTGTCAGACTGAGTGTCTGACTGT and
704 CTGAGTCACAGTCTGACTCACAGACTGACA) plus the PhiX Control V3 (Cat.
705 FC-110-3001, Illumina, CA, USA) were spiked in to balance the nucleotide distribution
706 within the library. Samples were sequenced using a custom sequencing primer
707 (GCGACCACCGAGATCTACACACTGACTGCAGTCTGAGTCTGACAG) and the
708 NextSeq® 500/550 Mid Output Kit v2 - 150 cycles (FC-404-2001, Illumina, CA, USA)
709 on the Illumina NextSeq® platform.
710
711 Barcode Analysis
712 Barcode composition analysis and calculation of barcode overlap between samples was
713 performed as indicated in the original protocol (*30*) and updated Python scripts
714 available from the Addgene website (https://www.addgene.org/pooled-
715 library/clontracer/).
716 Further analysis was performed using R statistical framework and packages
717 EntropyExplorer for analysis of differential Shannon Entropy (*65*), DEBRA for

718 differential barcode expression (*66*), and libraries fishplot and UpSetR for visualisation
719 purposes. Pheatmap (v1.0.12) was used with default parameters. Clustering distances
720 for rows and columns were euclidean and complete clustering linkage.
721
722 Generating clonal cell lines
723 Cells of interest were isolated from the barcoded parental population using a sub-
724 pooling approach.
725 The barcoded 4T1 BC5000 cells were seeded into a 96 well plate at a density of 150
726 cells per well. At approximately 80% confluence, cells were trypsinised and split
727 identically into two plates. One plate was viably frozen in freezing media (10% DMSO,
728 40% FCS, 50% 4T1 media). DNA was extracted from one plate using the Promega SV
729 Wizard Genomic DNA kit. Target barcodes of each sample were PCR amplified and
730 sequenced using the method described above.
731 After sequencing, wells containing cells with the target barcodes were thawed, pooled
732 and seeded at 40 cells/well in a 96 well plate. Media was changed every three days for
733 8 days before cells were split into two identical plates as above. One plate was viably
734 frozen in freezing media, while DNA was extracted and prepared for targeted
735 sequencing as above.
736 Wells with the highest proportion of target barcodes were revived into a 6 well plate
737 and grown for 4 days before being single-cell sorted by BD FACSAria II into 96 well
738 plate. Sorted single cells were grown in conditioned media for 5 days before being
739 changed to 4T1 media and grown until 80% confluence. As previously, cells were lifted
740 and split identically into two plates – one for freezing and one for targeted sequencing.

741 Once wells containing single cells clones of the cells of interest were identified, target
742 wells were revived. Cells were expanded before being aliquoted and viably frozen for
743 future experiments.
744 Barcoded sequences of isolated cells were confirmed by targeted Sanger sequencing of
745 barcode regions.
746
747 Bulk RNA sequencing
748 RNA was extracted from established subclonal cell lines using the QIAGEN RNeasy
749 Mini Kit. 3-4 unique clonal cell populations were sequenced for each barcode. Libraries
750 were prepared using the KAPA RNA HyperPrep Kit with RiboErase, and sequenced on
751 the NextSeq500 platform using a High Output V2.5 300 cycle kit.
752
753 Transcriptome analysis
754 FastQ files from sequencing libraries were first trimmed with FASTQC v0.11 *Andrews*
755 *S. (2010). FastQC: a quality control tool for high throughput sequence data. Available*
756 *online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc*. Raw reads were
757 subsequently mapped to the mouse transcriptome (Gencode release M9, GRCm38.p4),
758 to the mouse genome (mm10 assembly), with STAR aligner v.2.4.1d, allowing for
759 multimapping reads (*67*). The reads were counted over gene models with RSEM,
760 v.1.2.18 (*51*). Differentially expressed genes and repeat elements were defined with
761 EdgeR with FDR<0.01 (*68*). EdgeR uses genewise negative binomial generalised linear
762 models through the function glmQLFTest (*69*). Genes that with less than 10 reads
763 across three samples per group were omitted from the analysis. The model involved
764 comparison between IE samples, NT samples and the bulk 4T1 cell line.
765

766 Survival analysis

767 To assess the clinical relevance of our isolated immune evasion clones, we assessed the
768 association between the gene signatures derived from our bulk RNA-Sequencing
769 studies with the overall survival of basal (PAM50) breast cancer patients from the
770 METABRIC and The Cancer Genome Atlas (TCGA; https://www.cancer.gov/tcga)
771 cohorts. Mouse gene signatures were first converted to human orthologs using the
772 biomaRt package (*70*). Shared up-regulated genes across both immune evasion clones
773 IE1 and IE2 were then filtered, and only genes detected in each expression cohort were
774 considered. For each tumour from the bulk cohort, signature scores were computed
775 based on the average expression of the top 25 genes ranked by log fold change. Patients
776 were then stratified based on the signatures scores into the top 30%, middle 40% and
777 bottom 30% groups. Survival curves were generated using the Kaplan Meier method
778 with the 'survival' package in R (https://cran.r-project.org/package=survival). The Cox
779 proportional hazards model was used to compute Hazard Ratios. We assessed the
780 significance between groups using the log-rank test statistics.

781

782 Gene set enrichment analysis

783 Gene set enrichment analysis (GSEA) was carried out using the GSEA desktop app
784 (4.1.0) and DEGs generated as described above. GSEA was run using preranked list of
785 significantly differentially expressed genes, ranked by log fold change. The molecular
786 signatures database (MsigDB v7.5.1) hallmark and curated (C2) gene sets were used
787 for analysis.

788

789 Whole genome sequencing

790 DNA was extracted from established subclonal cell lines using the QIAGEN DNeasy
791 blood & tissue kit. Libraries were prepared using the Roche KAPA PCR-free library
792 preparation kit and genomes were sequenced on the HiSeq X platform to a depth of
793 ~30x.
794
795 Whole genome analysis
796 Fastq files from the WGS were firstly aligned to mouse genome reference mm10. The
797 output bam files were subsequently used for copy number analysis. Copy number
798 analysis was performed using a R package cn.mops (*71*) in "paired mode" with a
799 window length of 10kb. Reads were aligned to the BALB/c reference genome using
800 BWA before being indexed and sorted with Novosort. Reads that mapped incompletely
801 to the reference genome were then mapped to the barcode plasmid sequence with BWA
802 and sorted and indexed with Novosort. Read pairs where only one pair mapped to the
803 barcode plasmid sequence were blasted against mm10 to establish the barcode plasmid
804 insertion site.
805
806 Flow cytometry for MHC-1 and PD-L1
807 The 4T1 subclones (IE1, IE2, NT1, NT2), as well as the parental 4T1 bulk population
808 were revived and passaged three times before being seeded into a 6 well plate at a
809 density of 200 000 cells per well. At approximately 80% confluence, cells were
810 collected into FACS buffer (DPBS supplemented with 2% FCS and 2% HEPEs) for
811 flow cytometry. Cells were stained with a mastermix of APC conjugated anti-mouse
812 CD274 (Biolegend) and Alexa Fluor488 conjugated anti-mouse H2-kD (Biolegend) at
813 a final concentration of 1:200 for 20 minutes. Cells were washed three times with FACs
814 buffer before being stained with DAPI and run on the BD FACSCanto II flow

815 cytometer, utilising BD FACSDIVA software. Data was analysed in FlowJo (version
816 10.6.1) and median fluorescence intensity of live, single cells was calculated.
817
818 Treating cells with 5-Aza-2'-deoxycytidine and flow cytometry for MHC-I
819 5-Aza-2'-deoxycytidine (5aza) was sourced from Sigma and reconstituted in DMSO
820 according to manufacturer's instructions. Subclones (IE1, IE2) and the parental 4T1
821 cell line were seeded into a 24 well plate at a density of 8000 cells per well in 4T1
822 media (previously described). Cells were allowed to settle overnight before being
823 treated with 5aza at 200nM, 100nM or 50nM for 72 hours. 5aza was removed and cells
824 were cultured in media only for 24 hours before being collected for flow analysis. Cells
825 were stained with Alexa Fluor488 conjugated anti-mouse H2-kD (Biolegend) at a
826 concentration of 1:200 in FACS buffer for 20 minutes. Cells were washed three times
827 after staining before being stained with DAPI. Data was collected using the BD
828 FACSCanto II flow cytometer with BD FACSDIVA software. The resulting data was
829 analysed using FlowJo (version 10.6.1) and media fluorescence intensity of live, single
830 cells was calculated.
831
832 Treating cells with interferon gamma and flow cytometry for MHC-I
833 Active mouse interferon gamma (IFNγ) was sourced from Abcam and reconstituted in
834 sterile water, as per manufacturer's instructions. Subclones (IE1, IE2) and the parental
835 4T1 cell line were grown in a 24 well plate until approximately 70% confluence was
836 achieved. Cells were then treated with IFNγ (100ng/ml) for 24 hours. Cells were stained
837 with Alexa Fluor488 conjugated anti-mouse H2-kD (Biolegend) at a concentration of
838 1:200 in FACS buffer for 20 minutes. Cells were washed three times before being
839 stained with DAPI. Data was generated using the BD FCSCanto II flow cytometer with

840 BD FACSDIVA software. Analysis was carried out using FlowJo (version 10.6.1) and
841 median fluorescence intensity of live single cells was calculated.
842
843 Dissociation of metastatic lung tissue and flow cytometry for MHC-I and PD-L1
844 Lungs were minced with scissors and dissociated in GentleMACS C tubes (Miltenyi
845 Biotec) with collagenase (1mg/ml, Type 1A from clostridium histolyticum, Sigma, cat
846 # C9891) in collagenase buffer (RPMI-1640 supplemented with 2.5 % FBS and 1%
847 HEPES), shaking, 37°C for 50 minutes. After shaking, samples were passed through
848 70um MACS SmartStrainers (Miltenyi Biotec) and washed with PBS, samples were
849 treated with DNase (1mg/ml, from DNase 1 from bovine pancreas, Sigma, cat# CD25)
850 for 3 minutes and quenched in FACS buffer (DPBS, Gibco, supplemented with 2% FBS
851 and 2% 1M HEPES). Samples were washed twice in FACS before FC block with
852 Mouse BD Fc Block (BD, cat# 553141) for 15 minutes. Cells were stained with Alexa
853 Fluor488 conjugated anti-mouse H2-kD (clone SF1-1.1, Biolegend, cat# 116610) and
854 APC anti-mouse CD274 (B7-H1, PD-L1; clone 10F.9G2; Biolegend; cat# 124311) at a
855 concentration of 1:200 in FACS buffer for 20 minutes. Samples were washed three
856 times before being stained with DAPI. Data was generated using the BD LSR Fortessa
857 flow cytometer with BD FACSDIVA software. Analysis was carried out using FlowJo
858 (version 10.6.1).
859
860 Generation of poly-specific cytotoxic T lymphocyte lines
861 4T1-specific CD8+ T cell lines were derived as previously described (*37*) Briefly,
862 Balb/c mice were immunised with 1x106 irradiated 4T1 cells via intraperitoneal
863 injection. Spleens were harvested >30 days post-immunisation and re-stimulated in

vitro with irradiated 4T1 cells and cultured in RPMI containing 10% FBS and 10 IU IL-2/mL for 14 days before experimental use.

Functional assessment of tumor cell line immunogenicity

4T1 functional assessment was performed as previously described (*38*). Briefly, 4T1 cell line variants were cultured in the absence or presence of 100 ng/mL recombinant IFNγ in 37°C 5% CO2 for 24 h. 5 x 104 treated cells were washed twice with PBS and incubated with 5 x 103 4T1-specific CTL in the presence of 10 µg/mL Brefeldin A in 37°C 5% CO2 for 5 h. For anti-PD1 experiments: 4T1-specific CD8+ T cells were incubated in the presence of 10 µg/mL anti-PD1 (clone: RMP1-14, BioXcell) or isotype control (clone: 2A3, BioXcell) for 30 min, RT. Samples were then stained with cell surface antibodies for 20 min, 4°C (α-CD8, clone: 53-6.7, 1:300, eBioscience), washed with PBS, fixed with 1% paraformaldehyde (15 min, RT, dark), permeabilized and stained for intracellular proteins (α-IFNγ, clone: XMG1.2, 1:300, eBioscience) in the presence of 0.4% Saponin for 30 min, 4°C, and analysed on a FACSymphony A5 instrument (BD).

Data availability

RNA-seq data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-10027.

DNA-seq data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-10027.

889 **References:**

890 1. D. Hanahan, Robert A. Weinberg, Hallmarks of Cancer: The Next Generation.
891 *Cell* **144**, 646-674 (2011).
892 2. R. D. Schreiber, L. J. Old, M. J. Smyth, Cancer Immunoediting: Integrating
893 Immunity's Roles in Cancer Suppression and Promotion. *Science* **331**, 1565-
894 1570 (2011).
895 3. M. Angelova *et al.*, Evolution of Metastases in Space and Time under
896 Immune Selection. *Cell* **175**, 751-765.e716 (2018).
897 4. S. C. Casey *et al.*, MYC regulates the antitumor immune response through
898 CD47 and PD-L1. **352**, 227-231 (2016).
899 5. D. M. Pardoll, The blockade of immune checkpoints in cancer
900 immunotherapy. *Nature Reviews Cancer* **12**, 252-264 (2012).
901 6. P. Sharma, S. Hu-Lieskovan, J. A. Wargo, A. Ribas, Primary, Adaptive, and
902 Acquired Resistance to Cancer Immunotherapy. *Cell* **168**, 707-723 (2017).
903 7. L. Galluzzi, T. A. Chan, G. Kroemer, J. D. Wolchok, A. López-Soto, The
904 hallmarks of successful anticancer immunotherapy. *Science Translational*
905 *Medicine* **10**, (2018).
906 8. P. Schmid *et al.*, Atezolizumab and Nab-Paclitaxel in Advanced Triple-
907 Negative Breast Cancer. *New England Journal of Medicine* **379**, 2108-2121
908 (2018).
909 9. H. R. Ali, L. Chlon, P. D. P. Pharoah, F. Markowetz, C. Caldas, Patterns of
910 Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-
911 Expression-Based Retrospective Study. *PLOS Medicine* **13**, e1002194
912 (2016).
913 10. E. T. E. Gross *et al.*, Immunosurveillance and immunoediting in MMTV-PyMT-
914 induced mammary oncogenesis. *Oncoimmunology* **6**, e1268310 (2017).
915 11. C. M. Koebel *et al.*, Adaptive immunity maintains occult cancer in an
916 equilibrium state. *Nature* **450**, 903-907 (2007).
917 12. M. Efremova *et al.*, Targeting immune checkpoints potentiates immunoediting
918 and changes the dynamics of tumor evolution. *Nature Communications* **9**, 32
919 (2018).
920 13. I. Milo *et al.*, The immune system profoundly restricts intratumor genetic
921 heterogeneity. *Science Immunology* **3**, eaat1435 (2018).
922 14. H. Matsushita *et al.*, Cancer exome analysis reveals a T-cell-dependent
923 mechanism of cancer immunoediting. *Nature* **482**, 400-404 (2012).
924 15. K.-J. Malmberg *et al.*, Natural killer cell-mediated immunosurveillance of
925 human cancer. *Seminars in Immunology* **31**, 20-29 (2017).
926 16. K. A. Lawson *et al.*, Functional genomic landscape of cancer-intrinsic evasion
927 of killing by T cells. *Nature* **586**, 120-126 (2020).
928 17. R. T. Manguso *et al.*, In vivo CRISPR screening identifies Ptpn2 as a cancer
929 immunotherapy target. *Nature* **547**, 413 (2017).
930 18. C.-Q. Song *et al.*, Genome-Wide CRISPR Screen Identifies Regulators of
931 Mitogen-Activated Protein Kinase as Suppressors of Liver Tumors in Mice.
932 *Gastroenterology* **152**, 1161-1173.e1161 (2017).
933 19. G. Wang *et al.*, CRISPR-GEMM Pooled Mutagenic Screening Identifies
934 KMT2D as a Major Modulator of Immune Checkpoint Blockade. *Cancer*
935 *Discovery* **10**, 1912-1933 (2020).
936 20. S. J. Patel *et al.*, Identification of essential genes for cancer immunotherapy.
937 *Nature* **548**, 537-542 (2017).
938 21. N. Anfossi *et al.*, Human NK Cell Education by Inhibitory Receptors for MHC
939 Class I. *Immunity* **25**, 331-342 (2006).

940 22. E. Mamessier *et al.*, Human breast cancer cells enhance self tolerance by
941 promoting evasion from NK cell antitumor immunity. *The Journal of Clinical*
942 *Investigation* **121**, 3609-3622 (2011).
943 23. N. McGranahan *et al.*, Clonal neoantigens elicit T cell immunoreactivity and
944 sensitivity to immune checkpoint blockade. *Science* **351**, 1463-1469 (2016).
945 24. N. A. Rizvi *et al.*, Mutational landscape determines sensitivity to PD-1
946 blockade in non-small cell lung cancer. *Science* **348**, 124-128 (2015).
947 25. C. A. Stewart *et al.*, Single-cell analyses reveal increased intratumoral
948 heterogeneity after the onset of therapy resistance in small-cell lung cancer.
949 *Nature Cancer* **1**, 423-436 (2020).
950 26. N. McGranahan *et al.*, Allele-specific HLA loss and immune escape in lung
951 cancer evolution. *Cell* **171**, 1259-1271. e1211 (2017).
952 27. Y. Wolf *et al.*, UVB-Induced Tumor Heterogeneity Diminishes Immune
953 Response in Melanoma. *Cell* **179**, 219-235.e221 (2019).
954 28. J. B. Williams *et al.*, Tumor heterogeneity and clonal cooperation influence
955 the immune selection of IFN-γ-signaling mutant cancer cells. *Nature*
956 *Communications* **11**, 602 (2020).
957 29. A. Gambotto *et al.*, Immunogenicity of enhanced green fluorescent protein
958 (EGFP) in BALB/c mice: identification of an H2-Kd-restricted CTL epitope.
959 *Gene Therapy* **7**, 2036-2040 (2000).
960 30. H.-e. C. Bhang *et al.*, Studying clonal dynamics in response to cancer therapy
961 using high-complexity barcoding. *Nature Medicine* **21**, 440 (2015).
962 31. Y. Lan *et al.*, Enhanced preclinical antitumor activity of M7824, a bifunctional
963 fusion protein simultaneously targeting PD-L1 and TGF-β. *Science*
964 *Translational Medicine* **10**, (2018).
965 32. M. Ouzounova *et al.*, Monocytic and granulocytic myeloid derived suppressor
966 cells differentially regulate spatiotemporal tumour plasticity during metastatic
967 cascade. *Nature Communications* **8**, 14979 (2017).
968 33. E. Wagenblast *et al.*, A model of breast cancer heterogeneity reveals vascular
969 mimicry as a driver of metastasis. *Nature* **520**, 358-362 (2015).
970 34. X. Peng, L. Zha, A. Chen, Z. Wang, HOXA5 is a tumor suppressor gene that
971 is decreased in gastric cancer. *Oncol Rep* **40**, 1317-1329 (2018).
972 35. R. E. Vance, J. R. Kraft, J. D. Altman, P. E. Jensen, D. H. Raulet, Mouse
973 CD94/NKG2A Is a Natural Killer Cell Receptor for the Nonclassical Major
974 Histocompatibility Complex (MHC) Class I Molecule Qa-1b. *Journal of*
975 *Experimental Medicine* **188**, 1841-1848 (1998).
976 36. N. Luo *et al.*, DNA methyltransferase inhibition upregulates MHC-I to
977 potentiate cytotoxic T lymphocyte responses in breast cancer. *Nature*
978 *Communications* **9**, 248 (2018).
979 37. D. Zanker *et al.*, An optimized method for establishing high purity murine
980 CD8+ T cell cultures. *Journal of immunological methods* **387**, 173-180 (2013).
981 38. D. J. Zanker *et al.*, Intratumoral administration of the Toll-like receptor 7/8
982 agonist 3M-052 enhances interferon-driven tumor immunogenicity and
983 suppresses metastatic spread in preclinical triple-negative breast cancer.
984 *Clinical & translational immunology* **9**, e1177 (2020).
985 39. D. Blanco-Melo *et al.*, Imbalanced Host Response to SARS-CoV-2 Drives
986 Development of COVID-19. *Cell* **181**, 1036-1045.e1039 (2020).
987 40. I. Ben-Porath *et al.*, An embryonic stem cell–like gene expression signature in
988 poorly differentiated aggressive human tumors. *Nature Genetics* **40**, 499-507
989 (2008).
990 41. C. Curtis *et al.*, The genomic and transcriptomic architecture of 2,000 breast
991 tumours reveals novel subgroups. *Nature* **486**, 346-352 (2012).
992 42. N. The Cancer Genome Atlas, Comprehensive molecular portraits of human
993 breast tumours. *Nature* **490**, 61 (2012).

43. S. Dushyanthen *et al.*, Relevance of tumor-infiltrating lymphocytes in breast cancer. *BMC Medicine* **13**, 1-13 (2015).
44. M. Obeid *et al.*, Calreticulin exposure dictates the immunogenicity of cancer cell death. *Nature Medicine* **13**, 54-61 (2007).
45. N. Ishikawa *et al.*, Characterization of SEZ6L2 cell-surface protein as a novel prognostic marker for lung cancer. *Cancer Science* **97**, 737-745 (2006).
46. L. Wang *et al.*, Upregulated Seizure-Related 6 Homolog-Like 2 Is a Prognostic Predictor of Hepatocellular Carcinoma. *Disease Markers* **2020**, 7318703 (2020).
47. P. Jiang *et al.*, Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nature Medicine* **24**, 1550-1558 (2018).
48. A. W. Zhang *et al.*, Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer. *Cell*, (2018).
49. M. Bosiljcic *et al.*, Targeting myeloid-derived suppressor cells in combination with primary mammary tumor resection reduces metastatic growth in the lungs. *Breast Cancer Research* **21**, 1-16 (2019).
50. Y. R. Miao *et al.*, Inhibition of established micrometastases by targeted drug delivery via cell surface–associated GRP78. *Clinical Cancer Research* **19**, 2107-2116 (2013).
51. S. Meng *et al.*, Circulating tumor cells in patients with breast cancer dormancy. *Clinical cancer research* **10**, 8152-8162 (2004).
52. G. V. Echeverria *et al.*, High-resolution clonal mapping of multi-organ metastasis in triple negative breast cancer. *Nature communications* **9**, 1-17 (2018).
53. G. V. Echeverria *et al.*, Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. *Science translational medicine* **11**, (2019).
54. D. Merino *et al.*, Barcoding reveals complex clonal behavior in patient-derived xenografts of metastatic triple negative breast cancer. *Nature communications* **10**, 1-12 (2019).
55. P. Schmid *et al.*, Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet Oncology* **21**, 44-59 (2020).
56. S. C. Casey *et al.*, MYC regulates the antitumor immune response through CD47 and PD-L1. *Science* **352**, 227-231 (2016).
57. N. McGranahan *et al.*, Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*, 1490-1501 (2016).
58. Y. Chen *et al.*, Sephin1, which prolongs the integrated stress response, is a promising therapeutic for multiple sclerosis. *Brain* **142**, 344-361 (2019).
59. K. B. Chiappinelli, C. A. Zahnow, N. Ahuja, S. B. Baylin, Combining Epigenetic and Immunotherapy to Combat Cancer. *Cancer Research* **76**, 1683-1689 (2016).
60. M. Guo, Y. Peng, A. Gao, C. Du, J. G. Herman, Epigenetic heterogeneity in cancer. *Biomarker Research* **7**, 23 (2019).
61. D. R. Leach, M. F. Krummel, J. P. Allison, Enhancement of antitumor immunity by CTLA-4 blockade. *Science* **271**, 1734-1736 (1996).
62. 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).
63. D. A. Lawson *et al.*, Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. **526**, 131 (2015).
64. D. Hanahan, Hallmarks of Cancer: New Dimensions. *Cancer Discovery* **12**, 31-46 (2022).

1048 65. K. Wang, C. A. Phillips, A. M. Saxton, M. A. Langston, EntropyExplorer: an R
1049 package for computing and comparing differential Shannon entropy,
1050 differential coefficient of variation and differential expression. *BMC Research*
1051 *Notes* **8**, 832 (2015).
1052 66. Y. Akimov, D. Bulanova, S. Timonen, K. Wennerberg, T. Aittokallio, Improved
1053 detection of differentially represented DNA barcodes for high-throughput
1054 clonal phenomics. *Molecular Systems Biology* **16**, e9195 (2020).
1055 67. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**,
1056 15-21 (2012).
1057 68. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor
1058 package for differential expression analysis of digital gene expression data.
1059 *Bioinformatics* **26**, 139-140 (2009).
1060 69. A. T. Lun, G. K. Smyth, No counts, no variance: allowing for loss of degrees
1061 of freedom when assessing biological variability from RNA-seq data.
1062 *Statistical Applications in Genetics and Molecular Biology* **16**, 83-93 (2017).
1063 70. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the
1064 integration of genomic datasets with the R/Bioconductor package biomaRt.
1065 *Nature Protocols* **4**, 1184-1191 (2009).
1066 71. G. Klambauer *et al.*, cn.MOPS: mixture of Poissons for discovering copy
1067 number variations in next-generation sequencing data with a low false
1068 discovery rate. *Nucleic Acids Research* **40**, e69-e69 (2012).
1069

1070 **Figure legends:**

1071 Figure 1. Experimental workflow schematic. Barcode library is introduced into
1072 mammary carcinoma cell lines *in vitro* at a low multiplicity of infection (MOI). Cells
1073 are sorted based on RFP expression to select for those having incorporated a barcode.
1074 Barcoded cells are then transplanted into the mammary fat pad of mice. Following
1075 immunoselection with either the endogenous immune system or immunotherapy then
1076 barcode abundance and diversity within the primary tumours and lungs with metastases
1077 are analysed.

1078

1079 Figure 2. Immune selection and clonal immunoediting of EMT6 primary tumours.
1080 A. Outline of experimental design. B. EMT6 primary tumour growth in wild-type
1081 Balb/c mice and in NSG mice plotted as tumour volume. Average volume +/- SEM;
1082 n=5-6 mice per group. C. Kaplan-Meier survival analysis comparing Balb/c and NSG
1083 mice bearing EMT6 tumour (Mantel-Cox p=0.009). D. EMT6 primary tumour growth
1084 plotted as tumour volume in wild-type Balb/c mice with or without immunotherapy

1085 (anti-PD1+anti-CTLA4) on days 10, 12, 14, and 17 as indicated with red arrows.
1086 Average volume +/- SEM; n=5-6 mice. E. Kaplan-Meier survival analysis of Balb/c
1087 mice bearing EMT6 tumour treated with immunotherapy or isotype control (Mantel-
1088 Cox p=0.0006). F. Number of unique barcodes identified in EMT6 primary tumours
1089 grown in NSG mice and in Balb/c mice treated with isotype control antibodies or anti-
1090 PD1 + anti-CTLA4. GLM with Tukeys HSD for multiple comparisons. G. Shannon
1091 diversity index analysis of EMT6 primary tumours grown in NSG mice and in Balb/c
1092 mice treated with isotype control antibodies or anti-PD1 + anti-CTLA4. One way
1093 ANOVA with Tukeys HSD. H. Dot plot of a subset of barcodes with enrichment
1094 following immunotherapy treatment.
1095
1096 Figure 3. Immune selection and clonal immunoediting in the metastatic setting utilising
1097 the 4T1 model.
1098 A. Outline of experimental design. B. Kaplan-Meier survival analysis comparing Balb/c
1099 mice and NSG mice bearing 4T1 tumour; n=5 mice/group (p = 0.0002, Mantel-Cox
1100 Log-rank test). C. Kaplan-Meier survival analysis of Balb/c mice bearing 4T1 tumour
1101 treated with immunotherapy or isotype control on days 15, 17, 19, and 21 as indicated
1102 with red arrows; n=5-6 mice/group (p = 0.0141, Mantel-Cox Log-rank test). D. Number
1103 of unique barcodes identified in 4T1 primary tumours (PT) and lung metastases grown
1104 in NSG mice or Balb/c mice. GLM with Tukey's HSD for multiple comparisons. E.
1105 Number of unique barcodes identified in 4T1 primary tumours and lung metastases
1106 grown in Balb/c mice treated with isotype control antibodies or anti-PD1 + anti-
1107 CTLA4. GLM with Tukey's HSD for multiple comparisons. F. Shannon diversity index
1108 analysis in primary tumours from NSG mice or Balb/c mice and in lung metastases
1109 from NSG mice, Balb/c mice, Balb/c mice treated with isotype control, and Balb/c mice

1110 treated with anti-PD1 + anti-CTLA4. Unpaired t-test. G. Analysis of number of unique
1111 barcodes in lung metastases from Balb/c mice treated with control antibodies, anti-
1112 CD8a to deplete CD8 T cells, or anti-asialo GM1 to deplete NK cells. One way
1113 ANOVA with Tukeys HSD for multiple comparisons, 5 mice/group. Figure 4. Analysis
1114 of specific barcodes enriched or depleted by the immune system and immunotherapy in
1115 the 4T1 model.
1116 A. Unsupervised hierarchical clustering heatmap of barcodes with an abundance of
1117 above 5% in at least one 4T1 primary tumour or lung metastasis sample; * indicate
1118 barcodes detected at a frequency above 0.1% in a particular sample. B. Dot plot of a
1119 subset of specific barcodes. C. Fishplot of the nine most abundant barcodes detected in
1120 lungs of WT and NSG mice, each of these nine barcodes is given a unique colour with
1121 the remaining barcodes being combined and represented in grey. A bottleneck has been
1122 introduced between primary tumour and lung metastases to depict transition through
1123 the blood stream. D. Fishplot of the nine most abundant barcodes detected in the lungs
1124 of WT mice treated with combination immunotherapy or control antibodies presented
1125 as in C, a bottleneck has been introduced between primary tumour and lung metastases
1126 to depict transition through the blood stream.
1127
1128 Figure 5: Gene expression analysis of immunotherapy resistant clones.
1129 A. Volcano plot showing differentially expressed genes between parental 4T1 cells and
1130 the immunotherapy enriched 1 (IE1) clone. B. Volcano plot showing differentially
1131 expressed genes between parental 4T1 cells and the immunotherapy enriched 2 (IE2)
1132 clone. C. Expression of indicated MHC related genes, measured as transcripts per
1133 million (TPM), in the parental 4T1 population and in indicated cell clones; FDR
1134 calculated using EdgeR with Benjamini-Hochberg multiple testing correction. D.

1135 MHCI protein expression as quantified by flow cytometry in the indicated clones and
1136 the parental 4T1 population measured as mean fluorescence intensity (MFI) (left) and
1137 representative histogram (right); one way ANOVA with Tukey HSD test for multiple
1138 comparisonst. E. Expression of immune related genes up-regulated in clone IE2,
1139 measured as transcripts per million (TPM), in the parental 4T1 population and in
1140 indicated cell clones. FDR calculated by EdgeR with Benhamini-Hochberg multiple
1141 testing correction F. PD-L1 protein expression as determined by flow cytometry in
1142 indicated clones and the parental 4T1 population measured as mean fluorescence
1143 intensity (MFI), representative plot of three independent experiments; one way
1144 ANOVA with Tukey HSD for multiple comparisons.

1145 Figure 6. The 4T1-IE2 clone can directly suppress cytotoxic T-cell activation and
1146 prevent further T-cell activation by anti-PD1 immunotherapy.

1147 The 4T1-IE2 clone can directly suppress cytotoxic T-cell activation and prevent further
1148 T-cell activation by anti-PD1 immunotherapy. A. Measurent of intracellular IFNγ
1149 production by 4T1 specific CD8+ T-cells following co-culture with no 4T1 cells (No
1150 4T1), bulk 4T1 cells (Bulk 4T1) and 4T1 clonal populations (4T1-N1; 4T1-NT2; 4T1-
1151 IE1; 4T1-IE2). For clonal populations n=3-4; one-way ANOVA with Dunnett's post-
1152 hoc test. B. Measurent of intracellular IFNγ production by 4T1 specific CD8+ T-cells
1153 following co-culture with clonal 4T1 populations with or without anti-PD1 (PD1) or an
1154 isotype control (iso). n=3-4; FWER calculated using flow cytometry proportions fitted
1155 to a GLM with Bonferroni post-hoc correction.

1156 Figure 7. Overlapping gene signatures of the immunotherapy resistant clones show
1157 prognostic significance in basal-like breast cancer patients.

1158 A. Upset plots and Venn diagrams showing the overlap in significantly up-regulated
1159 (right) and down-regulated (left) genes between the two immunotherapy enriched

1160 clones (IE1 and IE2) and two control clones (NT1 and NT2). B. Heatmap of the top 50
1161 overlapping up regulated and down regulated genes between IE1 and IE2 clones across
1162 all clonal cell lines. C. Kaplan-Meier analysis of overall survival of basal-like breast
1163 cancer patients in the METABRIC cohort based on whose breast cancer having the top
1164 30% or the bottom 30% expression of the overlapping upregulated 25 gene signature.
1165 D. Kaplan-Meier analysis of overall survival of basal-like breast cancer patients in the
1166 TCGA cohort based on whose tumour having the top 30% or the bottom 30%
1167 expression of the overlapping upregulated 25 gene signature. In C and D, the Cox
1168 proportional hazards model was used to compute Hazard Ratios. Significance between
1169 stratification groups were computed using log-rank test statistics.
1170
1171 **Supplementary Figure Legends:**
1172 Sup Figure 1. Gating strategy showing isolation and selection of barcode transfected
1173 Red Fluorescent Protein (RFP) positive EMT6 cells
1174
1175 Sup Figure 2. Unique barcode counts and barcode distributions across EMT6 primary
1176 tumours and matching cell pellets. A. Unique barcode counts from EMT6 tumour
1177 bearing mice. B. Shannon diversity index. C. Distribution plots of barcode counts from
1178 individual mice, grouped by strain and treatment group. D. Upset plot showing unique
1179 barcode counts and overlap between the EMT6cell pellet and tumour bearing NSG mice
1180 or immunocompetent immunotherapy treated (IT) or control treated (Cont) mice.
1181
1182 Sup Figure 3. Gating strategy showing isolation and selection of barcode transfected
1183 Red Fluorescent Protein (RFP) positive 4T1 cells.
1184

1185 Sup Figure 4. Unique barcode counts and barcode distributions across 4T1 primary
1186 tumours and matched lungs in NSG and Balb/c mice. A. Unique barcode counts
1187 detected in matched primary tumour (T) or lungs from 4T1 tumour bearing NSG or
1188 Balb/c mice. B. Unique barcode counts detected in matched primary tumour (T) and
1189 lungs from 4T1 tumour bearing mice treated with either combined immunotherapy (IT)
1190 or control antibodies (IgG) C. Distribution plot of barcode counts from all (NSG and
1191 Balb/c) mice. D. Distribution plot of barcode counts from Balb/c mice treated with
1192 combination immunotherapy or control antibodies.
1193
1194 Sup Figure 5. 4T1 primary tumour growth and barcode diversity in primary tumours.
1195 A. Primary tumour mass. B. Total number of unique barcodes identified in primary
1196 tumours of Balb/c or NSG mice. C. Shannon diversity analysis comparing primary
1197 tumours from Balb/c and NSG mice.
1198
1199 Sup Figure 6. Depletion of CD8+ T cells with anti-CD8, CD4+ T cells with anti-CD4
1200 or NK cells with anti-asialo GM1 does not significantly affect survival of mice bearing
1201 4T1 tumours. A. Kaplan-Meier survival analysis of Balb/c mice transplanted with 4T1
1202 cells, primary tumour was resected on day 15. Depletion of target cell types was
1203 initiated 1 day prior to resection, on day 14. 8-15 mice per treatment group. B-D. Flow
1204 cytometry to confirm depletion of target cell types. Buffy coat was collected on day 22,
1205 2 days after the final dose of depleting antibodies was given. One way ANOVA.
1206
1207 Sup Figure 6. Depletion of CD8+ T cells with anti-CD8, CD4+ T cells with anti-CD4
1208 or NK cells with anti-asialo GM1 does not significantly affect survival or barcode
1209 diveristy of mice bearing 4T1 tumours. A. Kaplan-Meier survival analysis of Balb/c

1210 mice transplanted with 4T1 cells, primary tumour was resected on day 15. Depletion of
1211 target cell types was initiated 1 day prior to resection, on day 14. 8-15 mice per
1212 treatment group. B-D. Flow cytometry to confirm depletion of target cell types. Buffy
1213 coat was collected on day 22, 2 days after the final dose of depleting antibodies was
1214 given. One way ANOVA with Tukey's HSD for multiple comparisons. E. Initial
1215 experiments showed a small increase in unique barcode number in the lungs when mice
1216 were treated with CD8-depleting antibodies. F. Repeated experiments showed depletion
1217 of key cell types does not significantly change barcode number in the lungs at endpoint.
1218 Sup Figure 7.Gating strategy to confirm depletion of target cell types in 4T1 tumour
1219 bearing mice, treated with either anti-CD4, anti-CD8 or anti-asialo GM1.
1220
1221 Sup Figure 8. Analysis of changes in barcode proportions in the 4T1 cells with the 300
1222 000 barcode library replicates the 5000 barcode library. A. Number of unique barcodes
1223 identified in 4T1 primary tumours and lung metastases grown in NSG mice or Balb/c
1224 mice. GLM fit with Tukey's HSD for multiple comparisons. n= 5 mice per group. B.
1225 Number of unique barcodes identified in 4T1 primary tumours and lung metastases
1226 grown in Balb/c mice treated with isotype control antibodies or anti-PD1 + anti-
1227 CTLA4. A trend to decreasing unique barcode number is seen in anti-PD1 and anti-
1228 CTLA4 treated lungs, although this does not reach significance. GLM fit with Tukey's
1229 HSD for multiple comparisons. 5-6 mice/group.
1230 Sup Figure 9. Kinetics of 4T1 clonal cell lines in vitro and in vivo. A: Growth kinetics
1231 in vitro as measured by percentage confluence over time (hr). B: Tumour weight at
1232 resection generated from clonal cell lines transplanted into BALB/c mice. C: Lung
1233 weight at endstage of untreated subclone-tumour bearing BALB/c mice
1234

1235 Sup Figure 10. Genomic alterations do not appear to explain phenotypes of the immune
1236 evasive clones A. Genome copy number at the MHC-I locus on chromosome 17 of
1237 indicated clones. Chromosomal location is indicated on the x-axis and level of copy
1238 number alteration is indicated on the y-axis. B. Expression levels of the two genes that
1239 have a single copy number increase in IE1 and IE2 in indicated cell populations.
1240
1241 Sup Figure 11. Distinct populations of subclones with varying PD-L1 and MHC I
1242 expression exist in immunotherapy treated end-stage lungs. A: Gating strategy. B: PD-
1243 L1 and MHC I expression of RFP+ barcoded cancer cells isolated from endstage 4T1
1244 immunotherapy treated lungs. Each plot represents an individual mouse.
1245
1246 Sup Figure 12. Regulation of MHC expression by 5-aza and IFN-γ in the clonal cell
1247 lines. A. MHC-I protein expression measured by flow cytometry in indicated cell lines
1248 treated with indicated concentrations of 5-aza. Two-way ANOVA with Tukey's HSD
1249 for multiple corrections. B. MHC-I protein expression in indicated cell lines without or
1250 with IFN-γ treatment. One way ANOVA with Tukey's HSD for multiple corrections.
1251 Sup Figure 13. Common gene signature from immunotherapy resistant clones. A. Gene-
1252 set enrichment analysis of the overlapping genes between the IE1 and IE2 clones. Red
1253 indicates upregulated gene-sets with a significant FDR q-value when multiple testing
1254 is taken into account, orange indicates significant upregulated gene-sets with a nominal
1255 p value, and blue indicates downregulated gene-sets with a significant FDR q-value. B.
1256 Heatmap of unsupervised hierarchical clustering of the immunotherapy resistance
1257 signature genes in the TCGA breast cancer dataset, with tumours grouped based on top
1258 30%, middle 40%, and bottom 30% overall gene signature score. C. Correlation plot of
1259 immunotherapy resistance signature score with cytotoxic T lymphocyte (CTL) levels

1260 in the METABRIC dataset. D. Correlation plot of immunotherapy resistance signature
1261 score with CTL levels in the TCGA dataset.
1262
1263 **Supplementary Table Legends:**
1264 Sup Table 1: Barcode overlap between two cell pellets from independent barcoding
1265 experiments using the 4T1 barcoded cell line. At a barcode cutoff of 5, more than 95%
1266 of barcodes detected were the same, and at a cutoff of 10, more than 97% of detected
1267 barcodes were identical.
1268
1269 Sup Table 2: DNA barcode insertion sites of IE1 and IE2. By searching the whole
1270 genome sequencing (WGS) data and identifying read pairs where only one read mapped
1271 to the barcode sequence, the matching mate was blasted against mm10 to identify the
1272 barcode plasmid insertion site. Both barcodes were found to reside in introns of cancer-
1273 related genes but there was no difference in expression of these genes detected.
1274
1275 Sup Table 3: Copy number variation (CNV) locations found within all subclones. By
1276 analysing whole genome sequencing data in R using the cn.mops pacakges, copy
1277 number variations could be determined. No major copy number aberrations were
1278 detected across the clones, although a single copy number gain was detected in IE1, IE2
1279 and NT2.
1280
1281 Sup Table 4. List of genes differentially expressed in IE1 compared to parental 4T1
1282 bulk. Differentially expressed genes (DEGs) were generated by analyzing bulk RNA
1283 sequencing data using R and the EdgeR package. DEGs were filtered for significance
1284 based on a FDR <0.05.

1285

1286 Sup Table 5. List of genes differentially expressed in IE2 compared to parental 4T1
1287 bulk. Differentially expressed genes (DEGs) were generated by analyzing bulk RNA
1288 sequencing data using R and the EdgeR package. DEGs were filtered for significance
1289 based on a FDR <0.05. A greater number of differentially expressed genes were
1290 detected in IE2 than IE1.

1291

1292 Sup Table 6. All significantly enriched gene sets found in IE1. Differentially expressed
1293 genes generated from comparing IE1 to bulk were preranked by fold change before
1294 searching for gene set enrichment using the Molecular Signature Database (MSigDB)
1295 across all available collections. Gene sets were filtered for significance based on a
1296 FDR<0.05.

1297

1298 Sup Table 7. All significantly enriched gene sets found in IE2. Differentially expressed
1299 genes generated from comparing IE2 to bulk were preranked by fold change. Gene set
1300 enrichment was carried out using Molecular Signature Database (MSigDB) across all
1301 available collections. Gene sets were filtered for significance based on a FDR<0.05

1302

1303 Sup Table 8: List of common differentially expressed genes found in IE1 and IE2.
1304 Commonly differentially expressed genes (DEGs) were determined by overlapping
1305 significant DEGs in IE1 and IE2 by gene name. An average fold change was calculated
1306 across the two samples.

1307

1308 Sup Table 9: Top significantly enriched gene sets found from common differentially
1309 expressed genes in IE1 and IE2. The fold change of the common differentially

1310 expressed genes in IE1 and IE2 were averaged together to generate an average fold
1311 change across both IE1 and IE2. The gene list was then preranked before searching for
1312 gene set enrichment using Molecular Signature Database (MSigDB) across the C2 All
1313 collection. Gene sets were filtered for significance based on a FDR<0.05. The majority
1314 of significant gene sets that were returned were negatively enriched in IE1 and IE2.
1315
1316
1317

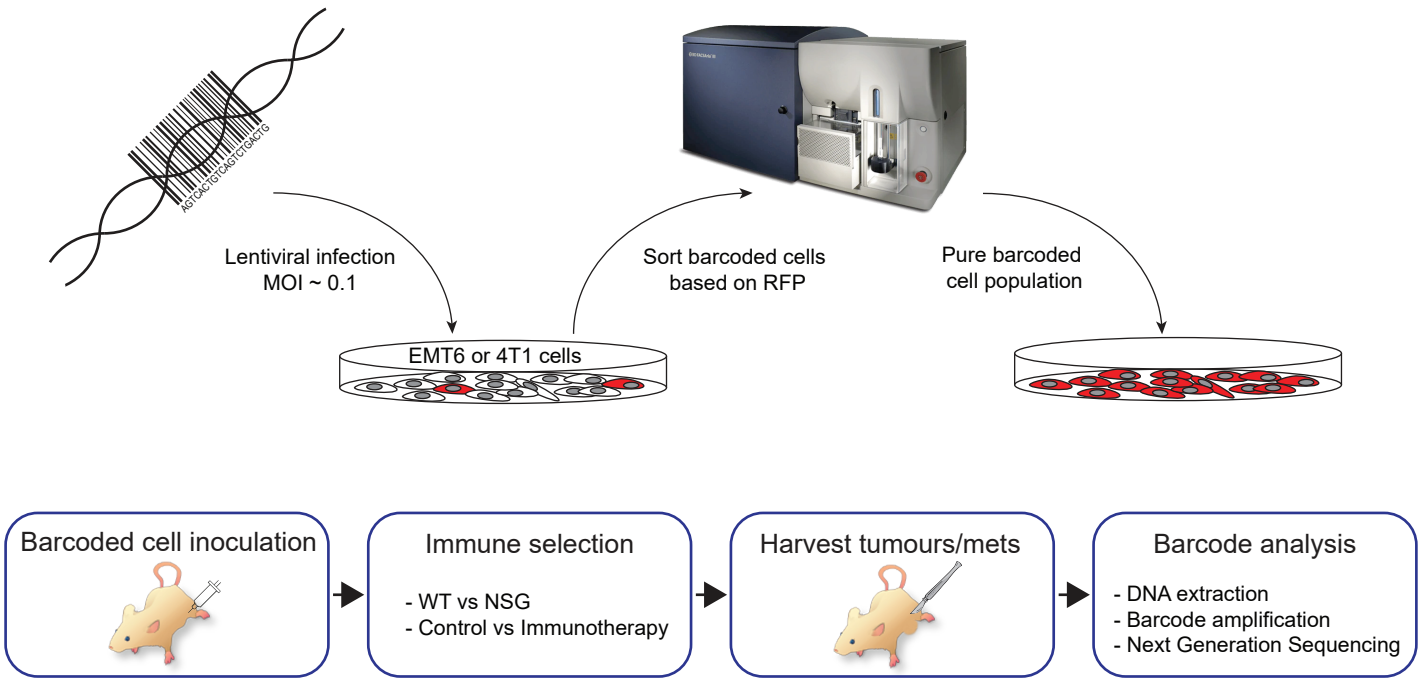


Figure 1: Experimental workflow schematic. Barcode library is introduced into mammary carcinoma cell lines in vitro at a low multiplicity of infection (MOI). Cells are sorted based on RFP expression to select for those having incorporated a barcode. Barcoded cells are then transplanted into the mammary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

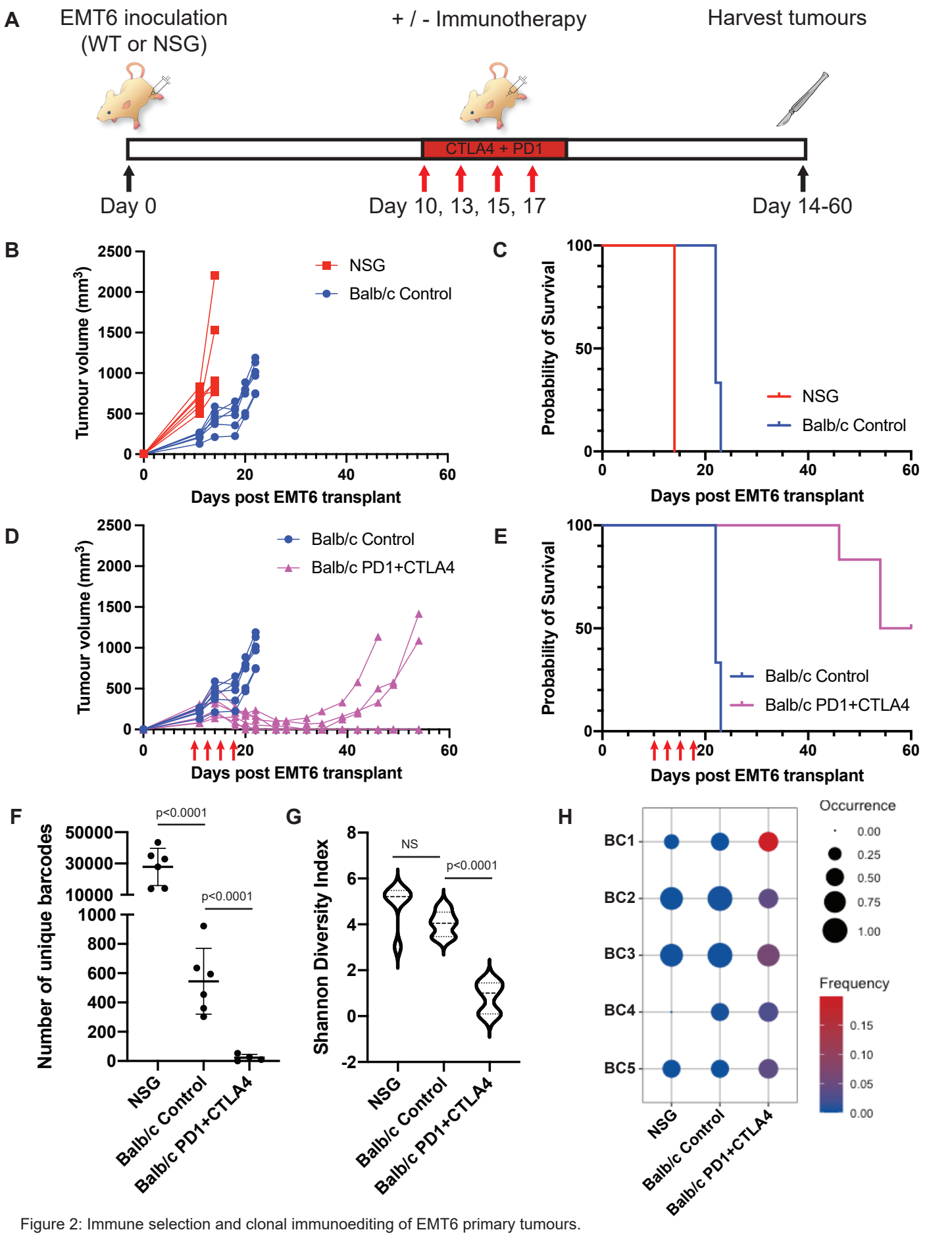


Figure 2: Immune selection and clonal immunoediting of EMT6 primary tumours.
A. Outline of experimental design. **B**. EMT6 primary tumour growth in wild-type Balb/c mice and in NSG mice plotted as tumour volume. n=6 mice per group. **C**. Kaplan-Meier survival analysis comparing Balb/c and NSG mice bearing EMT6 tumour (Mantel-Cox p=0.009). **D**. EMT6 primary tumour growth plotted as tumour volume in wild-type Balb/c mice with or without immunotherapy (anti-PD1+anti-CTLA4) on days 10, 12, 14, and 17 as indicated with red arrows. n=6 mice per group. **E**. Kaplan-Meier survival analysis of Balb/c mice bearing EMT6 tumour treated with immunotherapy or isotype control (Mantel-Cox p=0.0006). **F**. Number of unique barcodes identified in EMT6 primary tumours grown in NSG mice and in Balb/c mice treated with isotype control antibodies or anti-PD1 + anti-CTLA4. GLM fit with Tukeys HSD for multiple comparisons. **G**. Shannon diversity index analysis of EMT6 primary tumours grown in NSG mice and in Balb/c mice treated with isotype control antibodies or anti-PD1 + anti-CTLA4. One way ANOVA with Tukeys HSD. **H**. Dot plot of a subset of barcodes with enrichment following immunotherapy treatment.

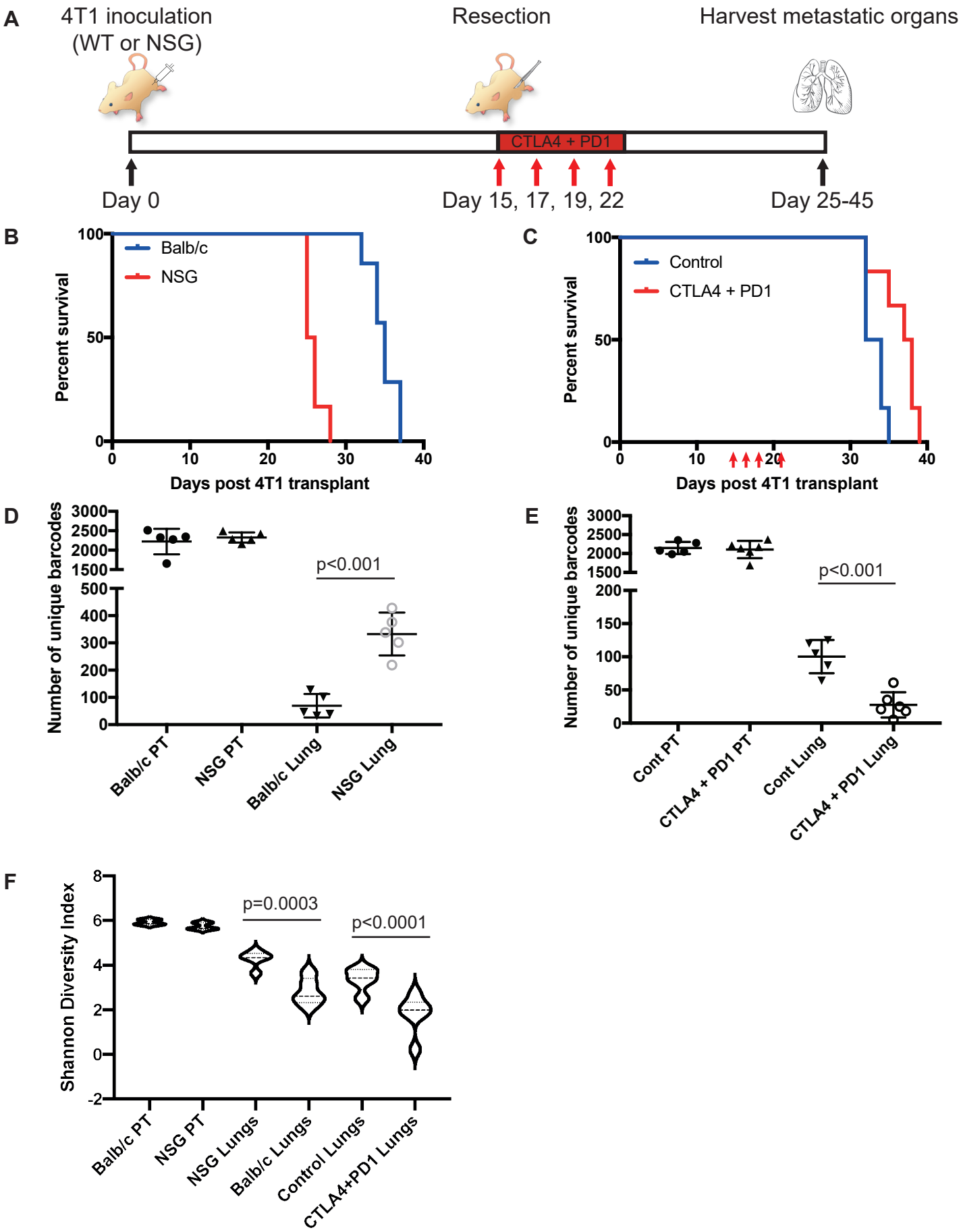


Figure 3: Immune selection and clonal immunoediting in the metastatic setting utilising the 4T1 model.
A. Outline of experimental design. **B**. Kaplan-Meier survival analysis comparing Balb/c mice and NSG mice bearing 4T1 tumour; n=5 mice/group (p = 0.0002, Mantel-Cox Log-rank test). **C**. Kaplan-Meier survival analysis of Balb/c mice bearing 4T1 tumour treated with immunotherapy or isotype control on days 15, 17, 19, and 21 as 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 (PT) and lung metastases grown in NSG mice or Balb/c mice. GLM fit with Tukey's HSD for multiple comparisons. **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. GLM fit with Tukey's HSD for multiple comparisons. **F**. Shannon diversity index analysis in primary tumours from NSG mice or Balb/c mice and in lung metastases from NSG mice, Balb/c mice, Balb/c mice treated with isotype control, and Balb/c mice treated with anti-PD1 + anti-CTLA4. One way ANOVA with Tukeys HSD for multiple comparisons.

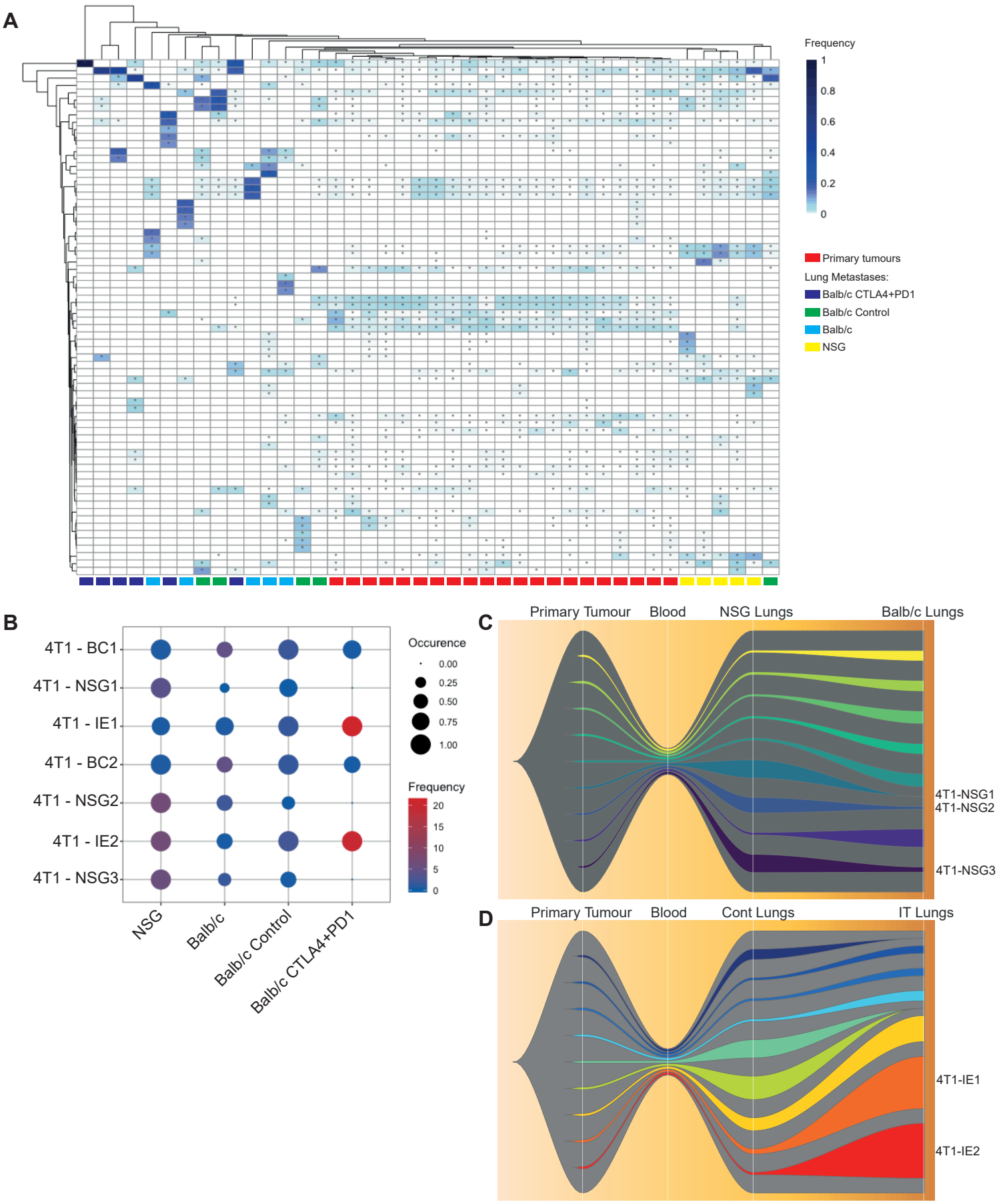


Figure 4: Analysis of specific barcodes enriched or depleted by the immune system and immunotherapy in the 4T1 model.
A. Unsupervised hierarchical clustering heatmap of barcodes with an abundance of above 5% in at least one 4T1 primary tumour or lung metastasis sample; * indicate barcodes detected at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

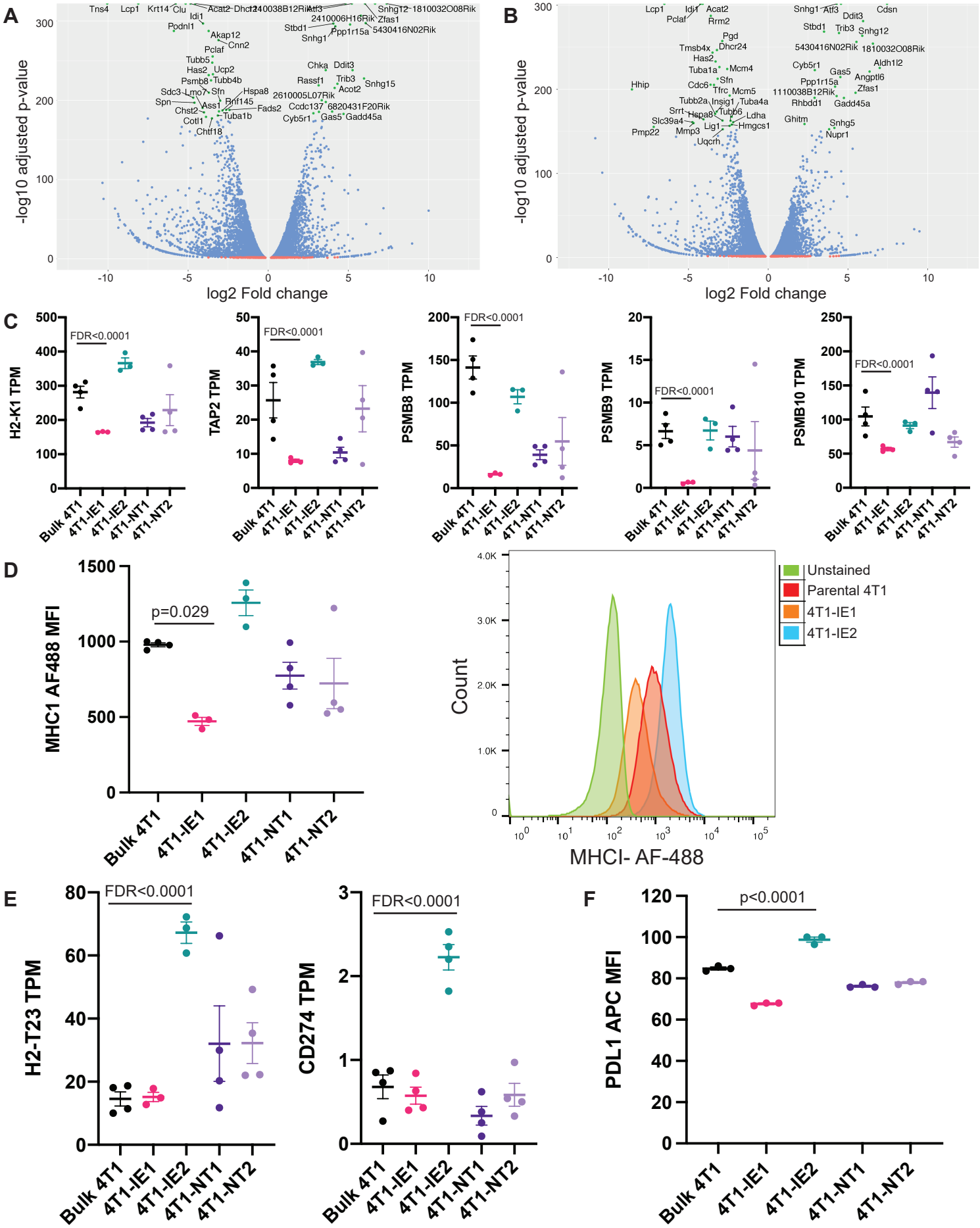


Figure 5: Gene expression analysis of immunotherapy resistant clones.
A. Volcano plot showing differentially expressed genes between parental 4T1 cells and the immunotherapy enriched 1 (IE1) clone. **B**. Volcano plot showing differentially expressed genes between parental 4T1 cells and the immunotherapy enriched 2 (IE2) clone. **C**. Expression of indicated MHC related genes, measured as transcripts per million (TPM), in the parental 4T1 population and in indicated cell clones; FDR calculated using EdgeR with Benjamini-Hochberg multiple testing correction. **D**. MHCI protein expression as quantified by flow cytometry in the indicated clones and the parental 4T1 population measured as mean fluorescence intensity (MFI) (left) and representative histogram (right); one way ANOVA with Tukey HSD test for multiple comparisonst. **E**. Expression of immune related genes up-regulated in clone IE2, measured as transcripts per million (TPM), in the parental 4T1 population and in indicated cell clones. FDR calculated by EdgeR with Benhamini-Hochberg multiple testing correction **F**. PD-L1 protein expression as determined by flow cytometry in indicated clones and the parental 4T1 population measured as mean fluorescence intensity (MFI), representative plot of three independent experiments; one way ANOVA with Tukey HSD for multiple comparisons.

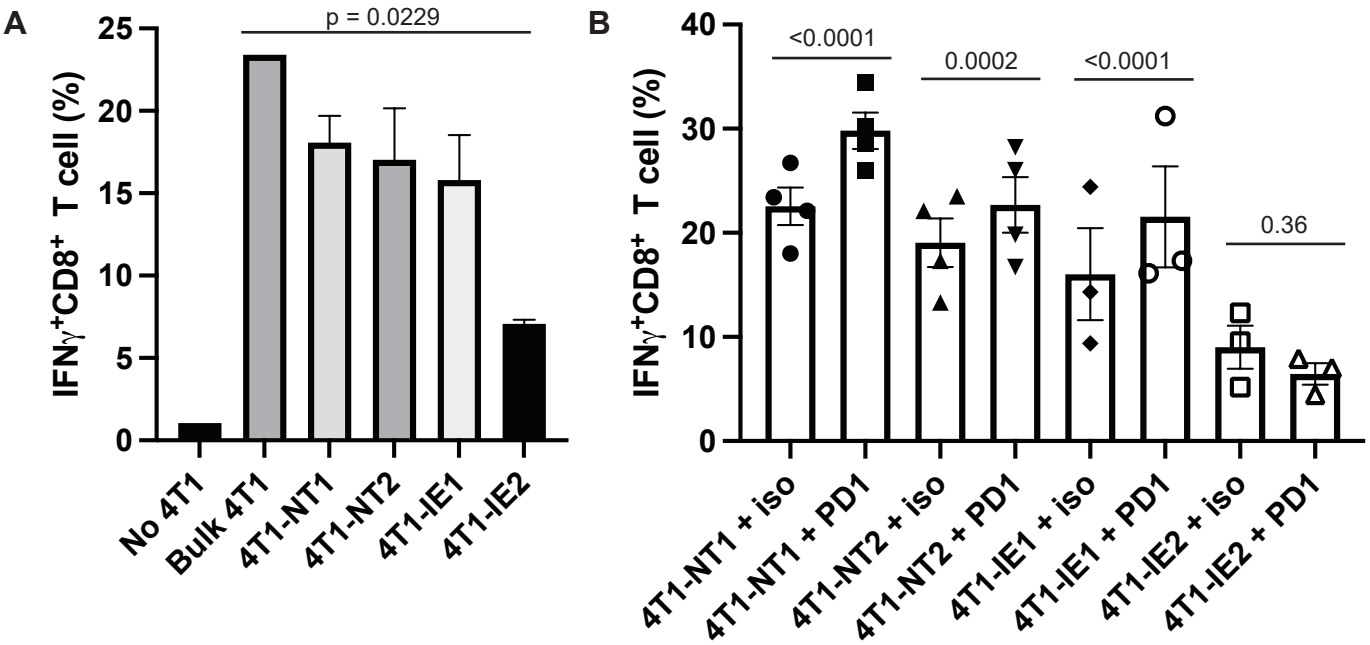


Figure 6: The 4T1-IE2 clone can directly suppress cytotoxic T-cell activation and prevent further T-cell activation by anti-PD1 immunotherapy. **A.** Measurent of intracellular IFNγ production by 4T1 specific CD8+ T-cells following co-culture with no 4T1 cells (No 4T1), bulk 4T1 cells (Bulk 4T1) and 4T1 clonal populations (4T1-N1; 4T1-NT2; 4T1-IE1; 4T1-IE2). For clonal populations n=3-4; one-way ANOVA with Dunnett's post-hoc test. **B.** Measurent of intracellular IFNγ production by 4T1 specific CD8+ T-cells following co-culture with clonal 4T1 populations with or without anti-PD1 (PD1) or an isotype control (iso). n=3-4; FWER calculated using flow cytometry proportions fitted to a GLM with Bonferroni post-hoc correction.

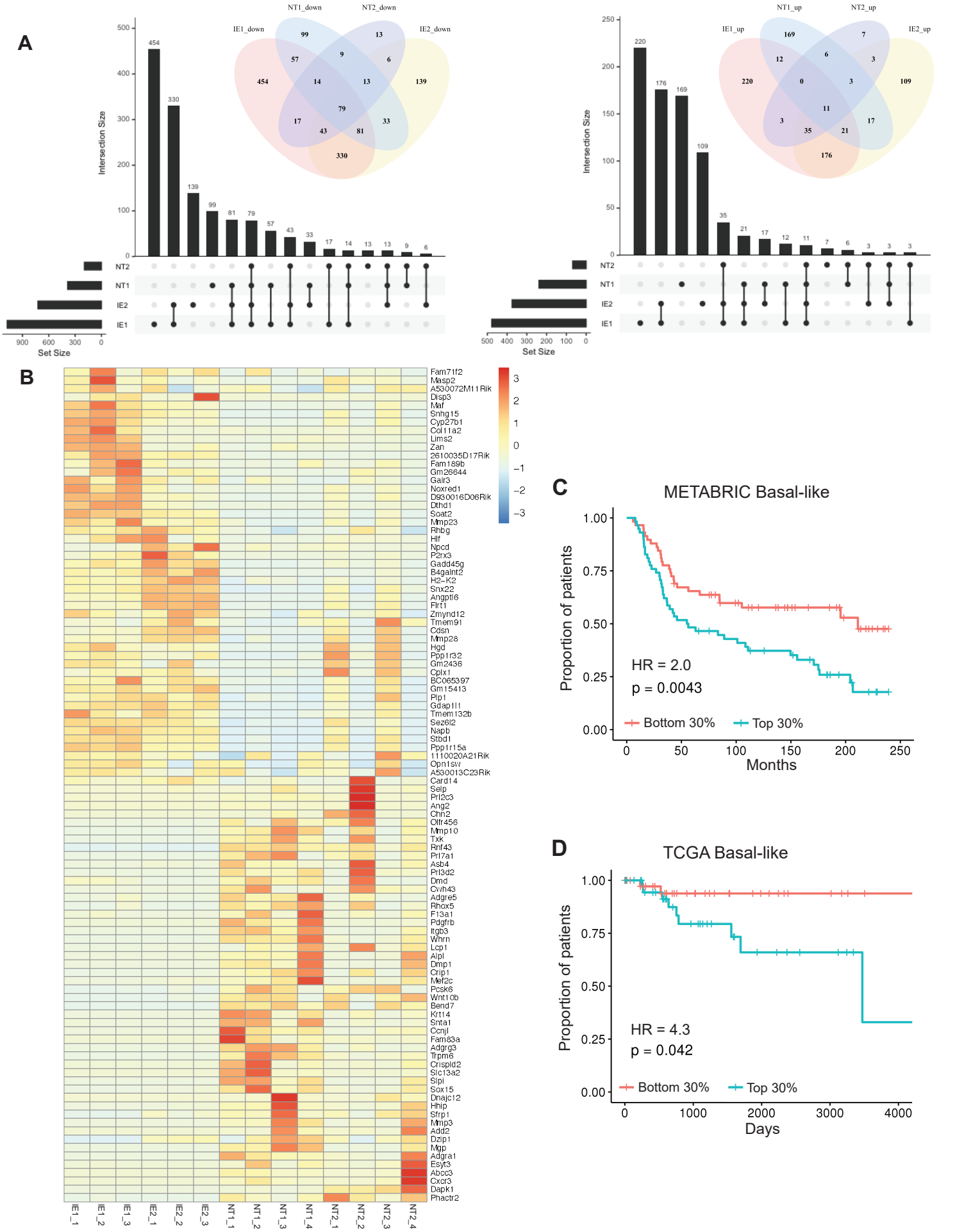
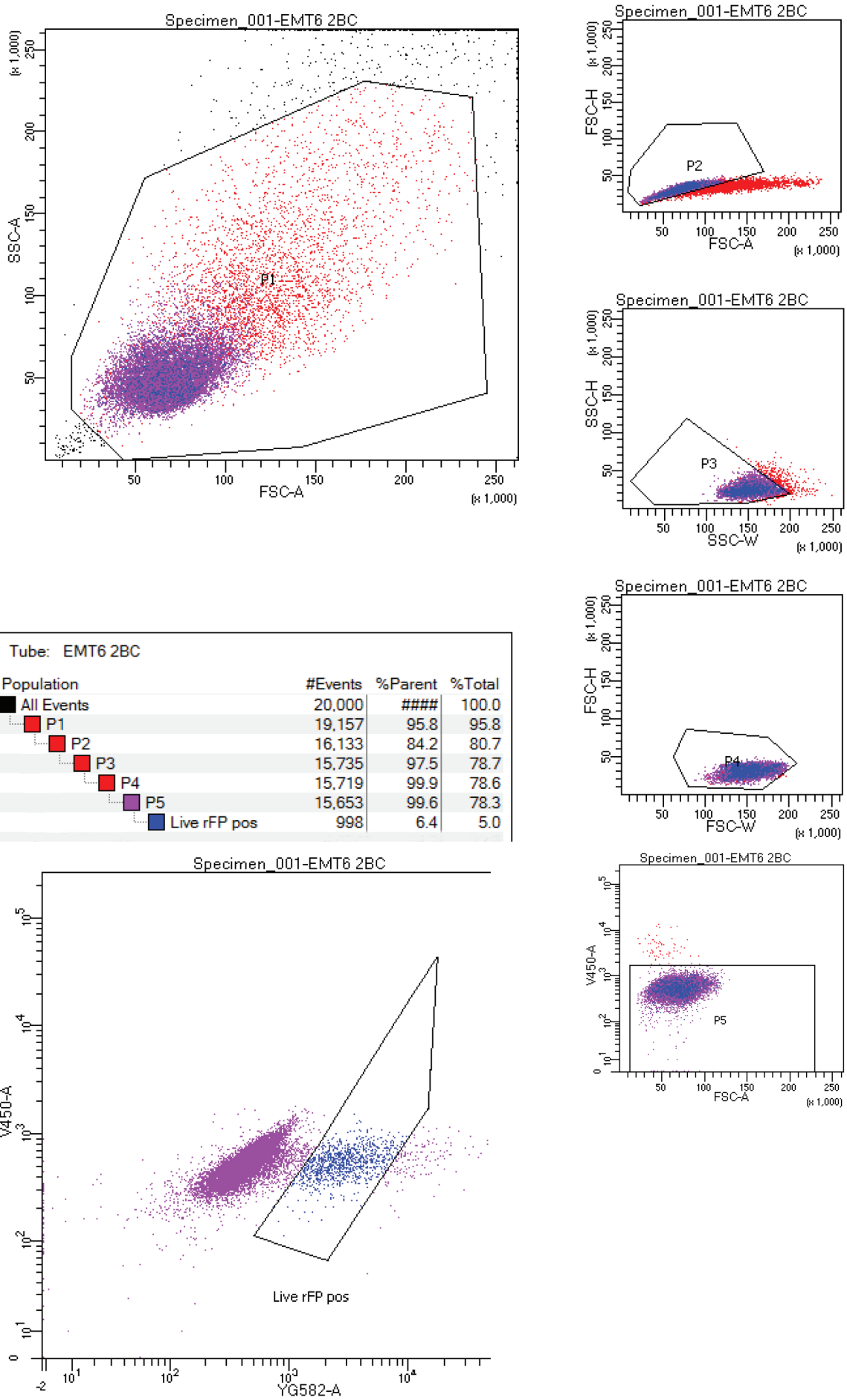


Figure 7: Overlapping gene signatures of the immunotherapy resistant clones show prognostic significance in basal-like breast cancer patients. **A**. Upset plots and Venn diagrams showing the overlap in significantly up-regulated (right) and down-regulated (left) genes between the two immunotherapy enriched clones (IE1 and IE2) and two control clones (NT1 and NT2). **B**. Heatmap of the top 50 overlapping up regulated and down regulated genes between IE1 and IE2 clones across all clonal cell lines. **C**. Kaplan-Meier analysis of overall survival of basal-like breast cancer patients in the METABRIC cohort based on whose breast cancer having the top 30% or the bottom 30% expression of the overlapping upregulated 25 gene signature. **D**. Kaplan-Meier analysis of overall survival of basal-like breast cancer patients in the TCGA cohort based on whose tumour having the top 30% or the bottom 30% expression of the overlapping upregulated 25 gene signature. In C and D, the Cox proportional hazards model was used to compute Hazard Ratios. Significance between stratification groups were computed using log-rank test statistics.

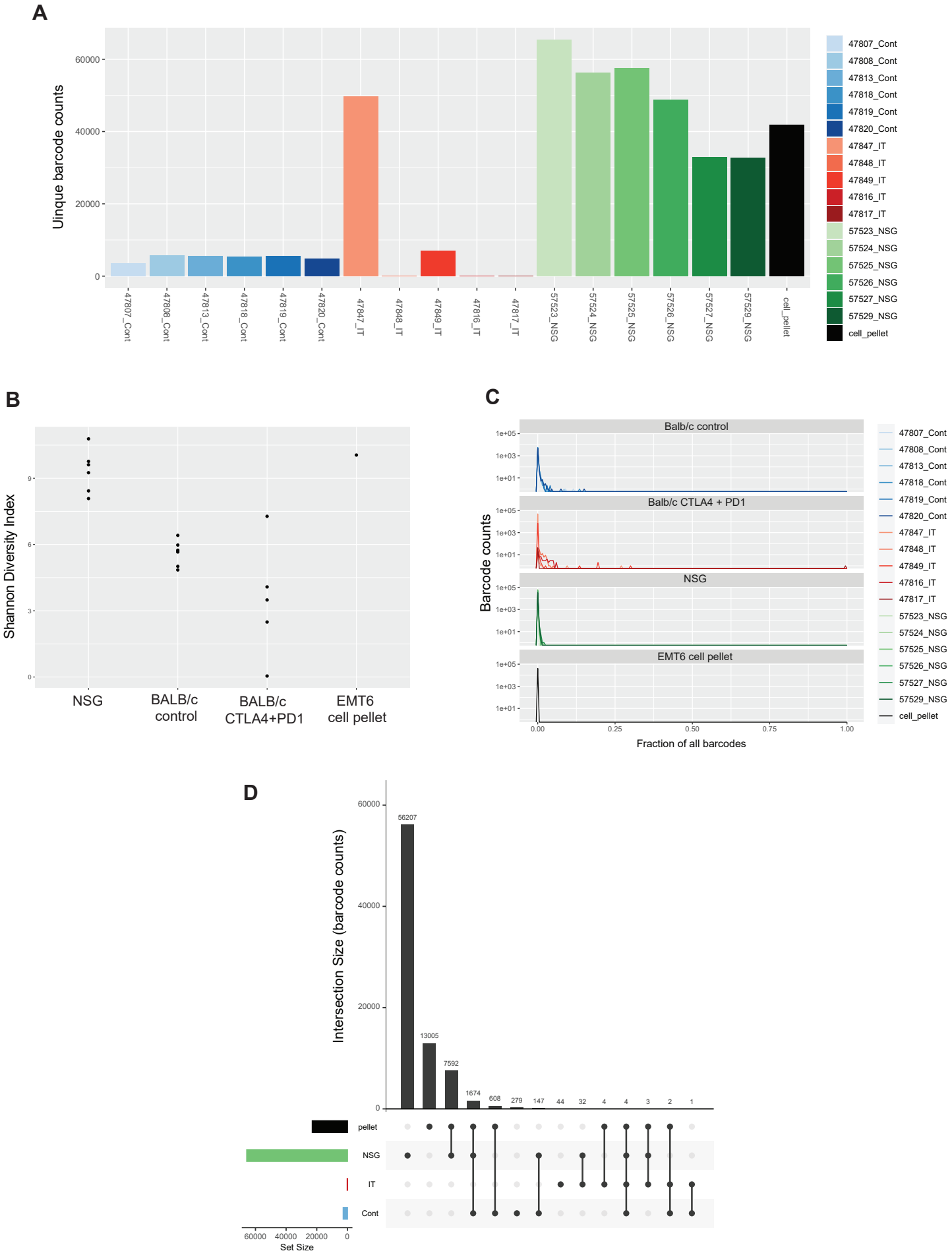
BD FACSDiva 8.0.1



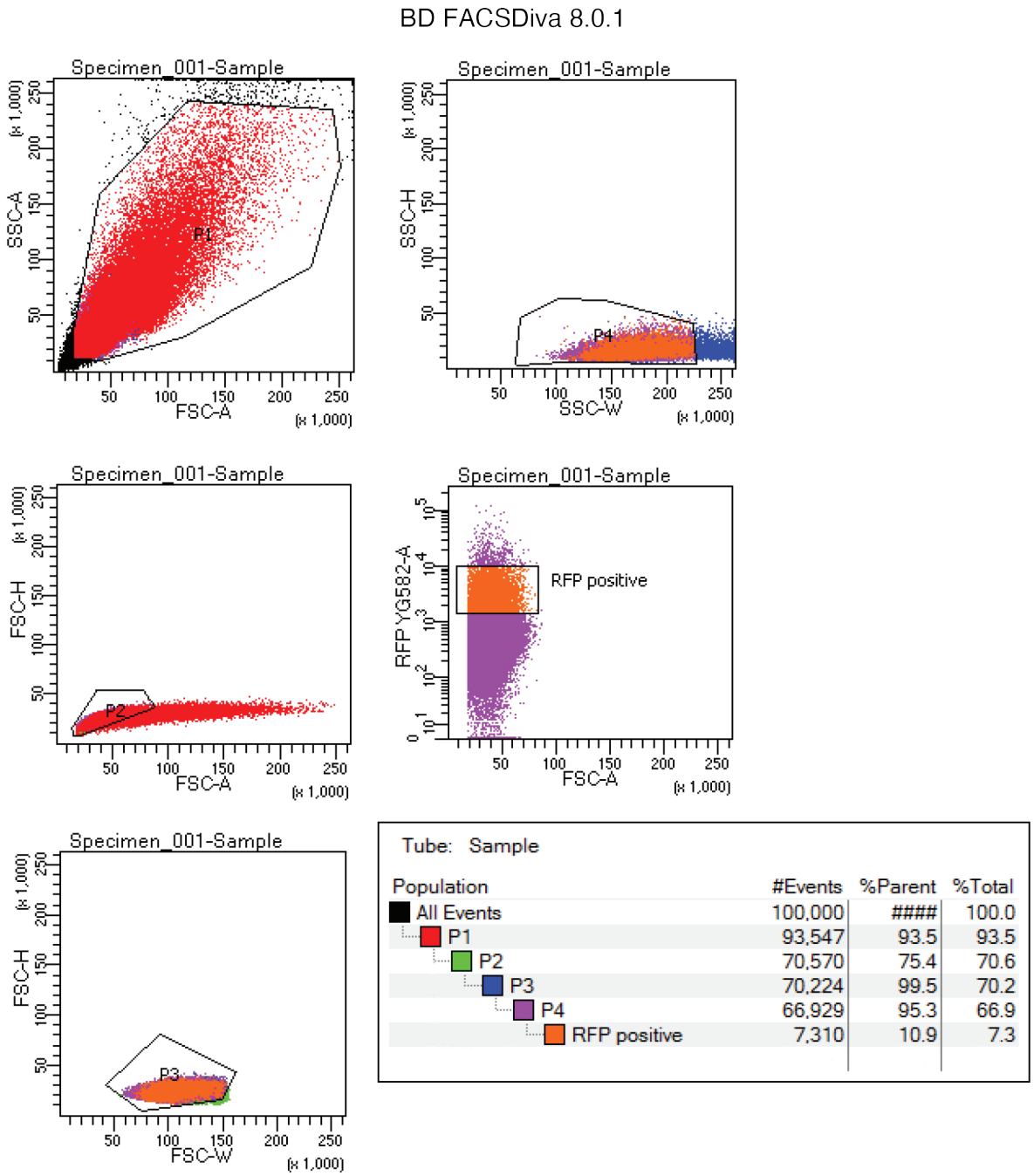
Tube: EMT6 2BC

Population	#Events	%Parent	%Total
All Events	20,000	####	100.0
P1	19,157	95.8	95.8
P2	16,133	84.2	80.7
P3	15,735	97.5	78.7
P4	15,719	99.9	78.6
P5	15,653	99.6	78.3
Live rFP pos	998	6.4	5.0

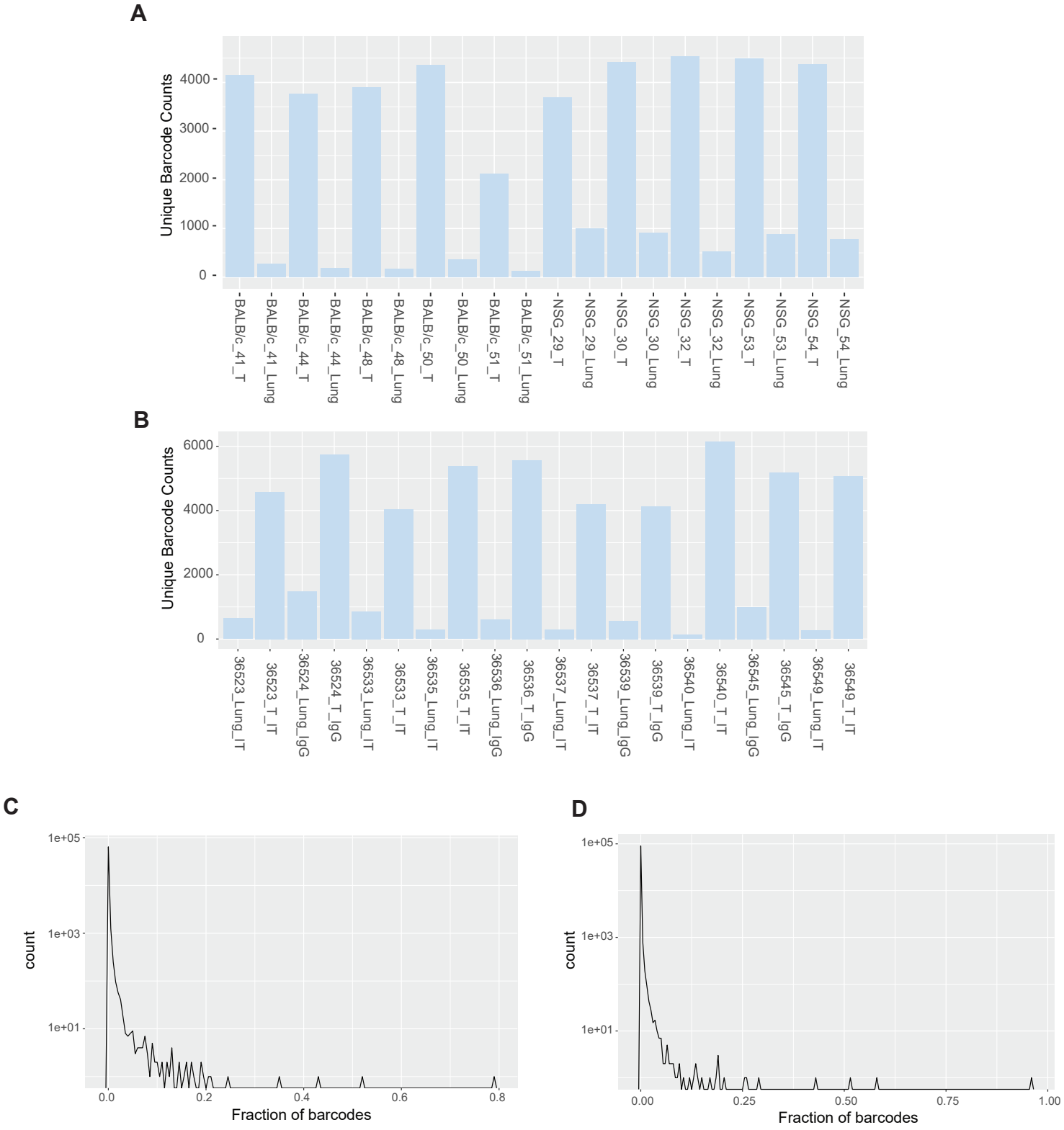
Sup Figure 1. Gating strategy showing isolation and selection of barcode transfected Red Fluorescent Protein (RFP) positive EMT6 cells.



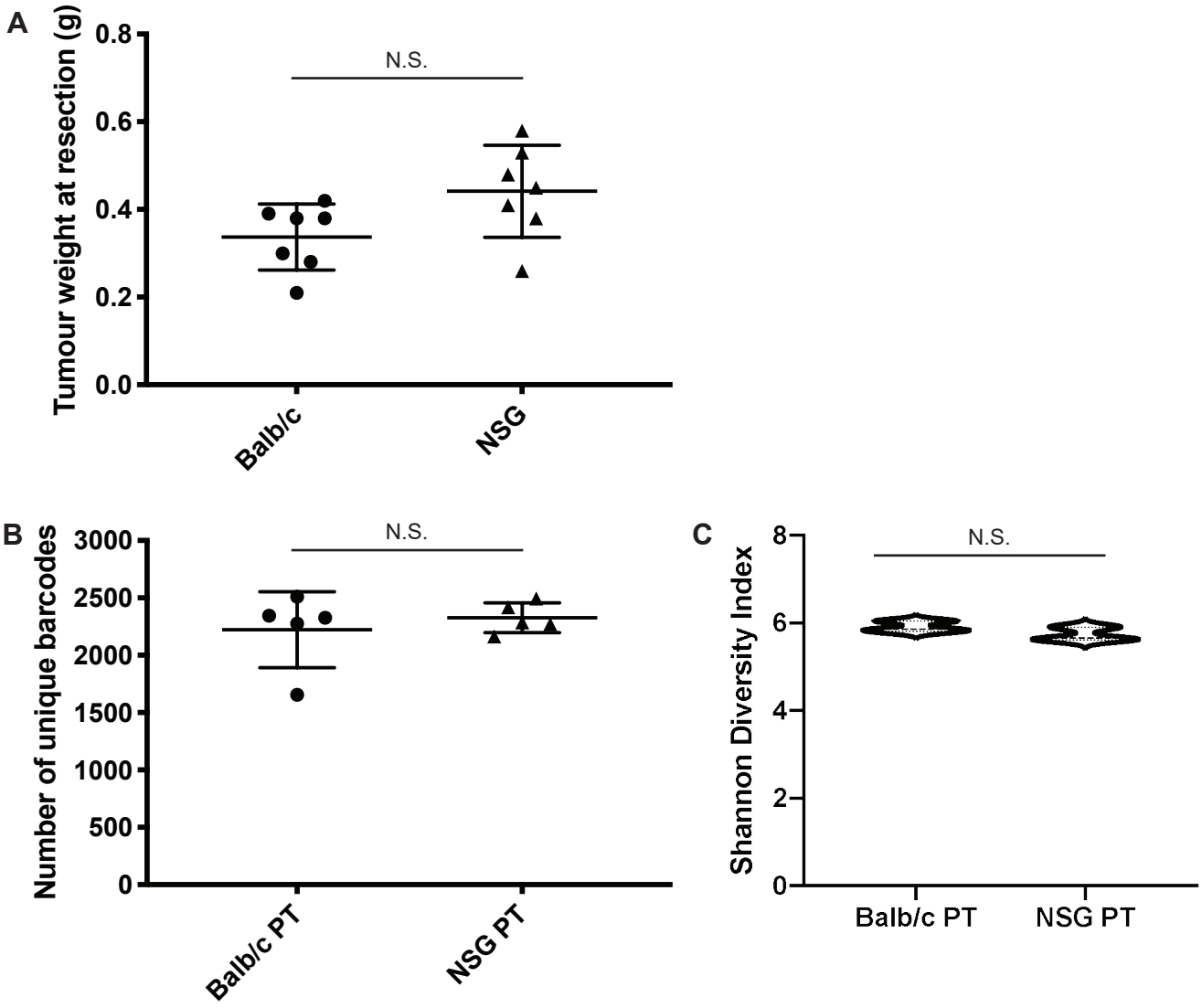
Sup Figure 2. Unique barcode counts and barcode distributions across EMT6 primary tumours and matching cell pellets. **A.** Unique barcode counts from EMT6 tumour bearing mice. **B.** Shannon diversity index. **C.** Distribution plots of barcode counts from individual mice, grouped by strain and treatment group. **D.** Upset plot showing unique barcode counts and overlap between the EMT6 cell pellet and tumour bearing NSG mice or immunocompetent immunotherapy treated (IT) or control treated (Cont) mice.



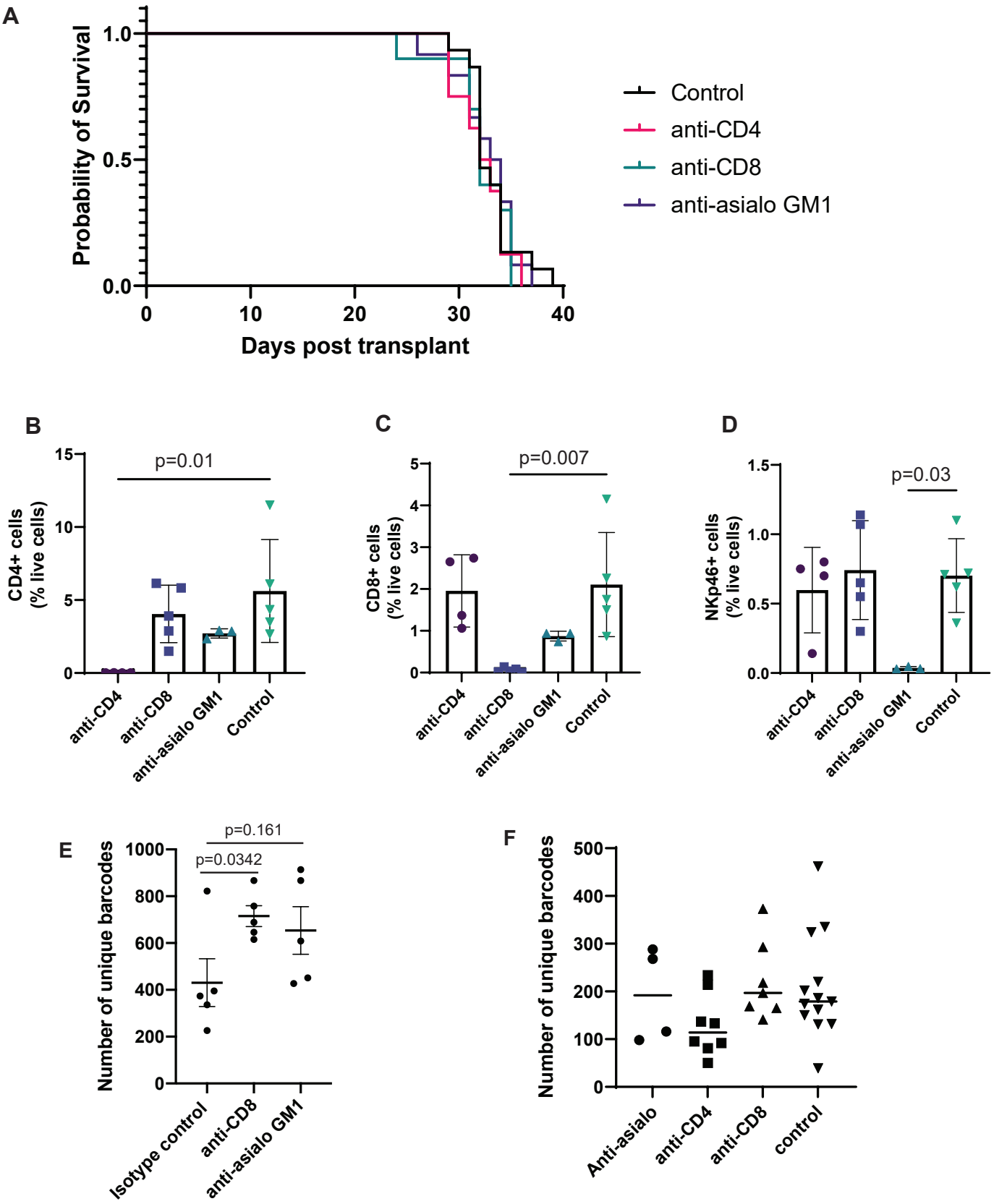
Sup Figure 3. Gating strategy showing isolation and selection of barcode transfected Red Fluorescent Protein (RFP) positive 4T1 cells.



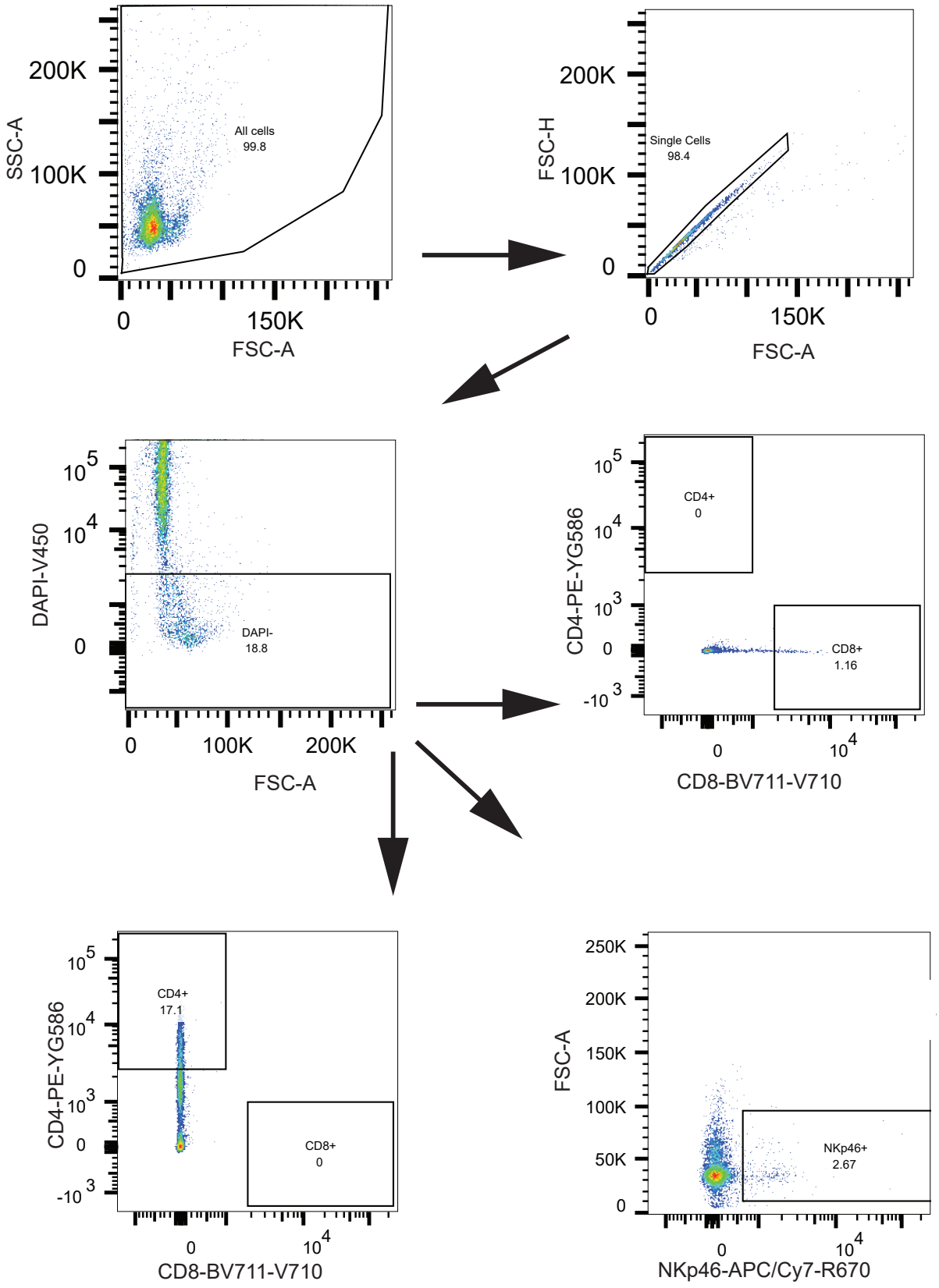
Sup Figure 4. Unique barcode counts and barcode distributions across 4T1 primary tumours and matched lungs in NSG and BALB/c mice. **A.** Unique barcode counts detected in matched primary tumour (T) or lungs from 4T1 tumour bearing NSG or BALB/c mice. **B.** Unique barcode counts detected in matched primary tumour (T) and lungs from 4T1 tumour bearing mice treated with either combined immunotherapy (IT) or control antibodies (IgG) **C.** Distribution plot of barcode counts from all (NSG and BALB/c) mice. **D.** Distribution plot of barcode counts from BALB/c mice treated with combination immunotherapy or control antibodies.



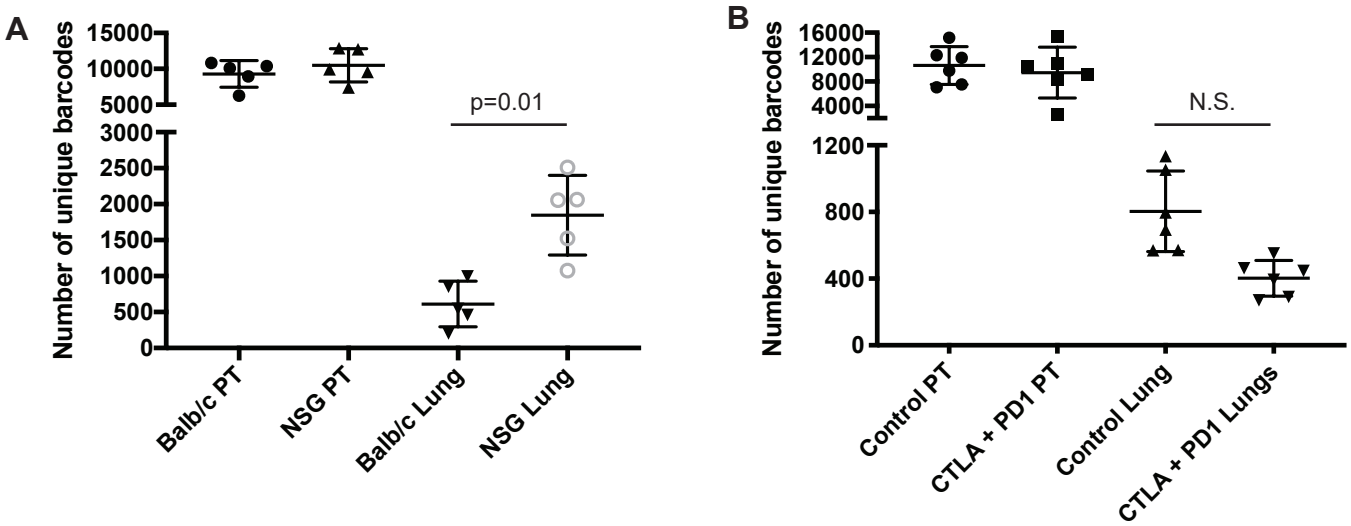
Sup Figure 5. 4T1 primary tumour growth and barcode diversity in primary tumours. **A.** Primary tumour mass. **B.** Total number of unique barcodes identified in primary tumours of Balb/c or NSG mice. **C.** Shannon diversity analysis comparing primary tumours from Balb/c and NSG mice.



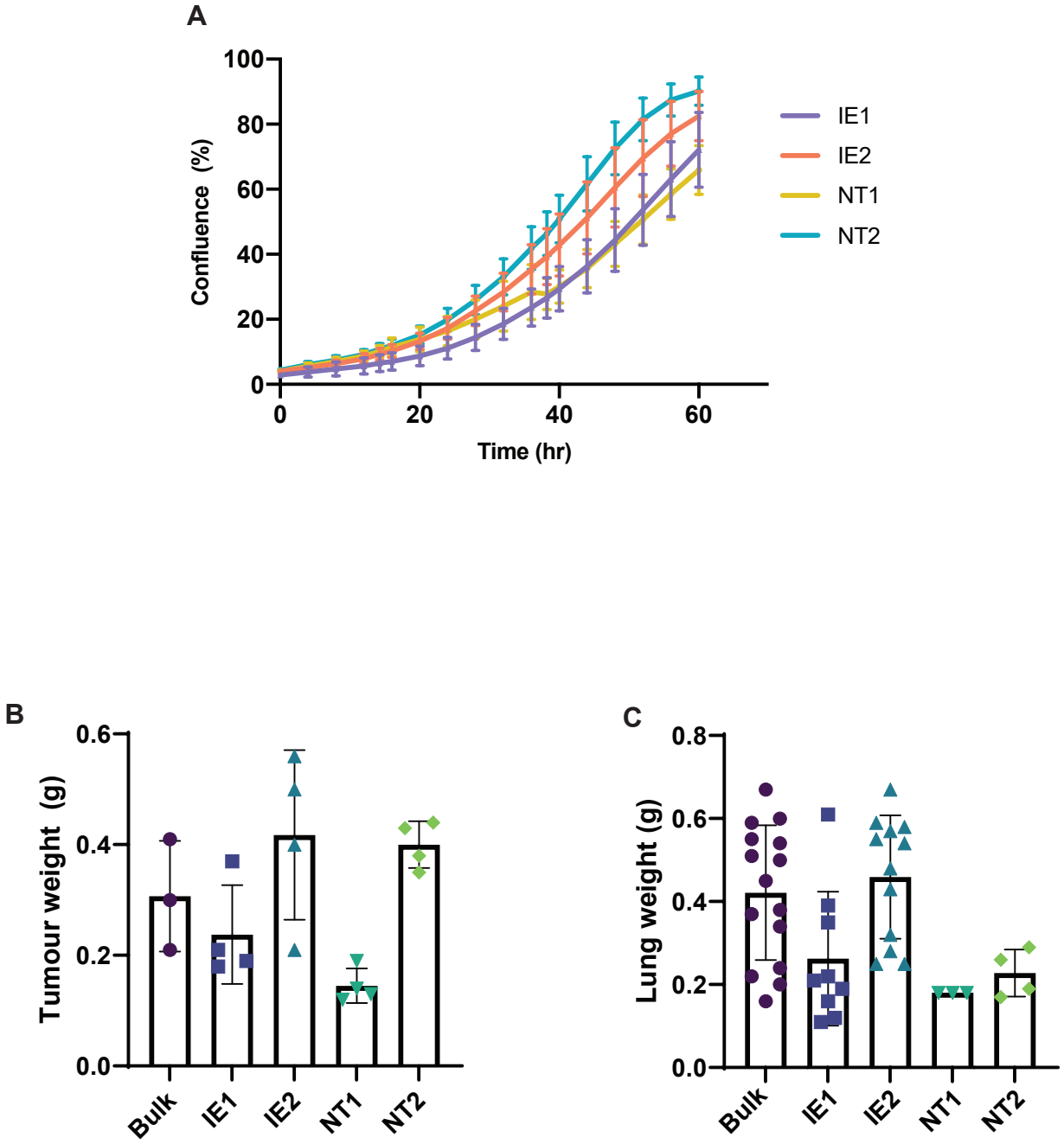
Sup Figure 6. Depletion of CD8+ T cells with anti-CD8, CD4+ T cells with anti-CD4 or NK cells with anti-asialo GM1 does not significantly affect survival or barcode diveristy of mice bearing 4T1 tumours. **A.** Kaplan-Meier survival analysis of Balb/c mice transplanted with 4T1 cells, primary tumour was resected on day 15. Depletion of target cell types was initiated 1 day prior to resection, on day 14. 8-15 mice per treatment group. **B-D.** Flow cytometry to confirm depletion of target cell types. Buffy coat was collected on day 22, 2 days after the final dose of depleting antibodies was given. One way ANOVA with Tukey's HSD for multiple comparisons. **E.** Initial experiments showed a small increase in unique barcode number in the lungs when mice were treated with CD8-depleting antibodies. **F.** Repeated experiments showed depletion of key cell types does not significantly change barcode number in the lungs at endpoint.



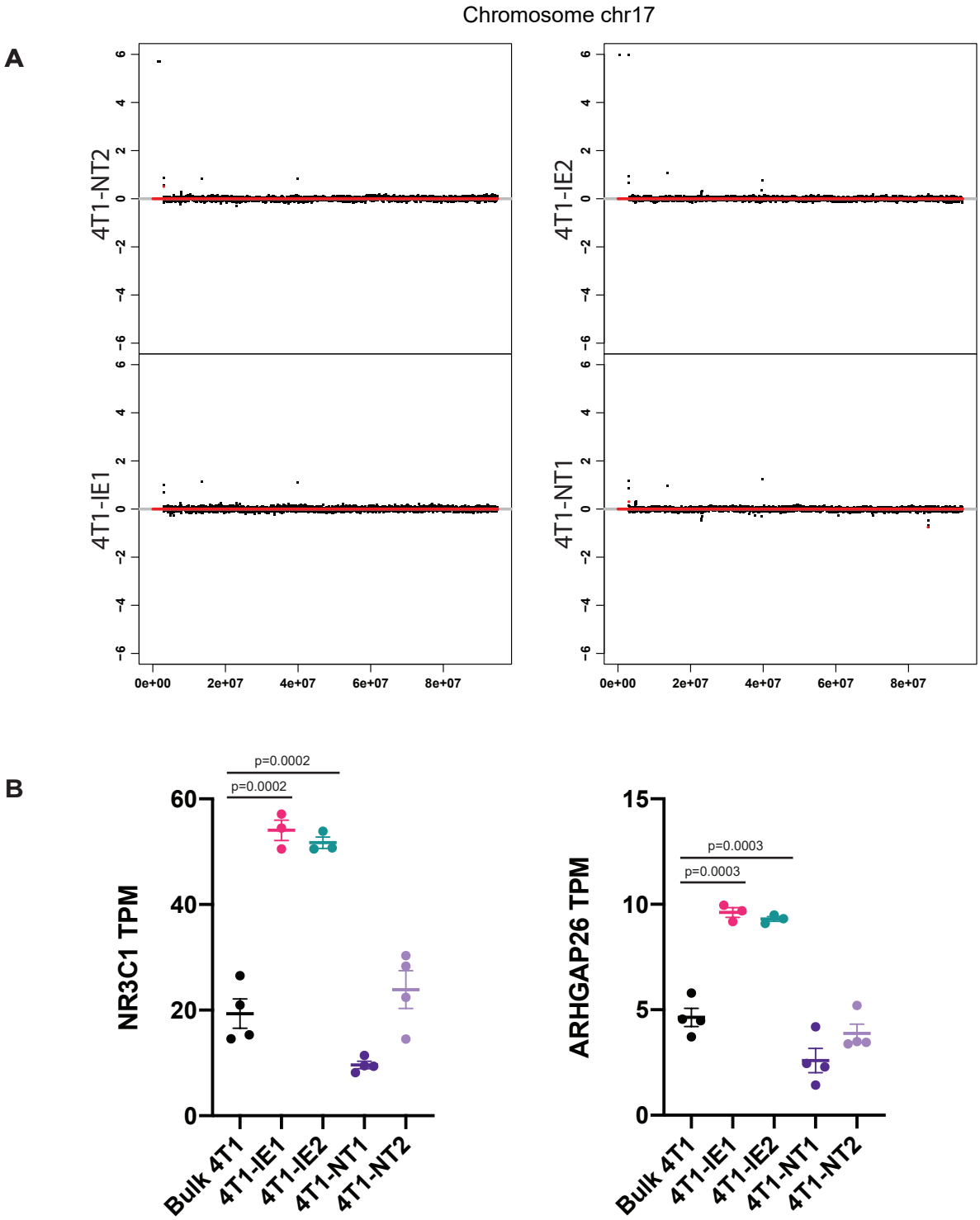
Sup Figure 7: Gating strategy to confirm depletion of target cell types in 4T1 tumour bearing mice, treated with either anti-CD4, anti-CD8 or anti-asialo GM1.



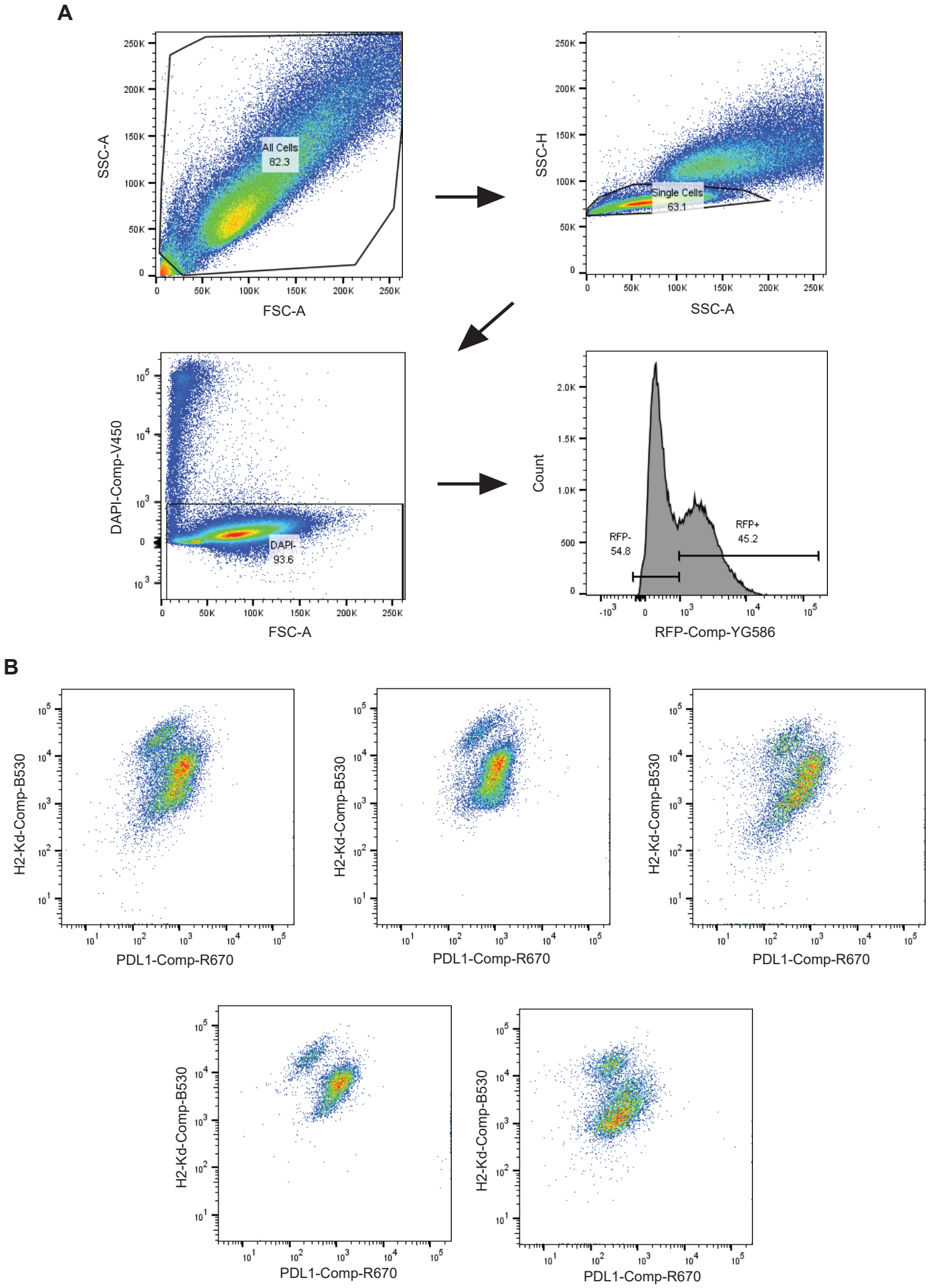
Sup Figure 8. Analysis of changes in barcode proportions in the 4T1 cells with the 300 000 barcode library replicates the 5000 barcode library. **A.** Number of unique barcodes identified in 4T1 primary tumours and lung metastases grown in NSG mice or Balb/c mice. GLM fit with Tukey's HSD for multiple comparisons. n= 5 mice per 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. A trend to decreasing unique barcode number is seen in anti-PD1 and anti-CTLA4 treated lungs, although this does not reach significance. GLM fit with Tukey's HSD for multiple comparisons. 5-6 mice/group.



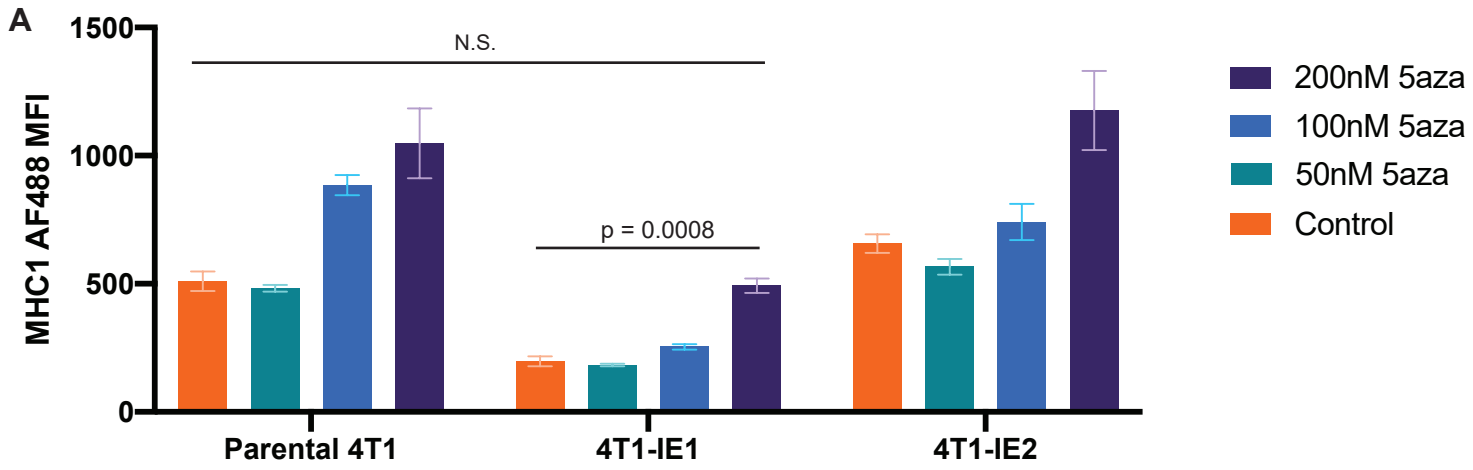
Sup Figure 9. Kinetics of 4T1 clonal cell lines in vitro and in vivo. **A:** Growth kinetics *in vitro* as measured by percentage confluence over time (hr). B: Tumour weight at resection generated from clonal cell lines transplanted into BALB/c mice. **C:** Lung weight at end-stage of untreated subclone-tumour bearing BALB/c mice

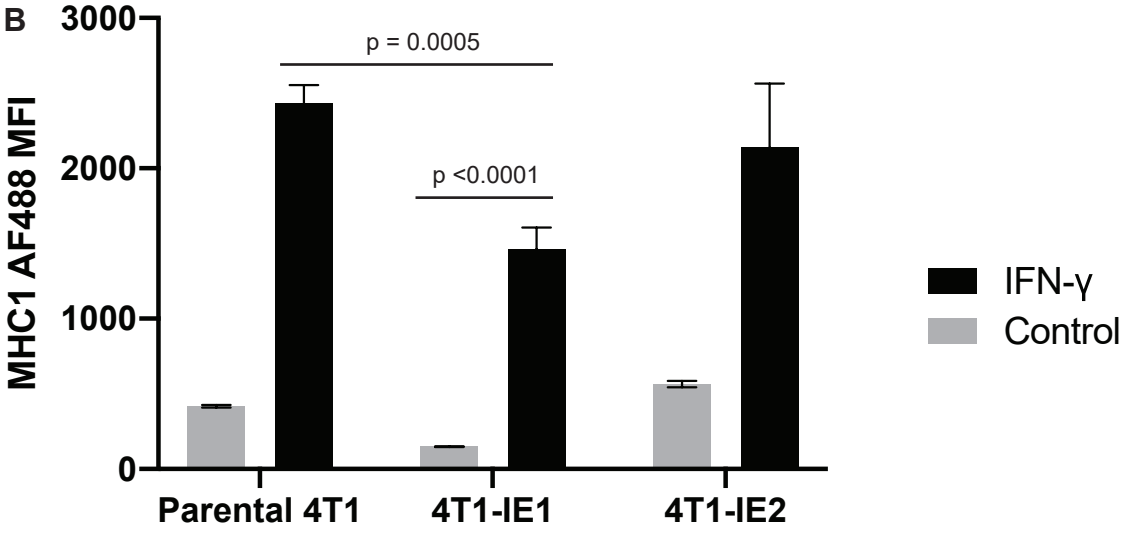


Sup Figure 10. Genomic alterations do not appear to explain phenotypes of the immune evasive clones. **A.** Genome copy number at the MHC-I locus on chromosome 17 of indicated clones. Chromosomal location is indicated on the x-axis and level of copy number alteration is indicated on the y-axis. **B.** Expression levels of the two genes that have a single copy number increase in IE1 and IE2 in indicated cell populations.

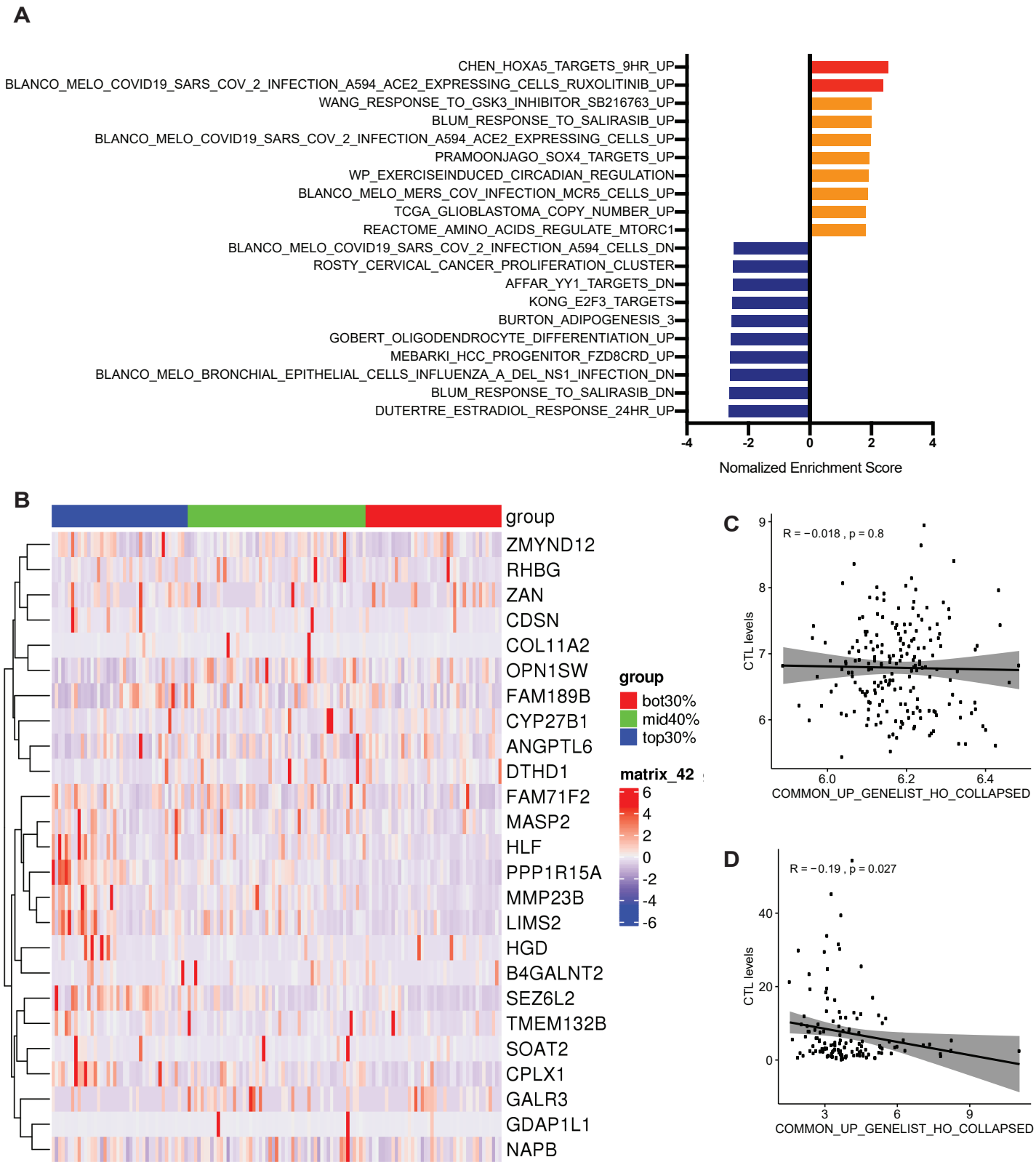


Sup Figure 11: Distinct populations of subclones with varying PD-L1 and MHC I expression exist in immunotherapy treated end-stage lungs. **A:** Gating strategy. **B:** PD-L1 and MHC I expression of RFP+ barcoded cancer cells isolated from endstage 4T1 immunotherapy treated lungs. Each plot represents an individual mouse.





Sup Figure 12. Regulation of MHC expression by 5-aza and IFN-γ in the clonal cell lines. **A.** MHC-I protein expression measured by flow cytometry in indicated cell lines treated with indicated concentrations of 5-aza. Two-way ANOVA with Tukey's HSD for multiple corrections. **B.** MHC-I protein expression in indicated cell lines without or with IFN-γ treatment. One way ANOVA with Tukey's HSD for multiple corrections.



Sup Figure 13. Common gene signature from immunotherapy resistant clones. **A.** Gene-set enrichment analysis of the overlapping genes between the IE1 and IE2 clones. Red indicates upregulated gene-sets with a significant FDR q-value when multiple testing is taken into account, orange indicates significant upregulated gene-sets with a nominal p value, and blue indicates downregulated gene-sets with a significant FDR q-value. **B.** Heatmap of unsupervised hierarchical clustering of the immunotherapy resistance signature genes in the TCGA breast cancer dataset, with tumours grouped based on top 30%, middle 40%, and bottom 30% overall gene signature score. **C.** Correlation plot of immunotherapy resistance signature score with cytotoxic T lymphocyte (CTL) levels in the METABRIC dataset. **D.** Correlation plot of immunotherapy resistance signature score with CTL levels in the TCGA dataset.