Immune surveillance in clinical regression of pre-invasive squamous cell lung cancer

Adam Pennycuick[1]*, Vitor H. Teixeira[1]*, Khalid AbdulJabbar[2,3]†, Shan E Ahmed Raza[2,3,4]†, Tom Lund[5,6,7]†, Ayse U. Akarca[8], Rachel Rosenthal[9], Lukas Kalinke[1], Deepak P. Chandrasekharan[1], Christodoulos P. Pipinikas[10], Henry Lee-Six[11], Robert E. Hynds[1,9,10], Kate H. C. Gowers[1], Jake Y. Henry[3,5], Fraser R. Millar[12], Yeman B. Hagos[2,3], Celine Denais[1], Mary Falzon[8], David A. Moore[7,8], Sophia Antoniou[1], Pascal F. Durrenberger[1], Andrew J. Furness[5,13], Bernadette Carroll[1], Claire Marceaux[14], Marie-Liesse Asselin-Labat[14,15], William Larson[16], Courtney Betts[16], Lisa M. Coussens[16], Ricky M. Thakrar[1], Jeremy George[1], Charles Swanton[7,9,10], Christina Thirlwell[10,17], Peter J. Campbell[11], Teresa Marafioti[8], Yinyin Yuan[2,3], Sergio A. Quezada[5,6,7], Nicholas McGranahan[10,18]#, Sam M. Janes[1,7]#

1. Lungs for Living Research Centre, UCL Respiratory, University College London, London, U.K.
2. Centre for Evolution and Cancer, The Institute of Cancer Research, London, U.K.
3. Division of Molecular Pathology, The Institute of Cancer Research, London, U.K.
4. Department of Computer Science, University of Warwick, Coventry, UK
5. Cancer Immunology Unit, University College London Cancer Institute, University College London, London, U.K.
6. Research Department of Haematology, University College London Cancer Institute, University College London, London, U.K.
7. UCL Manchester Lung Cancer Centre of Excellence
8. Department of Cellular Pathology, University College London Hospital, London, U.K.
9. Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, U.K.
10. University College London Cancer Institute, London, U.K.
11. The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, U.K.

28 12. Cancer Research UK Edinburgh Centre, Institute of Genetics and Molecular
29 Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.
30 13. The Royal Marsden NHS Foundation Trust
31 14. Personalised Oncology Division, The Walter and Eliza Hall Institute of Medical
32 Research, Melbourne, Australia
33 15. Knight Cancer Institute, Cancer Early Detection and Advanced Research (CEDAR)
34 Center, Oregon Health & Science University, Portland, OR USA
35 16. Knight Cancer Institute, Department of Cell, Developmental and Cancer Biology,
36 Oregon Health & Science University, Portland, OR USA
37 17. University of Exeter College of Medicine and Health, UK
38 18. Cancer Genome Evolution Research Group, University College London Cancer
39 Institute, London, U.K.
40
41 [#] **Corresponding authors:**
42
43 **Professor Sam M. Janes**
44 **Address:** Lungs for Living Research Centre, UCL Respiratory, University College London, 5
45 University Street, London, WC1E 6JF, U.K.
46 **Phone:** (+44) 020 3549 5979
47 **E-mail:** s.janes@ucl.ac.uk
48
49 **Dr Nicholas McGranahan**
50 **Address:** Cancer Research UK Lung Cancer Centre of Excellence, University College
51 London Cancer Institute, University College London, London, WC1E 6AG, UK
52 **Phone:** (+44) 020 7679 6296
53 **E-mail:** nicholas.mcgranahan.10@ucl.ac.uk
54
55 *,[†] These authors contributed equally to this work.

56 **Running Title:** Immune surveillance in regression of preinvasive lung cancer

57

58 **Conflict of Interest Statement**

59 S.A.Q. and C.S. are co-founders of Achilles Therapeutics. C.S. is a shareholder of
60 Apogen Biotechnologies, Epic Bioscience, GRAIL, and has stock options in Achilles
61 Therapeutics. R.R. and N.M. have stock options in and have consulted for Achilles
62 Therapeutics. L.M.C. is a paid consultant for Cell Signaling Technologies, received reagent
63 and/or research support from Plexxikon Inc., Pharmacyclics, Inc., Acerta Pharma, LLC,
64 Deciphera Pharmaceuticals, LLC, Genentech, Inc., Roche Glycart AG, Syndax
65 Pharmaceuticals Inc., Innate Pharma, and NanoString Technologies, and is a member of the
66 Scientific Advisory Boards of Syndax Pharmaceuticals, Carisma Therapeutics, Zymeworks,
67 Inc, Verseau Therapeutics, Cytomix Therapeutics, Inc., and Kineta Inc.

68

69 **Abstract**

70 Before squamous cell lung cancer develops, pre-cancerous lesions can be found in
71 the airways. From longitudinal monitoring, we know that only half of such lesions become
72 cancer, whereas a third spontaneously regress. While recent studies have described the
73 presence of an active immune response in high-grade lesions, the mechanisms
74 underpinning clinical regression of pre-cancerous lesions remain unknown. Here, we show
75 that host immune surveillance is strongly implicated in lesion regression. Using
76 bronchoscopic biopsies from human subjects, we find that regressive carcinoma *in-situ*
77 lesions harbour more infiltrating immune cells than those that progress to cancer. Moreover,
78 molecular profiling of these lesions identifies potential immune escape mechanisms
79 specifically in those that progress to cancer: antigen presentation is impaired by genomic
80 and epigenetic changes, CCL27/CCR10 signalling is upregulated, and the immunomodulator
81 TNFSF9 is downregulated. Changes appear intrinsic to the CIS lesions as the adjacent
82 stroma of progressive and regressive lesions are transcriptomically similar.

83

84 **Statement of Significance**

85 Immune evasion is a hallmark of cancer. For the first time, this study identifies mechanisms
86 by which pre-cancerous lesions evade immune detection during the earliest stages of
87 carcinogenesis and forms a basis for new therapeutic strategies that treat or prevent early
88 stage lung cancer.

89

90

91 **Introduction**

92 Before the development of lung squamous cell carcinoma (LUSC), pre-invasive lesions
93 can be observed in the airways. These evolve stepwise, progressing through mild and
94 moderate dysplasia (low-grade lesions) to severe dysplasia and carcinoma *in-situ* (CIS;
95 high-grade lesions), before the development of invasive cancer(1). In cross-sectional
96 studies, markers of immune sensing and escape have been associated with increasing
97 grade(2). However, longitudinal bronchoscopic surveillance of such lesions has shown that
98 progression of pre-invasive lesions to cancer is not inevitable; only half of high-grade CIS
99 lesions will progress to cancer within two years, whereas a third will spontaneously
100 regress(3). Our previous work defined the genomic, transcriptomic and epigenetic landscape
101 of carefully phenotyped airway CIS lesions(4). Here, we combine these data with
102 immunohistochemistry (IHC), imaging and transcriptomic analysis of adjacent stroma (**Table**
103 **S1; Figure S1**) to assess the role of immune surveillance in lesion regression. We identify
104 key immune escape mechanisms enriched in pre-invasive lesions which later progressed to
105 cancer. Understanding these mechanisms may offer new therapeutic strategies to induce
106 regression and prevent the development of invasive disease.

107

108 **Results**

109 To assess our hypothesis that lesion regression is driven by immune surveillance, we
110 used a deep-learning approach(5) to quantify lymphocytes from hematoxylin and eosin
111 (H&E) stained slides in a large dataset of 112 samples from 62 patients, which contained
112 more infiltrating lymphocytes in regressive lesions than progressive (**Figure 1a**; p=0.049).
113 We next performed immunohistochemistry (IHC) on 28 progressive and 16 regressive CIS
114 lesions from 29 patients (**Figure 1b-c**). Regressive lesions showed higher concentrations of
115 intra-lesional CD8+ cytotoxic T-cells (**Figure 1a**; p=0.055) but no significant difference in
116 CD4+ T helper cells (p=0.26) or FOXP3+ regulatory T cells (p=0.42). We then quantified
117 immune cells in stromal regions adjacent to CIS lesions, but found no significant differences
118 between progressive and regressive lesions for CD8+ (p=0.50), CD4+ (p=0.43) or FOXP3+

119 (p=0.64) cells. We confirmed these findings in an independent dataset of 19 progressive and
120 9 regressive samples subjected to multiplex IHC(6,7) (mIHC) using a wider antibody panel of
121 lymphoid biomarkers (**Table S2**), in which we again observed that regressive lesions had an
122 increased proportion of infiltrating lymphocytes (**Figure 2a**; p=0.032). Specifically, regressive
123 lesions showed significantly more infiltrating CD3+CD8+ cytotoxic T-cells (p=0.017) but no
124 significant difference in CD3+CD4+ T helper cells (p=0.18), T regulatory cells (p=0.12), B-
125 cells (p=0.12), macrophages (p=0.79) or neutrophils (p=0.53). In the mIHC cohort, the
126 proportion of CD3+CD8+ cells positive for granzyme B and EOMES was similar between
127 progressive and regressive lesions (p=0.63 and p=0.18 respectively) which may indicate that
128 disruption of T-cell infiltration into lesions has a greater impact on their capacity for immune
129 evasion than impairment of cytotoxic function or differentiation. Again, stromal regions in this
130 cohort showed no significant differences between progressive and regressive lesions.

131 For a broader assessment of transcriptomic differences between CIS lesions and
132 their adjacent stroma, we isolated epithelial tissue and paired stroma separately using laser
133 capture microdissection for 10 progressive and 8 regressive CIS lesions. Similarly to IHC
134 data, cell type deconvolution analysis using the Danaher method(8) demonstrated higher
135 infiltrating lymphocytes within regressive lesions (**Figure 2b**; p=0.0046), as did
136 deconvolution of methylation data from 36 progressive and 18 regressive CIS lesions using
137 methylCIBERSORT(9) (**Figure 2c**; p=0.0081). Comparing predictions for individual cell
138 types across gene expression and methylation data found an increase in most immune cell
139 types in regressive lesions compared to progressive (**Table S3**).

140 Analysis of cytokines classically considered to be pro- or anti-inflammatory within the
141 epithelial compartment (**Table S4**) demonstrated an increase in pro-inflammatory ($p=3.7\times10^{-}$
142 ^{4}) but not anti-inflammatory (p=0.32) response in regressive lesions compared to
143 progressive (**Figure S2a-f**). *IL2*, *TNF, IL12A* and *IL23A* were all increased in regressive
144 lesions (**Figure S3a-b**; FDR=0.0081, FDR=0.00051, FDR=0.00078, FDR=0.011
145 respectively). Only *CXCL8* was upregulated in progressive samples compared to regressive
146 (FDR=0.0063); produced by macrophages, the expression of *CXCL8* correlated strongly with

147 macrophage quantification from deconvoluted gene expression data (r^2=0.62, p=0.007).
148 Taken together, these data are in keeping with a model in which inflammation via pathways
149 including IL-2 and TNF fosters effective immune surveillance, whilst lesion-associated
150 macrophages – similar to tumor-associated macrophages in advanced cancers – have an
151 immunosuppressive effect.

152 Given the well-known immunosuppressive effects of smoking, we hypothesised that
153 patients who were current smokers were more likely to show reduced immune infiltrate and
154 therefore a higher chance of progression. Smoking status was available for 132 CIS lesions
155 from 59 patients (24 lesions from 13 current smokers; 104 from 43 former smokers; 4 from 3
156 never smokers; **Figure S4a-j**). Using a Cochrane-Armitage test to look for a trend from
157 Current to Former to Never smokers, we found a trend towards higher chance of regression
158 (p=0.002) and more infiltrating lymphocytes (p=0.095). This trend is still observed, yet no
159 longer statistically significant, using a bootstrapping method to account for samples from the
160 same patient (p=0.069 for regression; p=0.12 for infiltrating lymphocytes). Interestingly,
161 within the former-smoker group we did not observe increasing lymphocytes or chance of
162 regression with increasing time since quitting smoking, suggesting that the observed
163 differences in outcome are driven by the active process of smoking and its direct effects on
164 the immune response, rather than by chronic processes of airway remodelling and
165 repair(10).

166 Recent advances have demonstrated heterogeneity of lung cancer immune
167 infiltration, with patients whose tumors have predominantly infiltrated 'immune hot' regions
168 having improved survival as compared to those with abundant poorly infiltrated, 'immune
169 cold' regions(11,12). Hierarchical clustering of immune cell quantification by mIHC and by
170 deconvolution of both transcriptomic and epigenetic data demonstrated clear clusters of
171 'cold' lesions, almost all of which progressed to cancer (**Figure 2d-f**). However, we also
172 observed some 'hot' progressive lesions, suggesting the presence of other immune evasion
173 mechanisms in these lesions. We therefore sought to address two questions: firstly, could
174 deficits in antigen presentation and immune recruitment in progressive lesions be identified,

175 which could explain the observed 'cold' lesions? Secondly, could disordered immune cell
176 function explain the existence of progressive immune 'hot' lesions?
177 The acquisition of mutations that result in clonal neoantigens drives T cell
178 immunoreactivity in cancer(13). We hypothesised that immune-active regressive lesions
179 may contain more neoantigens than progressive lesions, however, this was not supported by
180 whole-genome sequencing data(4) (n=39). Predicted neoantigens correlated very closely
181 with mutational burden (r^2=0.94), and progressive lesions have been shown to have
182 significantly higher mutational burden than regressive lesions(4), therefore more
183 neoantigens were identified in progressive than regressive lesions (**Figure S5a-b**; p=0.088).
184 This remained true when the analysis was limited to clonal neoantigens (**Figure S5c;**
185 p=0.023) and there was no difference in the proportion of neoantigens that were clonal
186 (**Figure S5d**; p=0.76). Further, there were no significant differences in binding affinity
187 (p=0.46) or differential agretopicity index(14)(p=0.58) and the ratio of observed to expected
188 neoantigens ("depletion score"(15)) was not significantly different (**Figure S5e-h**; p=0.94),
189 therefore the putative neoantigens themselves were not qualitatively different in the
190 regressive group. The increased number of neoantigens identified in progressive lesions
191 suggests that immune escape mechanisms must be active in these lesions; indeed, these
192 antigens may act as a selection pressure to promote the development of immune
193 escape(16). Importantly, no overlap in putative tumor neoantigens was observed between
194 different patients suggesting that vaccine-based approaches aiming to prevent progression
195 will most likely need to be designed on a personalised basis.
196 Given that neoantigens are present in progressive lesions, we assessed the ability of
197 these lesions to present antigens to the immune system. In cancer, genomic alterations
198 have previously been associated with modulation of immune response(17,18). We studied
199 mutations and copy number burden of 62 genes expressed by cancer cells which are
200 involved in the following pathways: antigen presentation by MHC mechanisms, antigen
201 processing and immunomodulation (stimulation and inhibition of T-cell responses) (**Figure**
202 **3**). Mutations and CNAs in these genes were more prevalent in progressive than regressive

203 samples (p=0.003). Four of these genes – *B2M, CHUK, KDR* and *CD80* – had a significantly
204 elevated dN/dS ratio (19) – comparison of the rates of non-synonymous to synonymous
205 mutations – indicating positive selection for acquisition of mutations in these genes. We
206 observe that expression of immunostimulatory genes predominantly positively correlates
207 with infiltrating lymphocytes in CIS, and these genes are mostly downregulated in
208 progressive compared to regressive CIS. Conversely, inhibitory genes predominantly
209 correlate negatively with infiltrating lymphocytes and are upregulated in progressive lesions.

210 Loss of heterozygosity (LOH) in the HLA region, which is found in 61% of LUSC
211 patients(20), was identified in 34% of patients with CIS lesions. Interestingly, a similar
212 proportion of LUSC patients (28%) demonstrated *clonal* HLA LOH(20), suggesting that such
213 clonal events may often occur prior to tumor invasion; future longitudinal studies will be
214 required to confirm this. We did not find a statistically significant difference in the prevalence
215 of HLA LOH between progressive and regressive lesions (p=0.25) although sample numbers
216 were small. Expression of *HLA-A* was significantly reduced in progressive compared to
217 regressive lesions (p=1.9×10^{-10}).

218 Additionally, hypermethylation of the HLA region, which is well-described in invasive
219 cancers(21,22), was commonly observed, suggesting that epigenetic HLA silencing may be
220 an important immune escape mechanism in pre-invasive disease. Genome-wide differential
221 methylation analysis between progressive and regressive lesions identified differentially
222 methylated regions (DMRs) including a striking cluster of hypermethylation in chromosome 6
223 ((4); **Figure S6a-b**), covering a region containing all of the major HLA genes. This cluster
224 was also identified in analysis of 370 LUSC versus 42 control samples published by the
225 Cancer Genome Atlas(23). Further analysis of TCGA data demonstrate strong evidence for
226 epigenetic silencing of multiple genes in the antigen presentation pathway: mean
227 methylation beta value over the gene is inversely correlated with expression for *HLA-A* (r^2=-
228 0.32, p=2.5×10^{-10}), *HLA-B* (r^2=-0.42, p<2.2×10^{-16}), *HLA-C* (r^2=-0.18, p=3.6×10^{-4}), *TAP1* (r^2=-
229 0.53, p<2.2×10^{-16}) and *B2M* (r^2=-0.38, p=1.1×10^{-14}). Similar trends were observed in CIS

230 data (**Figure S7a-b**). The methylation pattern affecting these genes is predominantly
231 promoter hypermethylation (**Figure S8**).
232 Demethylating agents have been shown to promote immune activation through
233 improved antigen presentation, immune migration and T cell activity(24–26). These data
234 support the case for moving on-going trials of demethylating agents in combination with
235 immunotherapy from advanced lung cancer into early disease (examples of such trials
236 include NCT01928576 and NCT03220477, registered at https://clinicaltrials.gov/).
237 Additionally, several other cancer-associated pathways are known to be affected by
238 methylation changes(4), therefore the benefits of these drugs may extend beyond immune
239 activation. Nevertheless, we note with caution that some key immune genes demonstrate
240 *positive* correlations in TCGA data between gene expression and methylation, including the
241 immune co-stimulating ligand *TNFSF9* (coding for 4-1BBL) (r^2=0.32, p=1.7×10^{-10}) and the
242 MHC class II transcriptional activator *CIITA* (r^2=0.39, p=2.5×10^{-15}) (**Figure S7**). Further
243 studies will be required to demonstrate that immunological benefits of demethylating agents
244 are not outweighed by effects on these important pathways.
245 Despite this evidence for impairment of antigen presentation mechanisms in CIS, we
246 do observe 'immune hot' CIS lesions which progress to cancer. We therefore next
247 considered functional and microenvironment-related mechanisms to explain how these
248 lesions were able to evade immune predation.
249 To study microenvironment effects on the immune response, we performed gene
250 expression profiling on laser-captured stromal tissue taken from regions adjacent to CIS
251 lesions. In contrast to data from gastrointestinal pre-invasive lesions(27), no genes were
252 significantly differentially expressed on comparing stromal expression between progressive
253 (n=10) and regressive (n=8) lesions when a FDR of <0.1 was applied. This result holds true
254 with restricted hypothesis testing considering only genes that are related to immunity and
255 inflammation (**Figure 4a-b; Table S4**).
256 Targeting immunomodulatory molecules such as PD-1 now forms part of first-line
257 lung cancer management(28). PD-L1 expression is common in invasive LUSC with

258 estimates of positivity ranging from 34% to 52%, depending on criteria(29). Whilst we did not
259 identify transcriptional upregulation of the PD-L1 gene (*CD274*; **Figure 4c-d**), IHC data
260 identified 3 samples with >25% of epithelial cells (PanCK+) also positive for PD-L1 (**Figure**
261 **4e**), all of which progressed to cancer, suggesting that targeting this pathway early in the
262 clinical course may have therapeutic benefit in selected patients.

263 To investigate the role of immunomodulatory molecules more broadly in pre-invasive
264 immune escape, we performed differential expression analysis between progressive and
265 regressive lesions, focused on 28 known immunomodulatory genes (**Table S4**). *TNFSF9* (4-
266 1BBL, CD137L) was significantly downregulated in progressive lesions ($FDR=4.34x10^{-5}$;
267 **Figure 4c-d**) with no corresponding change identified in its receptor *TNFRSF9* (FDR=0.6).
268 These findings were corroborated by IHC (**Figure 4e-f**). *TNFSF9* promotes activation of T
269 cells and natural killer (NK) cells(30); in CIS lesions *TNFSF9* expression correlates with
270 cytotoxic cell (r^2=0.77, p=0.0002) and NK cell infiltration (r^2=0.54, p=0.02), as predicted from
271 gene expression data. Agonists of the *TNFSF9* receptor have been shown to be clinically
272 efficacious in several cancers(31–33) and these data support their investigation in targeted
273 early lung cancer cohorts. Furthermore, individual lesions showed notably high or low
274 expression of other immunomodulatory genes, raising the possibility that other
275 immunomodulators may be targets for therapy in individual cases (**Figure S9**).

276 To identify differences in cytokine responses between progressive and regressive
277 lesions, we calculated the ligand:receptor mRNA expression ratio for 52 known
278 cytokine:receptor pairs(34). Only one, *CCL27:CCR10*, was significant with FDR < 0.01 (Fold
279 change 1.55, FDR 0.003); progressive samples express more *CCL27* ($p=2.6x10^{-6}$) and less
280 *CCR10* ($p=0.1x10^{-4}$) than regressive (**Figure 4c-d**). Whilst sample numbers were small,
281 these findings were broadly supported by IHC (**Figure 4e-g**). *CCL27:CCR10* signaling has
282 been associated with immune escape in melanoma through PIK/Akt activation in a mouse
283 model(35); in CIS, *CCL27* expression correlates with expression of both *PIK3CA* (r2=0.61,
284 p=0.008) and *AKT1* (r2=0.68, p=0.002) (**Figure S10a-b**). *CCL27* is minimally expressed in
285 both normal lung tissue and invasive squamous cell lung cancer(23,36), suggesting that this

286 effect is specific to early carcinogenesis and therefore warrants further investigation as a
287 target for preventative therapy.

288 Our previous research highlighted occasional cases of 'late progressive' lesions,
289 which met a clinical endpoint of regression (defined by the subsequent biopsy at the same
290 site showing resolution to normal epithelium or low-grade dysplasia) but the index CIS
291 biopsy had the molecular appearance of a progressive lesion, and it indeed subsequently
292 developed cancer months or years later. Clinical review identified 11 lesions across the 53
293 regressive lesions in our current cohort (20.7%) that at later clinical follow up subsequently
294 progressed to cancer, and hence are termed 'late progressive'. These included 4 previously
295 published lesions subjected to whole-genome sequencing and/or methylation and shown to
296 display the genomically unstable appearance of progressive lesions, as well as 7 with
297 immunohistochemistry data and 10 with lymphocyte quantification performed from H&E
298 slides (**Table S1; Figure S1**). Interestingly, based on these data, late progressive lesions
299 appear immunologically similar to regressive lesions, showing increased infiltration with
300 lymphocytes and CD8+ T-cells compared to progressive lesions (**Figure S11**).

301 Whilst we acknowledge that sample numbers are small when examining subgroups
302 of regressive lesions in this way, our data support a model in which lesions can be
303 considered on two axes: genomic stability and immune competence. Our previous work
304 predicts that chromosomally unstable lesions will usually progress, implying that they have
305 escaped immune predation. Yet some may regress if they remain immune competent only to
306 later progress, potentially due to their genomic instability making them more likely to evolve
307 immune escape mechanisms during regression, and hence become 'late progressors'. Of 11
308 late progressors in this cohort, median time from regressive index biopsy to progression was
309 3.2 years (range 0.8-4.6 years). This time period represents a change from a point of known
310 immune competence to demonstrated immune escape. Hence, we might estimate that a
311 successful therapeutic strategy to block a particular immune escape mechanism might delay
312 the onset of cancer by around 3 years. Of the remaining 42 regressive samples in this
313 cohort, median follow-up time was 4.73 years (range 0.42-13.5 years), suggesting that

314 genomically 'stable' samples are likely to regress and remain regressed long-term. Given
315 their immunological competence, late progressors are included in the regressive cohort
316 when analysing immune escape mechanisms in this study.
317
318
319 **Discussion**
320 In summary, we present evidence that immune surveillance may play a
321 critical role in spontaneous regression of pre-cancerous lesions of the airways. Whilst recent
322 cross-sectional studies have greatly furthered our understanding of immune signals prior to
323 cancer invasion, and indeed at earlier disease stages than CIS(2,12), we have for the first
324 time shown an association with lesion regression. Including such outcome data offers insight
325 into the dynamics of immune surveillance and evasion; assuming that lesion regression is
326 driven by immune surveillance – which is likely based on our data – we are able to directly
327 compare preinvasive lesions which are immune competent (regressed) with those that are
328 able to evade immune predation (progressed). Analysis of 'late progressive' samples
329 furthers this model by providing estimates of timescales over which immune evasion
330 evolves. Hence we provide a roadmap for manipulation of the immune system as a cancer
331 intervention strategy, by identifying and targeting differences between these two immune
332 states.
333 To this end, we identify mechanisms of immune escape present before the point of
334 cancer invasion, many of which offer potential therapeutic targets. These data present an
335 opportunity to induce regression and prevent cancer development. Demethylating agents, 4-
336 1BB agonists and CCL27 blockade are therapeutic candidates that warrant further research,
337 as well as targeting the PD-1/PD-L1 axis in highly selected patients. As a result of field
338 carcinogenesis, patients with pre-invasive lesions are at risk of synchronous cancers at
339 other sites, which are likely to be clonally related(4,37) and therefore may benefit from
340 systemic immunomodulatory treatment. The data presented here support a new paradigm of
341 personalised immune-based systemic therapy in early disease.

342

343 **Methods**

344 Additional methods are provided in a supplementary file accompanying this manuscript.

345

346 **Ethical approval**

347 All tissue and bronchial brushing samples were obtained under written informed patient
348 consent and were fully anonymized. Study approval was provided by the UCL/UCLH Local
349 Ethics Committee (REC references 06/Q0505/12 and 01/0148). All relevant ethical
350 regulations were followed.

351

352 **Cohort description and patient characteristics**

353 For over 20 years, patients presenting with pre-invasive lesions, which are
354 precursors of squamous cell lung cancer (LUSC), have been referred to the UCLH
355 Surveillance Study. As previously described(3), patients undergo repeat bronchoscopy every
356 four months, with definitive treatment performed only on detection of invasive cancer.
357 Autofluorescence bronchoscopy is used to ensure the same anatomical site is biopsied at
358 each time point. Gene expression, methylation and whole genome sequencing data of
359 carcinoma in-situ (CIS) samples have been performed on this cohort, and data have been
360 published(4). These data are used in this study.

361 All patients enrolled in the UCLH Surveillance Study who met a clinical end point of
362 progression or regression were included; by definition they underwent an 'index' CIS biopsy
363 followed by a diagnostic cancer biopsy (progression) or a normal/low-grade biopsy
364 (regression) four months later. Index lesions were identified between 1999 and 2017. Cases
365 meeting an end-point of regression underwent clinical review to identify those which
366 subsequently progressed; 11 samples (20.7%) were identified, which are described as 'late
367 progressors' in the main text. Of these 11, median time from 'regressive' index biopsy to
368 progression was 3.2 years (range 0.8-4.6 years) whilst the remaining 42 samples had a
369 median follow up time of 4.73 years (range 0.42-13.5 years). Whilst we cannot fully exclude
370 that any regressive sample may later develop cancer, the fact that median follow up in the

371 study group was longer than the maximum follow up in the late progression group suggests
372 that late progression in included samples is unlikely.
373 All samples underwent laser capture microdissection (LCM) to ensure only CIS cells
374 underwent molecular profiling. Methods for sample acquisition, quality control and mutation
375 calling are as previously described, as are full details regarding patient clinical
376 characteristics.
377 Briefly, gene expression profiling was performed using both Illumina and Affymetrix
378 microarray platforms. Normalisation was performed using proprietary Illumina software and
379 the RMA method of the *affy*(38) Bioconductor package respectively. This study includes 18
380 previously unpublished gene expression arrays from stromal tissue. These samples were
381 collected using LCM to identify stromal regions adjacent to 18 already-published CIS
382 samples (corresponding to the 18 samples undergoing Affymetrix microarray profiling
383 described above). These new stromal samples underwent Affymetrix profiling using the
384 exact same methodology as previously described for CIS tissue samples. To avoid issues
385 related to batch effects between platforms, the analyses in this paper utilise only samples
386 profiled on Affymetrix microarrays, which include both CIS and matched stromal samples
387 (see **Supplementary Methods** and **Table S5**).
388 Methylation profiling was performed using the Illumina HumanMethylation450k
389 microarray platform. All data processing was performed using the ChAMP Bioconductor
390 package(39).
391 Whole genome sequencing data was obtained using the Illumina HiSeq X Ten
392 system. A minimum sequencing depth of 40x was required. BWA-MEM was used to align
393 data to the human genome (NCBI build 37). Unmapped reads and PCR duplicates were
394 removed. Substitutions, insertions-deletions, copy number aberrations and structural
395 rearrangements were called using CaVEMan(40), Pindel(41,42), ASCAT(43) and Brass(44)
396 respectively.
397
398

399
400
401
402 **Sample selection for profiling**

403 As previously described, all patients enrolled in the surveillance programme discussed
404 above were considered for this study. For a given CIS lesion under surveillance, when a
405 biopsy from the same site in the lung showed evidence of progression to invasive cancer or
406 regression to normal epithelium or low-grade dysplasia, we defined the preceding CIS
407 biopsy as a progressive or regressive 'index' lesion respectively. Due to the small size of
408 bronchoscopic biopsy samples, not all profiling techniques were applied to all samples.
409 Patients with Fresh Frozen (FF) samples underwent whole genome sequencing and/or
410 methylation analysis depending on sample quality. Patients with formalin-fixed paraffin-
411 embedded (FFPE) samples underwent gene expression analysis. Further detail is available
412 in our previous manuscript(4). Additionally, any patient with an available FFPE block
413 underwent image analysis as described below, and all patients with Affymetrix-based gene
414 expression profiling underwent further profiling of laser-captured adjacent stroma.
415
416
417 **Statistical Methods**

418 Unless otherwise specified, all analyses were performed in an R statistical
419 environment (v3.5.0; www.r-project.org/) using Bioconductor(45) version 3.7. Code to
420 reproduce a specific statistical test is publicly available at the Github repository above.

421 Unless otherwise stated, comparisons of means between two independent groups are
422 performed using a two-sided Wilcoxon test. In some cases, multiple samples have been
423 profiled from the same patient, although always from distinct sites within the lung. In such
424 cases we used mixed effects models to compare means between groups, treating the
425 patient ID as a random effect, as implemented in the Bioconductor *lme4* library(46), with p-
426 values calculated using the Anova method from the Bioconductor *car* library (available from

427 https://cran.r-project.org/web/packages/car). Differential expression was performed using the
428 *limma*(47) Bioconductor package to compare microarray data between two groups. When
429 adjustment for multiple correction is required we quote a False Discovery Rate (FDR) which
430 is calculated using the Benjamini-Hochberg method(48). Cluster analysis and visualization
431 was performed using the *pheatmap* Bioconductor package (available from https://cran.r-
432 project.org/web/packages/pheatmap/).
433

434 **Data Availability**

435 All raw data used in this study is publicly available. Previously published CIS gene
436 expression and methylation data is stored on GEO under accession number GSE108124;
437 matched stromal gene expression data is stored under accession number GSE133690.
438 Previously published CIS whole genome sequencing data is available from the European
439 Genome Phenome Archive (https://www.ebi.ac.uk/ega/) under accession number
440 EGAD00001003883. Annotated H&E images of all samples used for lymphocyte
441 quantification were deposited to the Image Data Resource (https://idr.openmicroscopy.org)
442 under accession number idr0082.
443

444 **Code Availability**

445 All code used in our analysis will be made available at http://github.com/ucl-
446 respiratory/cis_immunology on publication. All software dependencies, full version
447 information, and parameters used in our analysis can be found here.
448

449 **Author Contributions**

450 A.P. and V.H.T. contributed equally to this work, as did K.A., S.E.A.R. and T.L.. A.P.,
451 V.H.T., N.M. and S.M.J. co-wrote the manuscript. S.M.J., S.A.Q., V.H.T. and A.P. conceived
452 the study design. V.H.T., D.C., F.R.M. and S.A. performed stromal LCM and gene
453 expression profiling experiments. C.P.P. and C.T. performed methylation experiments. H.L-
454 S. and P.J.C. performed genomic experiments. A.A., T.L., J.Y.H., L.K. and T.M. designed

455 and performed IHC experiments. Further quantitative multiplex IHC was performed by C.M.,
456 M-L.A-L., W.L., C.B. and L.C.. K.A., S.E.A.R., Y.B.H. and Y.Y. performed cell quantification
457 on H&E and IHC images. S.M.J., P.J.G., B.C. and R.M.T. led the bronchoscopic surveillance
458 programme through which samples were obtained. M.F. and D.M. performed histological
459 review. P.F.D. performed pathological processing. A.P. performed bioinformatic analysis,
460 supported by R.R. and N.M.. R.E.H., K.H.C.G., C.D., A.F., N.M., C.S., C.T., S.A.Q. and N.M.
461 gave advice and reviewed the manuscript. S.M.J. provided overall study oversight.
462

463 **Acknowledgements**

464 We thank all of the patients who participated in this study. We thank P. Rabbitts, A.
465 Banerjee and C. Read for their early development of the study. The results published here
466 are in part based on data generated by a TCGA pilot project established by the National
467 Cancer Institute and National Human Genome Research Institute. Information about TCGA
468 and the investigators and institutions that constitute the TCGA research network can be
469 found at http://cancergenome.nih.gov. R.E.H., N.M., P.J.C., and S.M.J. are supported by
470 Wellcome Trust fellowships. S.M.J. is also supported by the Rosetrees Trust, the Welton
471 Trust, the Garfield Weston Trust, the Stoneygate Trust and UCLH Charitable Foundation, as
472 well as Stand Up to Cancer-LUNGevity-American Lung Association Lung Cancer
473 Interception Dream Team Translational Cancer Research Grant (Grant Number: SU2C-
474 AACR-DT23-17). Stand Up To Cancer is a division of the Entertainment Industry
475 Foundation. The research grant is administered by the American Association for Cancer
476 Research, the scientific partner of SU2C. V.T., C.P., R.E.H. and S.M.J. have been funded by
477 the Roy Castle Lung Cancer Foundation. A.P. and D.C. are funded by Wellcome Trust
478 clinical PhD training fellowships. H.L.-S. is funded by the Wellcome Trust Sanger Institute
479 non-clinical PhD studentship. C.T. was a CRUK Clinician Scientist. This work was partially
480 undertaken at UCLH/UCL, who received a proportion of funding from the Department of
481 Health's NIHR Biomedical Research Centre's funding scheme (S.M.J.). R.E.H., D.M., N.M.,
482 C.S., and S.M.J. are part of the CRUK Lung Cancer Centre of Excellence. Y.Y.

483 acknowledges funding from Cancer Research UK Career Establishment Award, Breast
484 Cancer, Children's Cancer and Leukaemia Group, NIH U54 CA217376 and R01 CA185138,
485 CDMRP Breast Cancer Research Program Award, CRUK Brain Cancer Award (TARGET-
486 GBM), European Commission ITN, Wellcome Trust, and The Royal Marsden/ICR National
487 Institute of Health Research Biomedical Research Centre. S.A.Q. is funded by a CRUK
488 Senior Cancer Research Fellowship, a CRUK Biotherapeutic Program Grant, the Cancer
489 Immunotherapy Accelerator Award (CITA-CRUK) and the Rosetrees Trust. LMC
490 acknowledges funding from the National Institutes of Health (1U01 CA224012, U2C
491 CA233280, R01 CA223150, R01 , R01 CA226909, R21 HD099367), the Knight Cancer
492 Institue, and the Brenden-Colson Center for Pancreatic Care at OHSU. The funders had no
493 role in study design, data collection and analysis, decision to publish or preparation of the
494 manuscript.

References

1. Nicholson AG, Perry LJ, Cury PM, Jackson P, McCormick CM, Corrin B, et al. Reproducibility of the WHO/IASLC grading system for pre-invasive squamous lesions of the bronchus: a study of inter-observer and intra-observer variation. Histopathology. 2001;38:202–8.

2. Mascaux C, Angelova M, Vasaturo A, Beane J, Hijazi K, Anthoine G, et al. Immune evasion before tumour invasion in early lung squamous carcinogenesis. Nature. 2019;571:570–5.

3. George PJ, Banerjee AK, Read CA, O'Sullivan C, Falzon M, Pezzella F, et al. Surveillance for the detection of early lung cancer in patients with bronchial dysplasia. Thorax. 2007;62:43–50.

4. Teixeira VH, Pipinikas CP, Pennycuick A, Lee-Six H, Chandrasekharan D, Beane J, et al. Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. Nat Med. 2019;25:517–25.

5. AbdulJabbar, K. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. Nature Medicine (accepted for publication). 2020;

6. Tsujikawa T, Kumar S, Borkar RN, Azimi V, Thibault G, Chang YH, et al. Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. Cell Reports. 2017;19:203–17.

7. Banik G, Betts CB, Liudahl SM, Sivagnanam S, Kawashima R, Cotechini T, et al. High-dimensional multiplexed immunohistochemical characterization of immune contexture in human cancers. Meth Enzymol. 2020;635:1–20.

8. Danaher P, Warren S, Dennis L, D'Amico L, White A, Disis ML, et al. Gene expression markers of Tumor Infiltrating Leukocytes. J Immunother Cancer. 2017;5:18.

9. Chakravarthy A, Furness A, Joshi K, Ghorani E, Ford K, Ward MJ, et al. Pan-cancer deconvolution of tumour composition using DNA methylation. Nature Communications. 2018;9:3220.

10. Yoshida K, Gowers KHC, Lee-Six H, Chandrasekharan DP, Coorens T, Maughan EF, et al. Tobacco smoking and somatic mutations in human bronchial epithelium. Nature. 2020;578:266–72.

11. Rosenthal R, Cadieux EL, Salgado R, Bakir MA, Moore DA, Hiley CT, et al. Neoantigen-directed immune escape in lung cancer evolution. Nature. 2019;567:479–85.

12. Beane JE, Mazzilli SA, Campbell JD, Duclos G, Krysan K, Moy C, et al. Molecular subtyping reveals immune alterations associated with progression of bronchial premalignant lesions. Nature Communications. 2019;10:1856.

530 13. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal
531 neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint
532 blockade. Science. 2016;351:1463–9.

533 14. Ghorani E, Rosenthal R, McGranahan N, Reading JL, Lynch M, Peggs KS, et al.
534 Differential binding affinity of mutated peptides for MHC class I is a predictor of
535 survival in advanced lung cancer and melanoma. Ann Oncol. 2018;29:271–9.

536 15. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of
537 tumors associated with local immune cytolytic activity. Cell. 2015;160:48–61.

538 16. Lakatos E, Williams MJ, Schenck RO, Cross WCH, Househam J, Werner B, et al.
539 Evolutionary dynamics of neoantigens in growing tumours. bioRxiv. 2019;536433.

540 17. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The Immune
541 Landscape of Cancer. Immunity. 2018;48:812-830.e14.

542 18. Wellenstein MD, de Visser KE. Cancer-Cell-Intrinsic Mechanisms Shaping the Tumor
543 Immune Landscape. Immunity. 2018;48:399–416.

544 19. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal
545 Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017;171:1029-1041 e21.

546 20. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, et al. Allele-
547 Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell. 2017;171:1259-
548 1271 e11.

549 21. Győrffy B, Bottai G, Fleischer T, Munkácsy G, Budczies J, Paladini L, et al. Aberrant DNA
550 methylation impacts gene expression and prognosis in breast cancer subtypes.
551 International Journal of Cancer. 2016;138:87–97.

552 22. Ye Q, Shen Y, Wang X, Yang J, Miao F, Shen C, et al. Hypermethylation of HLA class I
553 gene is associated with HLA class I down-regulation in human gastric cancer. Tissue
554 Antigens. 2010;75:30–9.

555 23. Network TCGAR. Comprehensive genomic characterization of squamous cell lung
556 cancers. Nature. 2012;489:519–25.

557 24. Wang L, Amoozgar Z, Huang J, Saleh MH, Xing D, Orsulic S, et al. Decitabine Enhances
558 Lymphocyte Migration and Function and Synergizes with CTLA-4 Blockade in a Murine
559 Ovarian Cancer Model. Cancer Immunol Res. 2015;3:1030–41.

560 25. Wang L-X, Mei Z-Y, Zhou J-H, Yao Y-S, Li Y-H, Xu Y-H, et al. Low dose decitabine
561 treatment induces CD80 expression in cancer cells and stimulates tumor specific
562 cytotoxic T lymphocyte responses. PLoS ONE. 2013;8:e62924.

563 26. Yang H, Bueso-Ramos C, DiNardo C, Estecio MR, Davanlou M, Geng Q-R, et al.
564 Expression of PD-L1, PD-L2, PD-1 and CTLA4 in myelodysplastic syndromes is enhanced
565 by treatment with hypomethylating agents. Leukemia. 2014;28:1280–8.

566 27. Saadi A, Shannon NB, Lao-Sirieix P, O'Donovan M, Walker E, Clemons NJ, et al. Stromal
567 genes discriminate preinvasive from invasive disease, predict outcome, and highlight
568 inflammatory pathways in digestive cancers. PNAS. 2010;107:2177–82.

569 28. Paz-Ares L, Luft A, Vicente D, Tafreshi A, Gümüş M, Mazières J, et al. Pembrolizumab
570 plus Chemotherapy for Squamous Non–Small-Cell Lung Cancer. New England Journal of
571 Medicine. 2018;379:2040–51.

572 29. Pawelczyk K, Piotrowska A, Ciesielska U, Jablonska K, Gletzel-Plucinska N, Grzegrzolka J,
573 et al. Role of PD-L1 Expression in Non-Small Cell Lung Cancer and Their Prognostic
574 Significance according to Clinicopathological Factors and Diagnostic Markers. Int J Mol
575 Sci. 2019;20.

576 30. Qi X, Li F, Wu Y, Cheng C, Han P, Wang J, et al. Optimization of 4-1BB antibody for
577 cancer immunotherapy by balancing agonistic strength with FcγR affinity. Nature
578 Communications. 2019;10:2141.

579 31. Melero I, Shuford WW, Newby SA, Aruffo A, Ledbetter JA, Hellström KE, et al.
580 Monoclonal antibodies against the 4-1BB T-cell activation molecule eradicate
581 established tumors. Nat Med. 1997;3:682–5.

582 32. Bartkowiak T, Curran MA. 4-1BB Agonists: Multi-Potent Potentiators of Tumor
583 Immunity. Front Oncol. 2015;5:117.

584 33. Segal NH, He AR, Doi T, Levy R, Bhatia S, Pishvaian MJ, et al. Phase I Study of Single-
585 Agent Utomilumab (PF-05082566), a 4-1BB/CD137 Agonist, in Patients with Advanced
586 Cancer. Clin Cancer Res. 2018;24:1816–23.

587 34. Zlotnik A, Yoshie O, Nomiyama H. The chemokine and chemokine receptor
588 superfamilies and their molecular evolution. Genome Biol. 2006;7:243.

589 35. Murakami T, Cardones AR, Finkelstein SE, Restifo NP, Klaunberg BA, Nestle FO, et al.
590 Immune Evasion by Murine Melanoma Mediated through CC Chemokine Receptor-10.
591 J Exp Med. 2003;198:1337–47.

592 36. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot
593 analysis: multitissue gene regulation in humans. Science. 2015;348:648–60.

594 37. Pipinikas CP, Kiropoulos TS, Teixeira VH, Brown JM, Varanou A, Falzon M, et al. Cell
595 migration leads to spatially distinct but clonally related airway cancer precursors.
596 Thorax. 2014;69:548–57.

597 38. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of Affymetrix GeneChip data
598 at the probe level. Bioinformatics. 2004;20:307–15.

599 39. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al.
600 ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics. 2014;30:428–30.

601 40. Jones D, Raine KM, Davies H, Tarpey PS, Butler AP, Teague JW, et al.
602 cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single
603 Nucleotide Variants in NGS Data. Curr Protoc Bioinformatics. 2016;56:15 10 1-15 10 18.

604 41. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to
605 detect break points of large deletions and medium sized insertions from paired-end
606 short reads. Bioinformatics. 2009;25:2865–71.

607 42. Raine KM, Hinton J, Butler AP, Teague JW, Davies H, Tarpey P, et al. cgpPindel:
608 Identifying Somatically Acquired Insertion and Deletion Events from Paired End
609 Sequencing. Curr Protoc Bioinformatics. 2015;52:15 7 1-12.

610 43. Raine KM, Van Loo P, Wedge DC, Jones D, Menzies A, Butler AP, et al. ascatNgs:
611 Identifying Somatically Acquired Copy-Number Alterations from Whole-Genome
612 Sequencing Data. Curr Protoc Bioinformatics. 2016;56:15 9 1-15 9 17.

613 44. Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, et al. RAG-mediated
614 recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1
615 acute lymphoblastic leukemia. Nat Genet. 2014;46:116–25.

616 45. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating
617 high-throughput genomic analysis with Bioconductor. Nature Methods. 2015;12:115–
618 21.

619 46. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using
620 lme4. Journal of Statistical Software. 2015;67:1–48.

621 47. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential
622 expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res.
623 2015;43:e47.

624 48. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing
625 under dependency. Ann Statist. 2001;29:1165–88.

626

Figure legends

Figure 1. Immune cell infiltration of lung carcinoma-in-situ lesions. a) Combined quantitative immunohistochemistry data of CD4, CD8 and FOXP3 staining (n=44; 28 progressive, 16 regressive) with total lymphocyte quantification from H&E images (n=112; 68 progressive, 44 regressive) shown. We observe increased lymphocytes (p=0.049) and CD8+ cells (p=0.055) per unit area of epithelium within regressive CIS lesions compared to progressive. Stromal regions adjacent to CIS lesions showed no significant differences in immune cells between progressive and regressive lesions. p-values are calculated using linear mixed effects models to account for samples from the same patient; $^{\#}p<0.1$, $^{*}p<0.05$. (b-c) Immunohistochemistry images of (b) progressive CIS lesion and (c) regressive CIS lesion with CD4+ T helper cells stained in brown, CD8+ cytotoxic T-cells in red and FOXP3+ T regulatory cells in blue. Immune cells are separately quantified within the CIS lesion and in the surrounding stroma.

Figure 2. Identification of immune ‘hot’ and ‘cold’ carcinoma in-situ lesions by immune cell clustering. Regressive lesions harbored significantly more infiltrating lymphocytes as assessed by multiplex immunohistochemistry (a; p=0.032 comparing percentage of all nucleated cells identified as T-cells (CD45+CD3+) or B-cells (CD45+CD3-CD20+) between 19 progressive and 9 regressive lesions). This finding was corroborated by molecular data in partially overlapping datasets; regressive lesions had higher gene-expression derived Tumor Infiltrating Lymphocyte (TIL) scores (b; p=0.0046; n=10 progressive, 8 regressive) and a higher proportion of immune cells as estimated from methylation data using methylCIBERSORT (c; p=0.0081; n=36 progressive, 18 regressive). d) Immune cell quantification from IHC data (n=28) shows an ‘immune cold’ cluster (left) in which most lesions progressed to cancer, and an ‘immune hot’ cluster (right) in which the majority regressed. Similar clustering patterns are seen in deconvoluted gene expression data (e; n=18) and on methylation-derived cell subtypes using methylCIBERSORT (f; n=54).

655 p-values are calculated using mixed effects models to account for samples from the same

656 patient.

657

Figure 3. Genomic aberrations affecting immune genes in lung carcinoma *in-situ* lesions. The mutational status is shown for 62 genes involved in the immune response, which are expressed by antigen presenting (tumor) cells. Genes are categorized as belonging to the Major Histocompatibility Complex (MHC) class I or II; stimulators (Stim) and inhibitors (Inhib) of the immune response, and genes involved in antigen processing (Ag-Proc). Mutations and copy number aberrations (CNAs) are shown for each of 29 progressive and 10 regressive samples. Loss of heterozygosity (LOH) events are shown as mutations to avoid confusion with copy number loss, relative to ploidy. The GXN PvR column displays the fold-change in expression of each gene between progressive and regressive samples, defined in a partially overlapping set of 18 samples. Significant genes, defined as False Discovery Rate < 0.05, are highlighted in blue. The TILcor column displays the Pearson's correlation coefficient between the expression of each gene and the gene-expression based tumour infiltrating lymphocyte (TIL) score, derived by the Danaher method. Progressive samples had more mutations (p=0.028) and CNAs (p=0.0038) than regressive in this gene set. dN/dS analysis identified *B2M, CHUK, KDR* and *CD80* as showing evidence of selection.

Figure 4. Immune escape mechanisms in CIS beyond antigen presentation. (a) Volcano plot of gene expression differential analysis of laser-captured stroma comparing progressive (n=10) and regressive (n=8) CIS samples. No genes were significant with FDR < 0.05 following adjustment for multiple testing. (b) Principle component analysis plot of the same 18 CIS samples, showing laser-captured epithelium and matched stroma. (c-d) RNA analysis of immunomodulatory molecules and cytokine:receptor pairs in n=18 CIS samples identified TNFSF9 and CCL27:CCR10 as significantly differentially expressed between progressive and regressive samples (p=0.0000058 and p=0.0000019 respectively). (e) Immunohistochemistry showed that TNFSF9 was similarly differentially expressed at the protein level (p=0.057; n=7 with successful staining). (f) Illustrative immunohistochemistry staining for TNFSF9. CCL27 and CCR10 showed a similar trend at the protein level to the

RNA level (e,g); whilst these data did not achieve a significance threshold (g; $p=0.49$ for CCL27:CCR10 ratio, $n=10$) we observe several outliers in the progressive group. Analysis of PD-L1 (encoded by CD274) and its receptor PD-1 (encoded by PDCD1) is included due to its relevance in clinical practice; again we do not achieve statistically significant results but do observe three marked outliers with PD-L1 expression >25%, all of which progressed to cancer. All p-values are calculated using linear mixed effects modeling to account for samples from the same patient; ***$p < 0.001$ **$p < 0.01$ *$p<0.05$ #$p<0.1$. Units for gene expression figures represent normalised microarray intensity values.