Context-dependent effects of *CDKN2A* and other 9p21 gene losses during the evolution of oesophageal cancer

Piyali Ganguli[1,2], Celia C. Basanta[1,2#], Akram Mendez[1,2#], Maria Armero[1,2], Aeman Zahra[1,2], Amelia A. Sagrado[1,2], Hrvoje Misetic[1,2], Ginny Devonshire[3], Gavin Kelly[4], Adam Freeman[3], Mary Green[5], Emma Nye[5], Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium[*], Manuel Rodriguez-Justo[6], Jo Spencer[7], Rebecca C. Fitzgerald[3], Francesca D. Ciccarelli[1,2**]

[1] Cancer Systems Biology Laboratory, The Francis Crick Institute, London NW1 1AT, UK

[2] Bart's Cancer Institute - Centre for Cancer Genomics & Computational Biology, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK

[3] Early Cancer Institute, Hutchison Research Centre, University of Cambridge, CB2 0XZ, UK

[4] Bioinformatics & Biostatistics STP, The Francis Crick Institute, London NW1 1AT, UK

[5] Experimental Histopathology STP, The Francis Crick Institute, London NW1 1AT, UK

[6] Department of Pathology, UCL Cancer Institute, London WC1E 6HX, UK

[7] School of Immunology and Microbial Sciences, King's College London; London WC2R 2LS, UK.

# = contributed equally

* = full list provided at the end of the manuscript

** = corresponding author (francesca.ciccarelli@crick.ac.uk)

ABSTRACT

CDKN2A is a tumour suppressor gene located in chromosome 9p21 and frequently lost in Barrett's oesophagus (BO) and oesophageal adenocarcinoma (OAC). How *CDKN2A* and other 9p21 gene co-deletions affect OAC evolution remains understudied. We explored the effects of 9p21 loss in OACs and cancer progressor and non-progressor BOs with matched genomic, transcriptomic, and clinical data. Despite its cancer driver role, *CDKN2A* loss in BO prevents OAC initiation by counter-selecting acquisition of *TP53* alterations. 9p21 gene co-deletions predict poor patient survival in OAC but not BO through context-dependent effects on cell cycle, oxidative phosphorylation, and interferon response. Immune quantifications using bulk transcriptome, RNAscope and high-dimensional tissue imaging showed that *IFNE* loss reduces immune infiltration in BO but not OAC. Mechanistically, *CDKN2A* loss suppresses the maintenance of squamous epithelium, contributing to a more aggressive phenotype. Our study demonstrates context-dependent roles of cancer genes during disease evolution, with consequences for cancer detection and patient management.

BACKGROUND

CDKN2A is among the most frequently damaged tumour suppressor genes in human cancer, with loss of function (LoF) reported in at least 35 different cancer types across 12 organ systems[1]. *CDKN2A* acts as a tumour suppressor by inducing cell cycle arrest[2] and cellular senescence[3] and preventing angiogenesis[4], oxidative stress[5], and metastasis[6]. Loss of *CDKN2A* is also associated with poor patient survival[7-9].

CDKN2A LoF may occur through damaging point mutations, small indels or large deletions of chromosome 9p21.3 locus (hereon 9p21), an event observed in around 15% of human cancers[10]. Depending on their length, deletions may involve up to 25 additional protein coding genes flanking *CDKN2A*, including other cell cycle regulators (*CDKN2B* and *KLHL9*), a metabolic enzyme (*MTAP*), and a cluster of 16 type-I interferon genes (**Figure 1A**). Recent studies have investigated the cancer role of other 9p21 genes in addition to *CDKN2A* and the loss of the whole locus, rather than *CDKN2A* alone, has been associated with poor survival and resistance to immunotherapy, possibly through the onset of an immune-cold tumour microenvironment (TME)[10].

Dissecting the consequences of individual 9p21 gene losses is not straightforward because of their co-occurrence. Recently, the induction of 9p21 deletions of different lengths in murine models of pancreatic cancer enabled observation of reduced $CD8^+$ T cell infiltration only when the IFN cluster was co-deleted with *CDKN2A*, *CDKN2B* and *MTAP*[11]. *IFNE*, one of 9p21 type-I interferon genes (**Figure 1A**), is a tumour suppressor in ovarian cancer due to its antiproliferative and apoptotic effects[12]. Moreover, IFNE treatment promotes $CD8^+$ T cell activation and reduces T regulatory cells (Tregs) and myeloid-derived suppressor cells (MDSCs)[12]. *MTAP* can also regulate $CD8^+$ and $CD4^+$ T cell infiltration in melanoma

mouse models by controlling methylthioadenosine accumulation in their TME[13]. All these studies have started to unveil that at least some of the effects previously ascribed to *CDKN2A* LoF are in fact due to the loss of other 9p21 genes.

CDKN2A LoF has long been known as an early event in the evolution of human oesophageal adenocarcinoma (OAC), occurring already in its metaplastic precursor Barrett's Oesophagus (BO)[14-17]. Consequently, *CDKN2A* LoF has been proposed to drive OAC initiation by favouring BO clonal selective sweeps and subsequent alterations of additional drivers, most frequently *TP53*[18-21]. Recently, this model has been challenged in favour of an alternative one where an early *TP53* LoF would enable whole genome doubling with consequent acquisition of additional oncogenic drivers[22,23]. The uncertainty about OAC initiation raises questions on the role of *CDKN2A* in BO and OAC evolution. Moreover, very little is known about the function of the remaining 9p21 genes.

Here we investigated how the loss of *CDKN2A* and other 9p21 genes affects OAC initiation and progression. We compared genomic, transcriptomic, and survival data from large and clinically annotated cohorts of OAC and BO patients who progressed or did not progress to cancer. We studied the effect of 9p21 loss on BO and OAC TME by quantifying immune infiltrates from bulk transcriptomic data and high-dimensional tissue profiling coupled with RNAscope. Finally, we rebuilt the causal gene regulatory networks linking *CDKN2A* gene loss to specific downstream functional effects. Our study showed the same genetic alterations of *CDKN2A* and other 9p21 genes have different effects in different contexts and stages of OAC evolution, with possible implications in future patient management.

RESULTS

CDKN2A LoF drives BO and OAC evolution but does not trigger OAC initiation.

We collected whole genome (WGS), whole exome (WES) and gene panel sequencing data for 1032 OACs sequenced *de novo* as part of the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium or collected from the literature[8,24-30] (**Table S1**). To ensure consistency, we annotated mutations and copy number variations in all datasets with the same approach (**Methods, Figure S1**) and identified the damaged drivers in each sample using a curated list of 54 known (canonical) OAC drivers (**Table S2**). In agreement with previous studies[24,31,32], *CDKN2A* was the second most frequently damaged gene (**Figure 1B**), with LoF alterations in 25% of OACs via either homozygous deletions (15%) or damaging mutations (10%). Interestingly, almost 56% of OACs (90% considering also *TP53*) had damaging alterations in cell cycle regulators (**Figure 1C, Table S2**), suggesting that cell cycle disruption is key in OAC evolution but does not always involve *CDKN2A*.

Next, we measured the frequency of *CDKN2A* LoF in 257 BOs that progressed to high-grade dysplasia or OAC (P-BOs), again sequenced for this study or gathered from published datasets[17,33-35] (**Table S1**). *CDKN2A* LoF occurred significantly more frequently in P-BO than OAC ($p = 0.01$, two-sided Fisher's exact test, **Figure 1D**), suggesting that OAC does not always originate from a *CDKN2A*-damaged BO. To further investigate this, we analysed 66 matched OAC-BO pairs with *CDKN2A* LoF in BO or OAC (**Table S1**). Only 15 matched lesions had either identical or clonally related *CDKN2A* alterations (**Figure 1E**), confirming that *CDKN2A* LoF is not required for precancer to cancer transition. Interestingly, 28 OACs lost *CDKN2A* independently of the paired BOs (**Figure 1E**), indicating that either OAC developed from a different

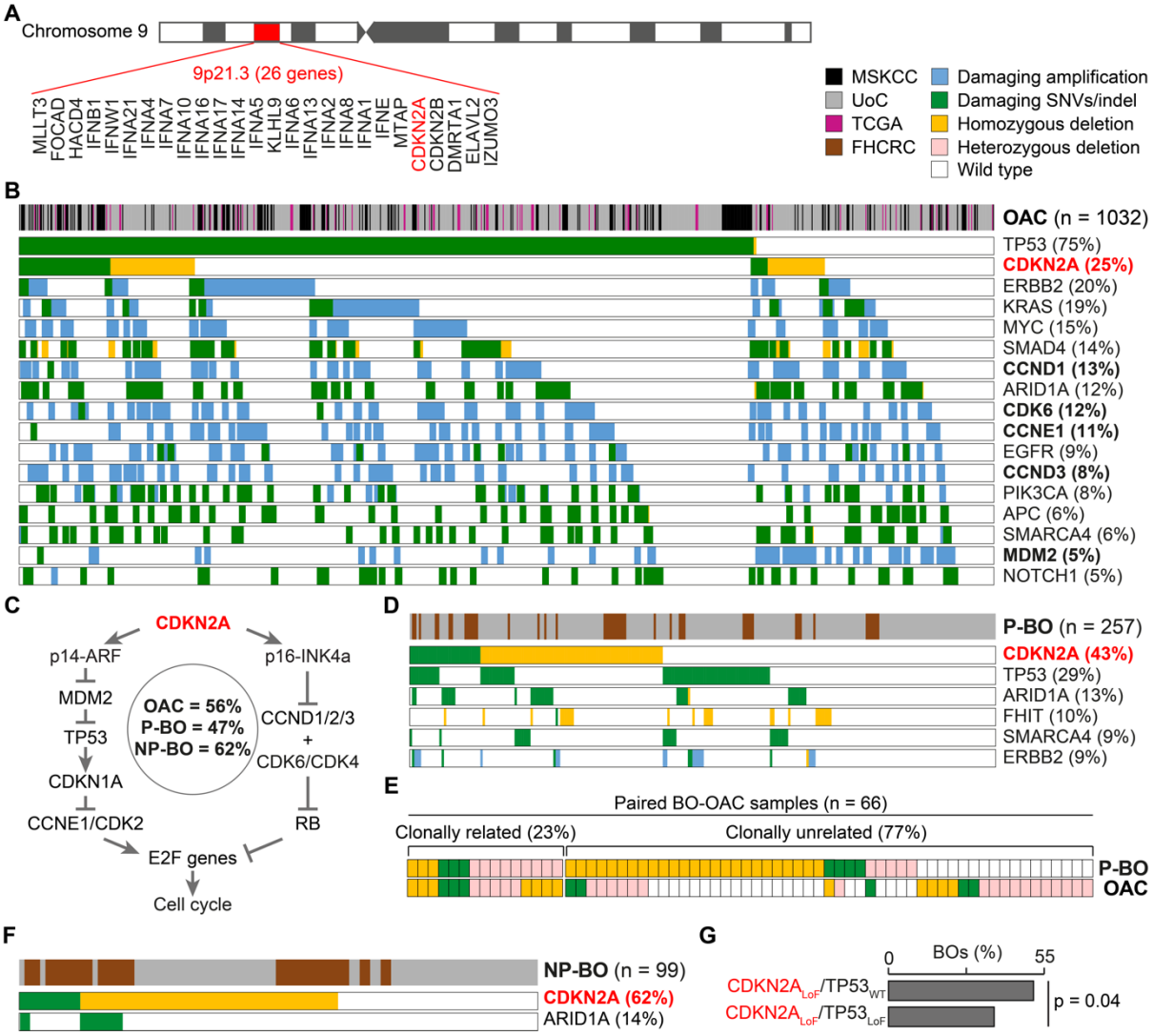
CDKN2A-damaged BO clone or *CDKN2A* LoF was acquired after BO transformed into OAC.

Finally, we analysed 99 BOs that did not progress to high-grade dysplasia or OAC (NP-BOs)[17,34,35] (**Table S1**). The frequency of *CDKN2A* LoF in NP-BO was even higher than P-BO and OAC ($p = 2x10^{-3}$ and $p = 2x10^{-7}$, respectively, two-sided Fisher's exact test, **Figure 1F**). Moreover, while in OAC the dysregulation of cell cycle could occur through alterations of other genes, *CDKN2A* was the only cell cycle gene damaged in BO (**Figure 1B**). Therefore, unlike OAC, it is specifically *CDKN2A* LoF to drive BO.

As observed previously[23,36], P-BOs had significantly more damaged drivers than NP-BOs ($p = 0.001$, two-sided Fisher's exact test, **Table S2**), indicating that OAC initiation requires additional events, most frequently *TP53* complete loss. We therefore used *TP53* LoF to assess the role of *CDKN2A* LoF in OAC initiation. Specifically, we calculated the odds of cancer progression based on the mutational status of *CDKN2A* and *TP53* in BO. As expected, the odds of cancer progression in BOs with *TP53* LoF was 1 irrespective of *CDKN2A* status (**Table S3**), confirming that *TP53* is a strong driver of OAC initiation. However, the odds of cancer progression in BOs with *CDKN2A* LoF and wild type *TP53* was lower than those of BOs with both wild type genes (0.58 and 0.72, respectively, **Table S3**). This suggested that an early occurrence of *CDKN2A* LoF in BO prevents OAC initiation. To test this further, we compared two logistic regression models, one assuming a role in OAC initiation only for *TP53* LoF (model 1) and the other for both *TP53* and *CDKN2A* LoFs (model 2, **Methods**). Model 2 was a significantly better predictor of OAC initiation than model 1 ($p = 0.01$, ANOVA test), with expected occurrences of P-BOs with any status of *TP53* and *CDKN2A* perfectly matching the observed occurrences (**Table S3**). The negative β coefficient

of *CDKN2A* in model 2 further confirmed an OAC-preventive role of *CDKN2A* LoF, with almost 50% reduced risk of progression (**Methods, Table S3**). This was likely a consequence of the negative selection acting on BO cells with *CDKN2A* LoF to lose also *TP53*, as suggested by the significantly lower proportion of BOs carrying both alterations (p = 0.04, two-sided Fisher's exact test, **Figure 1G**).

Figure 1. *CDKN2A* LoF occurrence in BO and OAC.



A. Gene composition of chromosome 9p21 locus.

B. Canonical OAC drivers damaged in at least 5% of 1032 OACs. All cell cycle regulators are reported in bold.

C. Alterations in cell cycle regulators in BO and OAC. *CDKN2A* gene products (p14-ARF and p16-INK4a) regulate the cell cycle through the E2F genes[2]. p14-ARF blocks MDM2 and TP53 degradation, which induces *CDKN1A* transcription. CDKN1A in turn inhibits the CCNE1/CDK2 complex ultimately blocking cell cycle through E2F1 inhibition. p16-INK4a directly inhibits the CCND/CDK6/CDK4 complex preventing RB1 phosphorylation. Unphosphorylated RB1 can bind E2F1 leading to cell cycle arrest.

CDKN2A LoF favours cell cycle progression resulting in uncontrolled cell proliferation. Values within the circle represent the proportion of OACs, P-BOs and NP-BOs with at least one damaged cell cycle regulator (except TP53).

D. Canonical OAC drivers damaged in at least 5% of 257 P-BOs.

E. Paired BO-OACs with *CDKN2A* LoF. Clonally related alterations refer to either identical alterations in both lesions or alterations in BO that could further evolve in OAC.

F. Canonical drivers damaged in at least 5% of 99 NP-BO. Alteration frequency of OAC canonical drivers in (**B**, **D** and **F**) is indicated in brackets. The alteration frequency of all OAC drivers in the three cohorts is available in **Table S2**.

G. Frequency of *CDKN2A* LoF in 356 BOs (257 P-BOs and 99 NP-BOs) with or without *TP53* LoF. Statistical significance was assessed using a two-sided Fisher's exact test.

FHCRC, Fred Hutchinson Cancer Research Center; LoF, loss of function; MSKCC, Memorial Sloan Kettering Cancer Center; NP-BO, non-progressor Barrett's oesophagus; OAC, oesophageal adenocarcinoma; P-BO, progressor Barrett's oesophagus; SNV, single nucleotide variant; TCGA, The Cancer Genome Atlas; UoC, University of Cambridge.

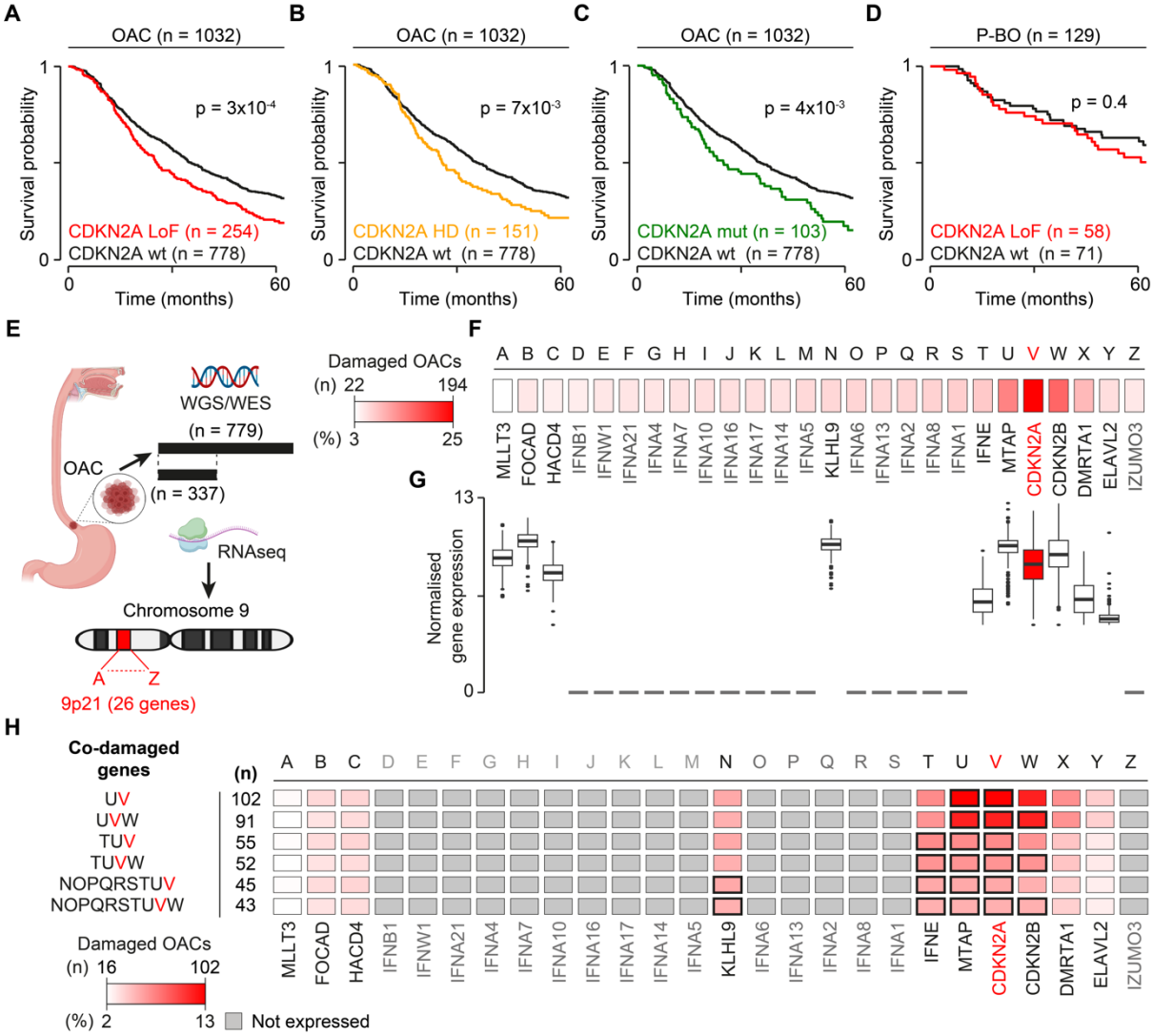
LoF of 9p21 genes predicts poor survival in OAC but not in BO.

Next, we investigated the impact of *CDKN2A* LoF on OAC patient survival. In keeping with previous reports[7,8], OAC patients with *CDKN2A* LoF showed significantly worse survival than those with the wild type gene (**Figure 2A**). This held true even when patients with *CDKN2A* homozygous deletions (**Figure 2B**) or damaging point mutations (**Figure 2C**) were considered separately. However, we did not observe lower survival in patients with *CDKN2A* heterozygous deletions (**Figure S2A**), indicating that *CDKN2A* complete loss is required. Interestingly, damaging alterations in *TP53* or other cell cycle regulators had no effect on survival (**Figure S2B-F**) despite their frequent OAC alterations (**Figure 1C**). Therefore, the survival effect of *CDKN2A* LoF does not depend on its function as cell cycle regulator. Moreover, *CDKN2A* LoF was not a predictor of worse survival in P-BO (**Figure 2D**), suggesting context-dependent consequences of its loss.

Although *CDKN2A* LoF through mutations alone was enough to predict poor survival (**Figure 2C**), we investigated whether the co-occurring loss of other 9p21 genes could also contribute. We restricted this analysis to 779 OACs with WGS or WES data (**Figure 2E**) because not all 9p21 genes were included in other gene panels. Although *CDKN2A* was the most frequently occurring alteration in the locus, confirming that it is the event under positive selection, the other 25 genes were frequently co-lost with it (**Figure 2F**). However, only ten 9p21 genes were expressed in OAC (**Figure 2G**) or normal oesophagus (**Figure S3**), indicating that the loss of the remaining 16 genes likely had no functional consequences. We therefore tested the potential impact on survival of the ten 9p21 expressed genes by dividing OAC patients into nine groups. Each of these represented at least 5% of the cohort and was composed of patients with the same 9p21 mutation and copy number profile (**Table**

S4). We then compared the survival of patients in each of the nine groups to that of 414 OAC patients with a wild type 9p21 locus. Patients in six groups had worse survival than 9p21 wild type patients (FDR <0.1, **Figure 2H, Table S4**). All of them lost *KLHL9, IFNE*, *MTAP*, *CDKN2A* and *CDKN2B* (**Figure 2H**), indicating that alterations in these genes may contribute to poor prognosis.

Figure 2. Effect of the LoF of *CDKN2A* and other 9p21 genes on survival



Kaplan-Meier survival curves of OAC patients with wild type *CDKN2A* compared to those with all types of LoFs (**A**), only homozygous deletions (**B**), and only LoF mutations (**C**).

D. Kaplan-Meier survival curves of P-BO patients with and without *CDKN2A* LoF.

E. Approach to test the effect of the co-damage in 9p21 genes on survival of OAC patients. Only 779 OACs with WGS or WES data were used for the survival analysis while 337 of them with RNAseq data were used to measure 9p21 gene expression. Letters correspond to the 26 genes according to their order in the chromosomal locus.

F. LoF frequency of 9p21 genes in 779 OACs.

G. Distribution of normalised expression values in the 9p21 genes in 337 OACs.

H. Kaplan-Meier survival analysis of OAC patients with co-alterations in the ten expressed 9p21 genes and 414 OAC patients with a wild type locus. Only groups with significantly poor survival (FDR <0.1) are shown and genes of interest are circled in black. All groups used in the analysis are listed in **Table S4**.

The minimum and maximum number and percent of damaged OACs in (**F**) and (**H**) are reported in the corresponding heatmap.

HD, homozygous deletion; LoF, loss-of-function; OAC, oesophageal adenocarcinoma; P-BO, progressor Barrett's oesophagus; WES, whole exon sequencing; WGS, whole genome sequencing; WT, wild type. Cartoon in **E** was created with BioRender.com.

LoF of 9p21 genes has distinct functional consequences in BO and OAC

Our results suggested that the LoFs of *CDKN2A* and other 9p21 genes have functional and survival consequences that depend on the context. Disentangling these context-dependent effects is challenging because 9p21 genes are often co-damaged (**Figure 2F**). To tease out the contribution of individual 9p21 genes, we divided 22 NP-BOs, 108 P-BOs and 337 OACs with matched genomic and transcriptomic data (**Table S1**) into four groups. Each of these had the same mutation and copy number profile of the five genes (*KLHL9, IFNE*, *MTAP*, *CDKN2A* and *CDKN2B*) whose loss impacted survival (**Figure 2H**). Group 1 included all samples with *CDKN2A* LoF independently on the status of the other genes, closely resembling the cohorts tested in the survival analysis (**Figure 2A-D**). The other three groups were subsets of group 1 with the highest frequency of LoFs in some of the five genes (**Figure 3B**). Namely, samples in group 2 had the highest possible LoF frequency in all five genes; those in group 3 had *CDKN2B, CDKN2A, and MTAP* LoF but were wild type for *IFNE* and *KLHL9*; and those in group 4 had only *CDKN2A* LoF but were wild type for the other four genes (**Figure 3A**).

We then identified the dysregulated biological processes in each group as compared to the corresponding 9p21 wild type group by performing a pre-ranked gene set enrichment analysis (GSEA)[37] in NP-BOs, P-BOs, and OAC separately. Overall, we detected 72, 53 and 28 unique pathways significantly dysregulated (FDR ≤0.01) in NP-BO, P-BO and OAC, respectively (**Table S5**). More than 85% of these pathways mapped to only five biological processes, namely cell cycle regulation, metabolism, immune response, signal transduction, and development. Overall NP-BO and P-BO showed a higher fraction of dysregulated pathways than OAC (**Figure 3C**), indicating that 9p21 LoF had higher impact in pre-malignant conditions.

We found an extensive dysregulation of cell cycle in NP-BO and P-BO but not OAC (**Table S5, Figure 3D-F**). This was consistent with the pervasive alteration of cell cycle regulators independently on *CDKN2A* in OAC but not BO (**Figure 1C**). Interestingly, group 4 (*CDKN2A* LoF only) did not show significant cell cycle alterations compared to wild type samples, indicating that the co-deletion of *KLHL9, CDKN2A* and *CDKN2B* maximises the effect.

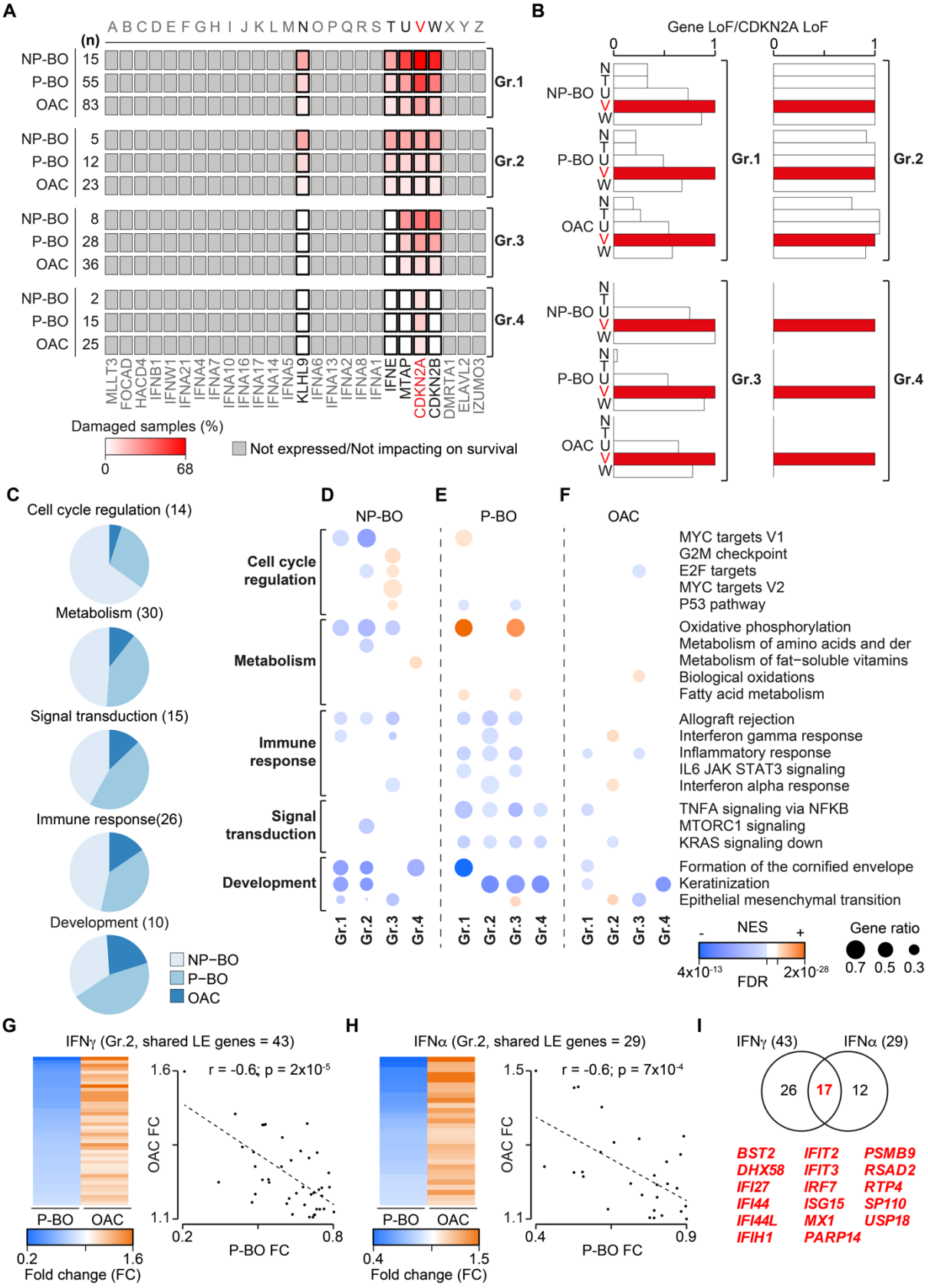
CDKN2A LoF alone was also not sufficient to modify metabolic or immune functions (**Table S5, Figure 3D-F**). Disruptions of the former were likely due to *MTAP*, given its role in metabolism[38] and metabolic reprogramming[39] and *IFNE*, since IFNs can activate an immune response through metabolic regulation[40]. Interestingly, oxidative phosphorylation was consistently downregulated in NP-BO, upregulated in P-BO, and showed no difference in OAC (**Table S5, Figure 3D-F**). This once again suggested that the same genetic alteration may trigger different functional responses depending on the context.

Similarly, the disruption of immune pathways differed between BO and OAC (**Table S5, Figure 3D-F**). While interferon alpha and gamma responses were consistently downregulated in NP-BO and P-BO, both were upregulated in OAC, particularly in group 2 patients (**Figure 3B**). Consistently, we observed a significant inverse correlation between expression fold changes of interferon gamma (**Figure 3G**) and alpha (**Figure 3H**) genes in BO and OAC groups 2 compared to 9p21 wild type samples. Moreover, there was substantial overlap between the altered genes in the two pathways (**Figure 3I**), suggesting a comprehensive transcriptional reprogramming. The most likely candidates responsible for this reprogramming were again *MTAP*, given its recently reported ability to regulate the TME[13], and *IFNE*, a type-1 interferon expressed in adult epithelia[41]. Since the effect was most visible in

group 2, which had LoF in both genes, but not in group 3, which had *MTAP* LoF and *IFNE* wild type (**Figure 3C**), the effect on interferon response was likely due to *IFNE* loss.

CDKN2A LoF played instead a direct role in signal transduction, contributing to the downregulation of TNFA and KRAS signalling, and in the pervasive downregulation of keratinisation genes (**Table S5, Figure 3D-F**).

Figure 3. Functional consequences of 9p21 gene LoF in BO and OAC.



A. Frequency of damaged and damaged 9p21 genes in the four groups of NP-BOs, P-BOs estimated over 22 NP-BOs, 108 P-BOs and 337 OACs with matched genomic and transcriptomic data.

B. Proportions of samples with LoF in *KLHL9* (N), *IFNE* (T), *MTAP* (U), and *CDKN2B* (W) over samples with *CDKN2A LoF* (V) in each group of NP-BOs, P-BOs and OACs. The number of samples in each group and condition is reported.

C. Relative proportion of dysregulated pathways in NP-BO, P-BO and OAC cohorts mapping to cell cycle regulation, metabolism, signal transduction, immune response, and development. Number in brackets represent the number of unique pathways.

Results of pre-ranked GSEA[37] showing the NES, FDR and gene ratio (number of leading-edge genes over the total expressed genes) of pathways dysregulated in each group of NP-BOs **(D),** P-BOs **(E)** and OACs **(F)**. NES >0 indicates pathway upregulation while NES <0 indicates downregulation.

Fold change of expression and correlation plot of the shared leading-edge genes of interferon gamma **(G)** and alpha **(H)** response pathways enriched in P-BO and OAC group 2 as compared to 9p21 wild type samples. Pearson's correlation coefficients and associated p-values are reported for both pathways.

I. Overlap of leading-edge genes between interferon gamma and alpha response pathways enriched in P-BO and OAC group 2. The 17 shared genes are listed.

FC, Fold change; FDR, False Discovery Rate; LoF, loss of function; NES, Normalised enrichment score; NP-BO, non- progressor Barrett's oesophagus; OAC, oesophageal adenocarcinoma; P-BO, progressor Barrett's oesophagus.

Loss of *IFNE* reduces immune infiltration in BO but not OAC.

To further investigate the opposite effect of *IFNE* on interferon alpha and gamma response in BO and OAC (**Figure 3D-F**), we quantified the infiltration of 18 immune cell populations in NP-BOs, P-BOs and OACs from their bulk transcriptomic data. We then compared the abundance of immune infiltrates between each of the four 9p21 LoF groups (**Figure 3A**) and the corresponding 9p21 wild type samples.

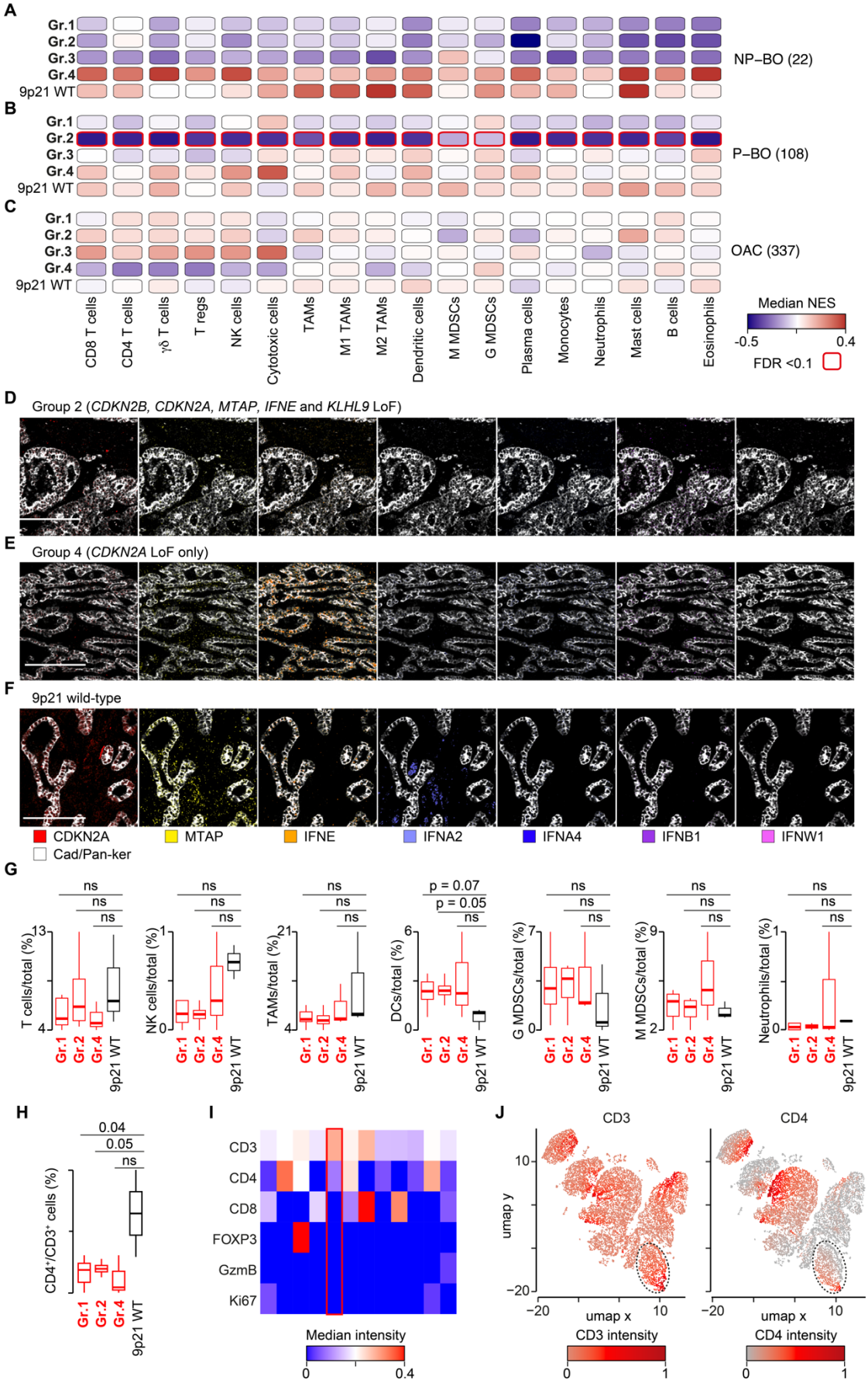
Groups 1, 2 and 3 showed overall depletion of immune infiltrates in NP-BO and P-BO (**Figure 4A, B, Table S6**). The signal was higher in group 2 where the impact of *IFNE* LoF was more appreciable (**Figure 3B**), again suggesting that the immune depletion is a consequence of *IFNE* loss. This was consistent with recent observations of a cold TME when *IFNE*[12] or the whole IFN locus[11] are lost in melanoma ovarian, or pancreatic cancers (**Table S7**). However, the same studies also reported an increased infiltration of T regs, MDSCs and B cells (**Table S7**) that we did not observe (**Figure 4A, B**). The TME of group 4 (*CDKN2A* LoF only) was similar to that of 9p21 wild type samples in both NP-BO and P-BO, confirming that *CDKN2A* LoF does not directly interfere with the immune system.

The most striking difference with other cancer types (**Table S7**) and BO (**Figure 4A, B**) was the absence of any significant TME difference between 9p21 LoF and wild type OACs (**Figure 4C, Table S6**). To investigate this at higher resolution, we performed high-dimensional imaging mass cytometry (IMC) on tissue sections representative of group 1, group 2, group 4 and 9p21 wild type OACs (**Table S8**). We used a panel of 26 antibodies targeting structural, immune, and 9p21-encoded proteins as well as RNAscope probes against *IFNE* and *IFNB1* mRNAs to increase the detection signal (**Table S9**). We confirmed that group 2 lost the expression of all 9p21 encoded proteins in the tumour, while group 4 lost CDKN2A only compared to

9p21 wild type OACs (**Figure 4D-F**). Moreover, IFNE was the only interferon clearly expressed in OAC epithelium (**Figure 4D-F**).

We performed single cell segmentation of the IMC images to quantify T cells, NK cells, macrophages, dendritic cells, monocytic (M) and granulocytic (G) MDSCs, and neutrophils (**Methods**). We then compared the relative abundance of each immune population over all cells in each slide across OAC groups. We confirmed no significant difference in immune infiltration between 9p21 LoF and wild type OACs, except for a borderline significant enrichment in dendritic cells in groups 1 and 2 (**Figure 4G**). We further applied unsupervised clustering to T cells and macrophages, for which we had multiple markers (**Table S9**), to test whether there was any difference in specific subpopulations. Again, we detected no major differences in any subpopulations of macrophages or T cells, except a borderline significant depletion of $CD4^+$ T cells in groups 1 and 2 compared to 9p21 wild type OAC (**Figure 4H-J**). These results confirmed that, unlike BO, the loss of *IFNE* or any other 9p21 genes does not lead to any major difference in the TME of OAC.

Figure 4. Impact of 9p21 gene loss on immune infiltration in BO and OAC.



Comparison of NES scores of 18 immune populations between 9p21 LoF and wild type samples in NP-BOs **(A)**, P-BOs **(B)** and OACs **(C)**. NES distributions were compared using a two-sided Wilcoxon's rank sum test and corrected for multiple testing using the Benjamini–Hochberg method. Number of samples are reported in brackets. Immune populations with significant differences (FDR < 0.1) are circled.

Representative IMC images from group 2 **(D)**, group 4 **(E)** and 9p21 wild type **(F)** OACs showing the expression of 9p21 targeted proteins and mRNAs. Cadherin-1 and pan-keratin denote tumour. Scale bar: 200 μm.

G. Relative abundance of immune cells over all cells in 9p21 LoF and wild type OACs. Samples in groups 2 and 4 were pooled together to form group 1 (n = 7). Distributions were compared using a two-sided Wilcoxon rank sum test.

H. Relative abundance of $CD4^+$ cells over all $CD3^+$ cells in 9p21 LoF and wild-type OACs. Distributions were compared using a two-sided Wilcoxon rank sum test.

I. Median marker intensity across the T cell clusters at a clustering resolution of 0.5.

J. UMAP map of 9750 T cells in ten OACs. Cells were grouped in 12 clusters based on the expression of six markers and coloured according to the mean intensities of CD3 and CD4. The cluster enriched in group 1 is circled.

FDR, False Discovery Rate; LoF, loss of function; NES, normalised enrichment score; NP-BO, non-progressor Barrett's oesophagus; P-BO, progressor Barrett's oesophagus; OAC, oesophageal adenocarcinoma.

***CDKN2A* LoF favours squamous to columnar epithelium transition.**

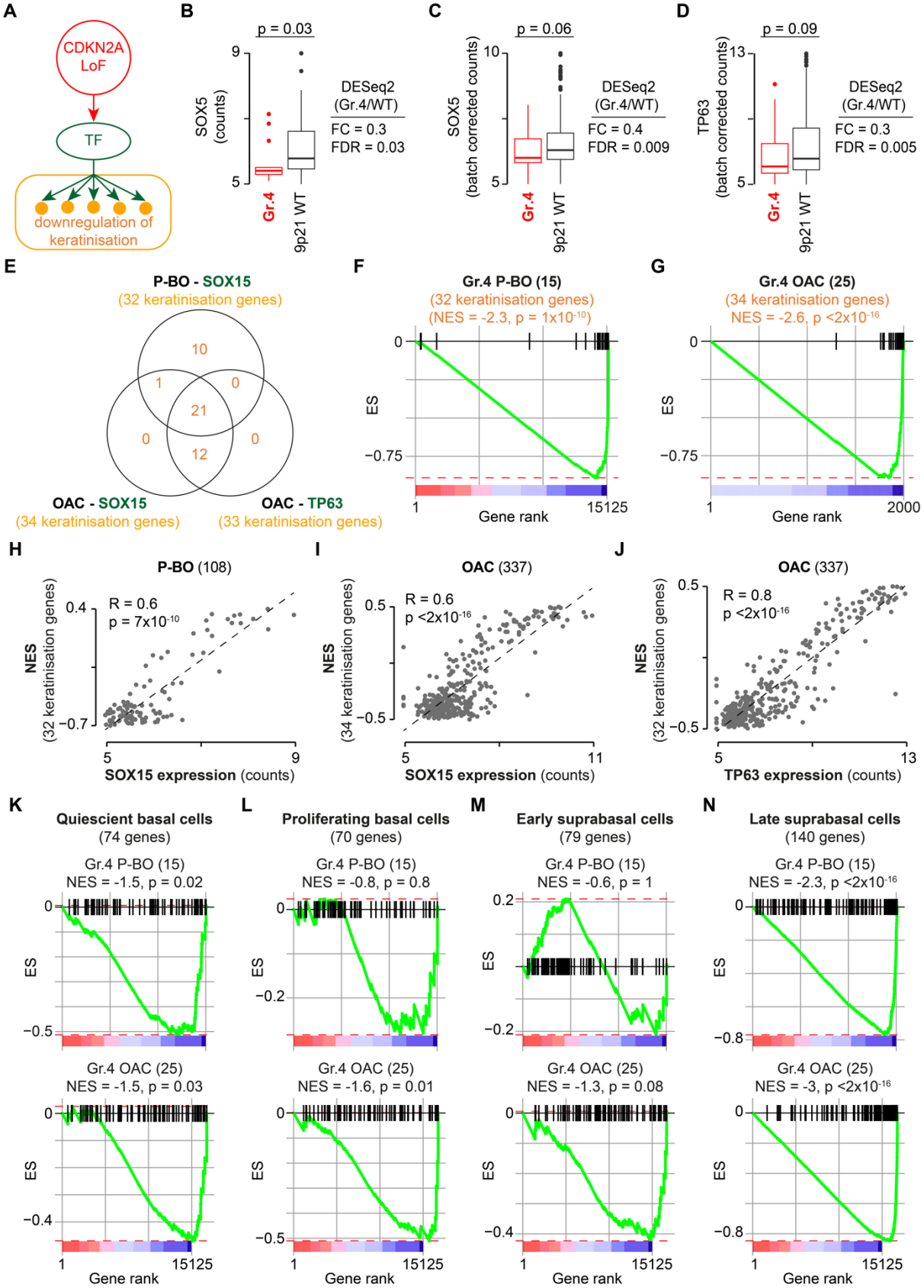
We observed a pervasive downregulation of processes responsible for terminal differentiation of keratinocytes, such as keratinisation and formation of the cornified envelope, across all 9p21 LoF groups **(Figure 3E-F**). In particular, P-BO and OAC groups 4 were associated with the downregulation of keratinisation, suggesting that *CD2KNA* LoF alone was sufficient for triggering this process. To gain further mechanistic insights, we rebuilt the gene regulatory network linking *CD2KNA* LoF to keratinisation in P-BO and OAC group 4 (**Figure 5A**). We applied a three-step protocol to (1) identify co-regulated keratinisation gene modules and associated transcription factors (TFs) in P-BO and OAC; (2) predict the causal models linking *CDKN2A* LoF directly to keratinisation gene downregulation through the perturbation of specific TFs; and (3) retain dysregulated causal models with robust experimental support (**Figure S4, Methods**).

The gene regulatory network analysis returned 17 and five causal models in P-BO and OAC, respectively, associated with only two TFs: SOX15 in P-BO and SOX15 and TP63 in OAC (**Table S10**). We further confirmed that these TFs were significantly downregulated in P-BO (**Figure 5B**) and OAC (**Figure 5C, D**) groups 4 as compared to 9p21 wild type samples. Overall, the gene modules controlled by SOX15 and TP63 included 44 keratinisation genes (**Table S10**), 21 (48%) of which were shared across all gene modules and all but one were shared between SOX15 and TP63 (**Figure 5E**). Therefore, the downregulation these two TFs in *CDKN2A* LoF samples led to a comprehensive downregulation of the keratinisation transcriptional programme, as confirmed by a pre-ranked GSEA[37] using a keratinisation gene-derived signature in P-BO (**Figure 5F**) and OAC (**Figure 5G**). Moreover, *SOX15* and *TP63* gene expression

was positively correlated with the enrichment score of the keratinisation genes **(Figures 5H-J)**, again confirming that the two TFs control their expression.

SOX15 regulates transcription of a large number of genes specific to oesophageal epithelium[42] and TP63 is essential for development and maintenance of all stratified epithelia[43]. The transition from oesophageal squamous epithelium to intestinal columnar epithelium is a key feature in the initiation of BO and OAC[44]. Our data indicated that *CDKN2A* LoF leads to a downregulation of the transcriptional programme responsible for the maintenance of the squamous epithelium more robust and persistent than in *CDKN2A* wild type samples. To further test the link between *CDKN2A* LoF and suppression of squamous epithelium, we performed pre-ranked GSEA[37] using four independent gene signatures characteristic of cells composing the oesophageal epithelium, namely quiescent basal cells, proliferating basal cells, early supra-basal cells and late supra-basal cells[45]. We observed global downregulation of all four signatures in OAC and quiescent basal cells and late supra-basal cells in P-BO (**Figure 5K-N**). These results confirmed that *CDKN2A* LoF exacerbates a phenotype typical of OAC and this may contribute to more aggressive tumours.

Figure 5. Impact of *CDKN2A* LoF on epithelium differentiation in P-BO and OAC.



A. Gene regulatory network linking *CDKN2A* LoF to the downregulation of keratinisation genes through TF deregulations.

Distributions of gene expression values of *SOX15* in P-BO **(B)** and *SOX15* **(C)** and *TP63* **(D)** in OAC in group 4 and 9p21 wild-type samples. Distributions were compared using two-sided Wilcoxon's rank sum test. FC and FDR from the differential gene expression analysis with DESeq2[46] are also shown.

E. Overlap between keratinisation genes targeted by *SOX15* and *TP63* in P-BO and OAC.

Pre-ranked GSEA plots using as signatures keratinisation genes targeted by *SOX15* in P-BO **(F)** and by *SOX15* and *TP53* in OAC **(G)**. Genes were ranked from the most upregulated to the most downregulated in group 4 compared to 9p21 wild type samples. For OAC, only the top 2000 downregulated genes are shown.

Correlation plots between keratinisation GSEA NES and the gene expression values of *SOX15* in P-BO **(H)** and *SOX15* **(I)** and *TP63* **(J)** in OAC. Spearman's correlation coefficient and associated p-values are shown.

Pre-ranked GSEA plots using gene signatures for quiescent basal cells **(K)**, proliferating basal cells **(L)**, early supra-basal cells **(M)** and late supra-basal cells **(N)** in P-BO and OAC group 4.

LoF, loss of function; NES, normalised enrichment score; OAC, oesophageal adenocarcinoma; P-BO, progressor Barrett's oesophagus; TF, transcription factor; GSEA, gene set enrichment analysis.

DISCUSSION

In this study, we dissected the role of *CDKN2A* and other 9p21 genes in OAC evolution, from the transformation of premalignant BO to the impact on patient survival.

Despite being an OAC driver, the early loss of *CDKN2A* has a tumour suppressive role since it occurs at a significantly higher frequency in NP-BO than P-BO and OAC. This is consistent with other drivers whose alterations are more frequent in normal tissues than cancer, including *ERBB2*, *ERBB3*, *KRAS*, and *NOTCH1*[47]. The anti-tumorigenic function of *NOTCH1* is exerted through an increased fitness of *NOTCH1* mutant cells that are able to outcompete early tumours[48]. For *CDKN2A* we propose a different mechanism whereby *TP53* mutations are counter-selected in *CDKN2A* mutant BO cells. Since *TP53* loss is a strong driver of OAC initiation, the decrease of its occurrence induced by *CDKN2A* LoF also decreases tumour formation. Like TP53 immunostain, which is currently one of the few molecular markers of OAC progression[49], *CDKN2A* mutational screen or protein immunostain could be developed as a novel marker of favourable prognosis in non-dysplastic BO. Endoscopic surveillance of BO is an integral component of the current OAC prevention paradigm, but the rate of progression to OAC is only 0.54/100 patient-years[50]. Identifying BOs with a lower risk of progression could substantially improve patient management, decreasing the burden of endoscopy for patients who have low chances to develop cancer.

CDKN2A LoF is the most frequent event in 9p21 locus, implying that the co-occurring loss of other 9p21 genes is due to genetic hitchhiking, with variable effects on cell cycle, oxidative phosphorylation, and interferon response depending on the stage and context of BO and OAC evolution. Most notably, *IFNE* exerts a tumour-suppressive role in BO but not OAC by reducing IFN response and inducing a cold

immune microenvironment. Despite several reports of a lower infiltration of immune cells in cancers with reduced *CDKN2A* expression[51,52], *CDKN2A* LoF alone does not change the immune composition of BO or OAC TME. This may be due to tumour-specific effects or to the fact that at least some cancer promoting roles previously attributed to *CDKN2A* LoF are in fact triggered by the loss of other 9p21 genes.

The association of *CDKN2A* LoF with bad prognosis is also context dependent and detectable only in OAC patients. It appears unrelated to the effect of *CDKN2A* LoF on cell cycle since alterations in other cell cycle regulators can drive OAC without affecting survival. A contribution towards a more aggressive OAC phenotype is likely due to a combination of effects, including the pervasive suppression of transcriptional programmes responsible for the maintenance of squamous epithelium. Although this is a common feature of BO and OAC[44], it is significantly more pronounced when *CDKN2A* is lost and is achieved through and *SOX15* downregulation. This could be an indirect effect of *CDKN2A* LoF on E2F transcriptional programme since iASPP, which controls *TP63* expression[53], is a target of E2F1[54] and *SOX15*, in turn, is a target of TP63[55].

CDKN2A LoF alone is already predictive of poor survival, but the loss of other 9p21 genes including *KLHL9*, *IFNE, MTAP* and *CDKN2B* enhances the effect. Understanding how each gene loss impacts on survival is challenging because they all co-occur with *CDKN2A* loss. Alternative approaches such as *in vitro* and *in vivo* cancer models where individual genes can be more easily manipulated may be informative. Our study introduces the intriguing concept that the functional consequences of alterations in cancer genes may change during the evolution of disease, from preventing cancer transformation in premalignant setting to favouring a more aggressive disease at later stages. If proven of general applicability, this may

lead to a paradigm shift with consequences on the understanding and treatment of cancer.

METHODS

Sample collection and ethical approval

Single Nucleotide Variants (SNVs), indels and copy number (CN) data for 1032 primary OACs were collected from published studies and *de novo* sequenced samples (**Table S1**). In particular, WGS from 706 OACs was performed at the University of Cambridge (UoC, EGAD00001011191 and EGAD00001006083, https://ega-archive.org/). WES for 73 TCGA OACs were downloaded from the Genomic Data Commons portal (https://portal.gdc.cancer.gov/). Damaged genes for 253 Memorial Sloan Kettering Cancer Center (MSKCC) OACs that underwent targeted re-sequencing of 528[29], 477[8] and 970[30] genes were downloaded from the cBioPortal (https://www.cbioportal.org/). In cases of multiple samples per patient, the sample with *CDKN2A* LoF was retained. Clinical data for the TCGA and MSKCC cohorts were obtained from the same sources. For the UoC cohort clinical data were derived from LabKey (https://occams.cs.ox.ac.uk/labkey). Bulk RNAseq data were available for 337 OACs, all of which had matched WGS or WES (**Table S1**). Of these, 264 were sequenced at the UoC (EGAD00001011190) and 73 derived from TCGA.

WGS, WES and clinical data for 356 BOs were obtained from UoC (EGAD00001011191 and EGAD00001011189, which also includes samples from[34] and[35]) and from the Fred Hutchinson Cancer Research Center (FHCRC)[17,33] (**Table S1**). As for OAC, in cases of multiple samples per patient, the sample with *CDKN2A* LoF was retained. BOs were classified as progressors (P-BO, 257) and non-progressors (NP-BO, 99) based on whether patients progressed or not to high-grade dysplasia or OAC in a follow-up period of up to 17 years (**Table S1**).

Paired WGS BO and OAC data were available for 86 cases (EGAD00001011191 and EGAD00001006083, which also include samples

from[26,27,34]). Bulk RNAseq data for 108 P-BOs and 22 NP-BOs were sequenced at the UoC (EGAD00001011190, including samples from[34]) (**Table S1**).

BO and OAC patients whose samples from UoC sequenced for this study were consented to research for the OCCAMS project (REC: 10/H0305/1 & IRAS:15757). Samples were collected either at endoscopy, staging laparoscopy, endoscopic mucosal resection, or surgical resection then snap frozen in liquid nitrogen. Samples were then embedded in optimal cutting temperature media for cutting of 1 x 3μM slide to be H&E stained and reviewed by a pathologist. Only tumour samples of >50% cellularity and BO samples with high intestinal metaplasia (IM) content proceeded to sequencing.

DNA and RNA extraction, library preparation and variant calling

DNA and RNA were extracted using Qiagen AllPrep Mini kits, using a Precellys for tissue dissociation after all excess OCT was removed. Extracted nucleic acids were quantified by Qubit. Libraries were then prepared using Illumina PCR Free methods and sequenced on HiSeq 4000 or NovaSeq platforms. Paired-end whole genome sequencing at 50X target depth for OACs, P-BOs and NP-BOs and 30X target depth for matched normal (blood) was performed by Illumina, the Sanger Institute, or the CRUK Cambridge Institute on Illumina platforms. Quality checks were performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For mutation calling, sequencing reads were aligned against the reference genome (hg19/GRCh37) using BWA-MEM[56]. Aligned reads were then sorted into genome coordinate order and duplicate reads were flagged using Picard MarkDuplicates (http://broadinstitute.github.io/picard). Strelka[57] 2.0.15 was used for calling single nucleotide variants and indels. Sample purity and ploidy values were estimated using

ASCAT-NGS 2.1[58]. Copy number alterations after correction for estimated normal-cell contamination were inferred using ASCAT from read counts at germline heterozygous positions estimated by GATK 3.2-2 HaplotypeCaller[59]. Shallow WGS data for 75 BOs[35] were processed with the QDNAseq package[60] using 50-kb bins. The QDNAseq package was also used for GC-bias correction, segmentation and generation of copy number calls and used to identify homozygously deleted and amplified genes. Since the read depth was only 0.4x, mutation calls could not be generated for these samples.

Annotation of damaged genes and OAC drivers

For WGS (UoC, FHCRC) and WES (TCGA) data, SNV, indel and copy number calls were taken from the original publications or derived as described above. ANNOVAR[61] (April 2018) and dbNSFP[62] v3. 0 were used to annotate the effect of mutations and indels. Only SNvs and indels with damaging effects on the proteins as previously described[1] were further retained. Briefly, these included (1) truncating (stopgain, stoploss, frameshift) mutations; (2) missense mutations predicted by at least seven methods[1].

Copy number alteration (CNA) segments from ASCAT were intersected with the exonic coordinates of 19,641 unique human genes[1] and a gene was considered amplified, homozygously or heterozygously deleted if at least 25% of its length overlapped with an amplified (CNA > twice sample ploidy) or homozygously (CNA = 0) or heterozygously deleted (CNA = 1) segment, respectively. Genes with at least one damaging SNV or indel as well as amplified and homozygously deleted genes were considered damaged. Genes with heterozygous deletion of one allele and at least a damaging SNV or indel in the other, were also considered damaged. The distribution of damaged genes across OAC and BO cohorts is shown in **Figure S1**.

Mutated, amplified and homozygously deleted genes for the MSKCC cohort[8,29,30] were downloaded from the cBioPortal.

A list of 40 OAC canonical drivers was obtained from the Network of Cancer Genes (NCG7.1, http://www. network-cancer-genes. org)[1]. Additionally, 34 OAC drivers that undergo CNA were collected through manual curation of the literature[22,32,63-65]. Only 54 of the resulting 74 OAC drivers were present also in the gene panel used in the MSKCC studies[8,29,30] and these were considered for further analysis **(Table S2)**.

Logistic regression and survival analysis

Logistic regression with Firth bias correction[66] was used to test the difference between two models of OAC initiation in the entire BO (P-BO and NP-BO) cohort. The first model assumed *TP53* LoF as the only driver (model 1), while the second model assumed that both *TP53* and *CDKN2A* LoF impacted on OAC initiation (model 2). The models were developed using the package logistf v1.25.0 and compared using the anova function in R. The two models were used to estimate the numbers of expected BOs that progressed to OAC according to corresponding genomic status of *TP53* and *CDKN2A*. The β coefficients for *TP53* and *CDKN2A* LoF were obtained from the regression models and the p-values were calculated using the chi-squared test. Negative or positive β coefficient values indicated cancer-protective or cancer-promoting roles, respectively. The β coefficient (β) of CDKN2A LoF in model 2 was used to estimate the odds of progression as:

$$odds = e^{\beta}$$

The results of the whole analysis are reported in **Table S3**.

Kaplan-Meier survival analysis was performed with survminer v.0.4.9 using the log-rank method. The analysis of the survival effect of *CDKN2A* co-damage with other 9p21 genes was performed only on 779 OAC patients with WGS or WES data as the information on the genomic alteration of all 9p21 genes was not available in the targeted re-sequencing studies. Log-rank method was used to estimate the p-values, which were then corrected for multiple hypothesis testing using the Benjamini–Hochberg method, when needed.

RNA sequencing, gene set enrichment and immune infiltration

Paired-end RNA sequencing for OAC, P-BO and NP-BO from UoC was performed at the CRUK Cambridge Institute on Illumina platforms and quality checks were performed using FastQC. Reads were aligned using STAR with ENSEMBL gene annotation. Reads per gene were quantified using the summariseOverlaps function from the GenomicRanges package[67]. Raw read counts of 18,846 human genes shared between the UoC and TCGA cohorts were extracted from the corresponding BO and OAC RNAseq datasets. SMIXnorm v0.0.0.9[68] was used to estimate the probability of expression of these genes across all samples. Genes with a probability of expression below 0.9 were filtered out, resulting in 16,901 retained genes in OAC, 15,134 in P-BO and 15,866 in NP-BO, respectively.

Twenty-two NP-BOs, 108 P-BOs and 337 OACs with matched genomic and transcriptomic data (**Table S1**) were divided into four groups depending on the mutation and copy number profiles of the five 9p21 genes (*CDKN2B, CDKN2A*, *MTAP*, *IFNE* and *KLHL9*) with impact on survival. Differential gene expression analysis was performed between each of these groups and the corresponding 9p21 wild type OACs (185), P-BOs (33), and NP-BOs (6) using DESeq2 v1.38.3[46] after correction for the

batch effect with DESeqDataSetFromMatrix. Genes were ordered according to log2 fold-change values and used for pre-ranked GSEA using fgsea v1.24.0[37] against 50 gene sets from MSigDB v7.5.1[69] and 1,303 level 2-8 pathways from Reactome v.72[70] containing between 10 and 500 expressed genes and excluding the 'Disease' hierarchical level. The resulting p-values were corrected for multiple testing in each analysis separately using the Benjamini-Hochberg method. Pathway redundancy was removed accounting for the extent of overlap between leading-edge genes, i.e. the genes that contributed the most to the enrichment. If the number of unique leading-edge genes in a pathway was higher than the shared and the unique leading-edge genes in the other pathway, the latter was removed. If the number of shared leading-edge genes between two pathways was higher than the unique leading-edge genes in both, the pathway with the higher FDR was removed. Retained processes are reported in **Table S5.**

To estimate the abundance of immune cell populations in the from bulk RNA-seq data, raw read counts of the expressed genes from 22 NP-BOs, 108 P-BOs and 337 OACs were normalized to transcripts per million (TPM) values after batch correction with ComBat-seq[71]. Resulting TPMs were used as input for ConsensusTME v0.0.1[72] as implemented in immunedeconv v2.1.0 to estimate the normalised enrichment score (NES) using 16 oesophageal carcinoma immune signatures. To further estimate the abundance of MDSCs, two M-MDSC and G-MDSC signatures[73] were used in ConsensusTME custom mode.

RNAScope and Imaging Mass Cytometry

A panel of 26 antibodies targeting structural markers, immune markers, three 9p21 proteins and three RNAScope probes against *IFNE*, *IFNB1* and *PPIB* mRNAs

was assembled (**Table S9**). RNAScope staining was detected using metal-tagged antibodies as previously described[74]. Sixteen of these antibodies were already metal-tagged (Standard Biotools), while eleven were carrier-free and tagged using the Maxpar X8 metal conjugation kit (Standard Biotools). The whole panel was tested in OAC FFPE sections using three dilutions ranging from 1:100 to 1:3500 and the dilution giving the highest signal-to-noise ratio was chosen for each antibody (**Table S9**).

Five μm thick sections were obtained from FFPE blocks of ten OAC patients selected based on their 9p21 gene profile (**Table S8**). Slides were incubated for one hour at 60°C, loaded on a Leica Bond autostainer (Leica Biosystems) and processed using the RNASCope LS Multiplex Fluorescent Assay following manufacturer's instructions and *IFNE*, *IFNB1*, and *PPIB* probes at a 1:50 dilution. C2 oligos were developed with TSA-digoxinenin, C3 oligos with TSA-biotin and C1 oligos with TSA-FITC (diluted 1:200 in TSA buffer). Slides were blocked for two hours at room temperature in a Sequenza rack (Thermo Fisher Scientific). Slides were incubated overnight at 4°C with the mix of metal-conjugated antibodies, washed, and incubated with the DNA intercalator Cell-ID Intercalator-Ir (Standard Biotools). Slides were removed from the Sequenza rack, air-dried and loaded into the Hyperion Imaging System (Standard Biotools). Regions of interest (ROI) were manually selected to contain areas with tumour and immune cells by a certified pathologist (M.R.J). Regions of about 1.44 mm^2 were laser-ablated within the pre-selected ROIs at 1 μm/pixel resolution and 400 Hz frequency.

IMC image analysis was performed using SIMPLI[75]. TIFF images for each metal-tagged antibody and DNA intercalator were obtained from the raw.txt files of the ablated regions. Pixel intensities for each channel were normalised to the 99th percentile of the intensity distribution. Background pixels of the normalised images

were removed with CellProfiler4[76] using global thresholding and processed images were verified by an expert histologist (J.S.). Single cell segmentation was performed using CellProfiler4[76] to identify cell nucleus (DNA1 channel) and membrane (cadherin-1, pan-keratin, CD3, CD8, CD4, CD11b, CD11c, NCAM1, CD68, CD27, CD163, CD16, CD15, CD14). Obtained cells were phenotyped based on at least 10% overlap with the masks of individual cell types in the following order: (1) $CD15^+$ and $CD16^+$ for neutrophils; (2) $NCAM1^+$ for NK cells; (3) $CD11c^+$ for dendritic cells; (4) $CD68^+$ for macrophages; (5) $CD14^+$ for M-MDSCs; (6) $CD15^+$ for G-MDSCs; (7) $CD3^+$ for T cells; (8) cadherin-1 and pan-keratin for tumour cells and (9) vimentin for stromal cells. Cells with <10% overlap with any mask were left unassigned.

Unsupervised clustering was performed separately on $CD3^+$ T cells and $CD68^+$ macrophages using Seurat v.2.4[77], with random seed = 123 and 0.3, 0.5, 0.7 and 0.9 cluster resolutions. Markers used for clustering were CD3, CD4, CD8, FOXP3, GzMB and Ki67 for T cells, and CD68, CD11c, HLA-DR/DP/DQ and CD163, CD11b and Ki67 for macrophages. Silhouette score of each cluster was calculated using v.2.1.6 package. The resolution with the highest median silhouette score was identified as the best clustering resolution for each cell type.

Keratinisation causal regulatory network analysis

Causal networks linking *CDKN2A* LoF to the downregulation of keratinization were inferred using a three-step protocol modified from[78], separately for P-BO and OAC (**Figures S3**). In the first step, co-regulated gene modules were identified using cMonkey2[79] based on gene co-expression, proximity in the protein-protein interaction network (PPIN) and enrichment in transcription factor (TF) targets. Co-expressed genes were identified from the top 50% most variably expressed genes in P-BO and

OAC after converting read counts into z-scores using DESeq2 v1.38.3[46]. Proximity in the PPIN was measured using the human weighted PPIN from STRING v11.5[80]. GO:0006355 term of Gene Ontology (release 2022-05) was used to identify 1,471 TFs. These were in turn used as input for ARACNE-AP[81] together with P-BO and OAC gene expression data to identify TF-target pairs. cMonkey2 was run with a fixed number of iterations (n = 2000) and seed value (n = 123) for the initialization step to ensure reproducibility. The number of gene modules (k) was determined as:

$$k = \frac{nAG * nBpG}{nGpB}$$

where *nAG* was the number of analysed genes, *nBpG* was the maximum number of gene modules each gene could appear in (fixed to 2), and *nGpB* was the average number of genes per gene module (fixed to 30). Identified gene modules were then filtered based on (i) co-expression quality according to the first principal component (FDR ≤0.1 and variance explained ≥0.32 for P-BO and ≥0.25 for OAC), (ii) functional enrichment in keratinisation-related genes (two-sided Fisher's test p-value ≤0.01), (iii) enrichment in TF target genes (two-sided Fisher's test p-value ≤0.01), (iv) correlation of TFs with gene module eigengenes, *i.e.* genes that explain the maximum expression variance. In the second step, the single.marker.analysis function of the Network Edge Orienting (NEO)[78,82] method was used to infer causal models where *CDKN2A* LoF causally affected the expression of specific TFs, which, in turn, altered keratinisation gene modules. To assess statistical significance, the next best single marker score was defined as the log10 probability of the causal model divided by the log10 probability of the next best fitting alternative model[82] and causal models with next best single marker score ≥0.3 were considered significant. In the third step, significant causal models were further retained if (1) TFs were differentially expressed (FDR < 0.1) in group 4 as compared to 9p21 wild type P-BOs and OACs and (2) there was

significant positive correlation (R >0.5 and FDR <0.1) between TF expression and the GSEA NES score of the predicted targets in P-BOs and OACs. The final list of significant causal models is reported in **Table S10**.

AVAILABILITY OF DATA AND MATERIALS

DNA and RNA sequence data for the UoC cohort are available from the European Genome-phenome Archive using the following accession IDs: WGS (EGAD00001011191, EGAD00001006083), shallow WGS (EGAD00001011189), bulk RNA-sequencing (EGAD00001011190).

CODE AVAILABILITY

No unique code was developed for this study.

ACKNOWLEDGEMENTS

The authors thank Patricia C. Galipeau (Fred Hutchinson Cancer Research Center, USA) for sharing somatic variant call files for the BO samples, Michele Bortolomeazzi (Deutsches Krebsforschungszentrum, Germany), Michael Pitcher and Lucia Montorsi (King's College London, UK) for help with the IMC analysis.

AUTHOR CONTRIBUTIONS

F.D.C. conceived and directed the study with the support of P.G.; R.C.F. recruited participants and led the OCCAMS Consortium; P.G., C.C.B., M.A., A.M., A.Z., H.M., G.D and F.D.C. analysed the data; G.D. constructed and managed the sequencing alignment and variant-calling pipelines. G.K. guided the statistical analysis. A.F. coordinated and carried out the processing of patient samples. A. F. and P.G. screened histopathological reports and identified samples; M.A., A.A.S., M.G., and E.N. performed the IMC experiments; M.R.J. and J.S. assessed tissue sections. P.G. and F.D.C. wrote the manuscript with contributions from C.C.B., M.A., A.M., A.Z., H.M., G.D. and A.F. All authors approved the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

OPEN ACCESS

For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

FUNDING

This work was supported by Cancer Research UK [C43634/A25487 to F. D. C.] and [EDDPJT-Nov21\100010 to F. D. C], the Cancer Research UK City of London Centre [C7893/A26233 to F. D. C], Bart's Charity and the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001002), the UK Medical Research Council (FC001002), and the Wellcome Trust (FC001002).

FULL LIST OF AUTHORS

Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium:

Rebecca C. Fitzgerald[1], Paul A.W. Edwards[1,2], Nicola Grehan[1,5], Barbara Nutzinger[1], Aisling M Redmond[1], Christine Loreno[1], Sujath Abbas[1], Adam Freeman[1] Elizabeth C. Smyth[5], Maria O'Donovan[1,3], Ahmad Miremadi[1,3], Shalini Malhotra[1,3], Monika Tripathi[1,3], Hannah Coles[1] Curtis Millington[1], Matthew Eldridge[2], Maria Secrier[2], Ginny Devonshire[1,2], Jim Davies[4], Charles Crichton[4], Nick Carroll[5], Richard H.Hardwick[5], Peter Safranek[5], Andrew Hindmarsh[5], Vijayendran Sujendran[5], Stephen J. Hayes[6,13], Yeng Ang[6,7,26], Andrew Sharrocks[26], Shaun R. Preston[8], Izhar Bagwan[8], Vicki Save[9], Richard J.E. Skipworth[9], Ted R. Hupp[20], J. Robert O'Neill[5,9,20], Olga Tucker[10,29],

Andrew Beggs[10,25], Philippe Taniere[10], Sonia Puig[10], Gianmarco Contino[10], Timothy J. Underwood[11,12], Robert C. Walker[11,12], Ben L. Grace[11], Jesper Lagergren[14,22], James Gossage[14,21], Andrew Davies[14,21], Fuju Chang[14,21], Ula Mahadeva[14], Vicky Goh[21], Francesca D. Ciccarelli[21], Grant Sanders[15], Richard Berrisford[15], David Chan[15], Ed Cheong[16], Bhaskar Kumar[16], L. Sreedharan[16] Simon L Parsons[17], Irshad Soomro[17], Philip Kaye[17], John Saunders[6, 17], Laurence Lovat[18], Rehan Haidry[18], Michael Scott[19], Sharmila Sothi[23], Suzy Lishman[2, 24], George B. Hanna[27], Christopher J. Peters[27],Krishna Moorthy[27], Anna Grabowska[28], Richard Turkington[30], Damian McManus[30], Helen Coleman[30], Russell D Petty[31] , Freddie Bartlett[32]

[1] Early Cancer Institute, University of Cambridge, Cambridge, UK

[2] Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK

[3] Department of Histopathology, Addenbrooke's Hospital, Cambridge, UK

[4]Department of Computer Science, University of Oxford, UK, OX1 3QD

[5]Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK, CB2 0QQ

[6]Salford Royal NHS Foundation Trust, Salford, UK, M6 8HD

[7]Wigan and Leigh NHS Foundation Trust, Wigan, Manchester, UK, WN1 2NN

[8]Royal Surrey County Hospital NHS Foundation Trust, Guildford, UK, GU2 7XX

[9]Edinburgh Royal Infirmary, Edinburgh, UK, EH16 4SA

[10]University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK, B15 2GW

[11]University Hospital Southampton NHS Foundation Trust, Southampton, UK, SO16 6YD

[12]Cancer Sciences Division, University of Southampton, Southampton, UK, SO17 1BJ

[13]Faculty of Medical and Human Sciences, University of Manchester, UK, M13 9PL

[14] Guy's and St Thomas's NHS Foundation Trust, London, UK, SE1 7EH

[15]Plymouth Hospitals NHS Trust, Plymouth, UK, PL6 8DH

[16]Norfolk and Norwich University Hospital NHS Foundation Trust, Norwich, UK, NR4 7UY

[17]Nottingham University Hospitals NHS Trust, Nottingham, UK, NG7 2UH

[18]University College London, London, UK, WC1E 6BT

[19]Wythenshawe Hospital, Manchester, UK, M23 9LT

[20]Edinburgh University, Edinburgh, UK, EH8 9YL

[21]Barts Cancer Institute, Queen Mary University of London, London, UK, EC1M 6BQ

[22]Karolinska Institute, Stockholm, Sweden, SE-171 77

[23]University Hospitals Coventry and Warwickshire NHS, Trust, Coventry, UK, CV2 2DX

[24]Peterborough Hospitals NHS Trust, Peterborough City Hospital, Peterborough, UK, PE3 9GZ

[25]Institute of Cancer and Genomic sciences, University of Birmingham, B15 2TT

[26]GI science centre, University of Manchester, UK, M13 9PL.

[27]Department of Surgery and Cancer, Imperial College, London, UK, W2 1NY

[28]Queen's Medical Centre, University of Nottingham, Nottingham, UK

[29]Heart of England NHS Foundation Trust, Birmingham, UK, B9 5SS.

[30]Centre for Cancer Research and Cell Biology, Queen's University Belfast, Northern Ireland BT7 1NN.

[31]Tayside Cancer Centre, Ninewells Hospital and Medical School, Dundee, DD1 9SY

[32] Portsmouth Hospitals NHS Trust, Portsmouth, PO6 3LY

References

1 Dressler, L. *et al.* Comparative assessment of genes driving cancer and somatic evolution in non-cancer tissues: an update of the Network of Cancer Genes (NCG) resource. *Genome Biol* **23**, 35 (2022). https://doi.org/10.1186/s13059-022-02607-z

2 Sherr, C. J. The INK4a/ARF network in tumour suppression. *Nat Rev Mol Cell Biol* **2**, 731-737 (2001). https://doi.org/10.1038/35096061

3 Collado, M. & Serrano, M. The power and the promise of oncogene-induced senescence markers. *Nat Rev Cancer* **6**, 472-476 (2006). https://doi.org/10.1038/nrc1884

4 Baruah, P. *et al.* Impact of p16 status on pro- and anti-angiogenesis factors in head and neck cancers. *Br J Cancer* **113**, 653-659 (2015). https://doi.org/10.1038/bjc.2015.251

5 Jenkins, N. C. *et al.* The p16(INK4A) tumor suppressor regulates cellular oxidative stress. *Oncogene* **30**, 265-274 (2011). https://doi.org/10.1038/onc.2010.419

6 Zeng, H. *et al.* Bi-allelic Loss of CDKN2A Initiates Melanoma Invasion via BRN2 Activation. *Cancer Cell* **34**, 56-68 e59 (2018). https://doi.org/10.1016/j.ccell.2018.05.014

7 Izadi, F. *et al.* Genomic Analysis of Response to Neoadjuvant Chemotherapy in Esophageal Adenocarcinoma. *Cancers (Basel)* **13** (2021). https://doi.org/10.3390/cancers13143394

8 Sihag, S. *et al.* Next-Generation Sequencing of 487 Esophageal Adenocarcinomas Reveals Independently Prognostic Genomic Driver Alterations and Pathways. *Clin Cancer Res* **27**, 3491-3498 (2021). https://doi.org/10.1158/1078-0432.CCR-20-4707

9 Gutiontov, S. I. *et al.* CDKN2A loss-of-function predicts immunotherapy resistance in non-small cell lung cancer. *Sci Rep* **11**, 20059 (2021). https://doi.org/10.1038/s41598-021-99524-1

10 Han, G. *et al.* 9p21 loss confers a cold tumor immune microenvironment and primary resistance to immune checkpoint therapy. *Nat Commun* **12**, 5606 (2021). https://doi.org/10.1038/s41467-021-25894-9

11 Barriga, F. M. *et al.* MACHETE identifies interferon-encompassing chromosome 9p21.3 deletions as mediators of immune evasion and metastasis. *Nat Cancer* **3**, 1367-1385 (2022). https://doi.org/10.1038/s43018-022-00443-5

12 Marks, Z. R. C. *et al.* Interferon-epsilon is a tumour suppressor and restricts ovarian cancer. *Nature* **620**, 1063-1070 (2023). https://doi.org/10.1038/s41586-023-06421-w

13 Gjuka, D. *et al.* Enzyme-mediated depletion of methylthioadenosine restores T cell function in MTAP-deficient tumors and reverses immunotherapy resistance. *Cancer Cell* (2023). https://doi.org/10.1016/j.ccell.2023.09.005

14 Barrett, M. T. *et al.* Allelic loss of 9p21 and mutation of the CDKN2/p16 gene develop as early lesions during neoplastic progression in Barrett's esophagus. *Oncogene* **13**, 1867-1873 (1996).

15 Weaver, J. M. J. *et al.* Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat Genet* **46**, 837-843 (2014). https://doi.org/10.1038/ng.3013

16 Killcoyne, S. & Fitzgerald, R. C. Evolution and progression of Barrett's oesophagus to oesophageal cancer. *Nat Rev Cancer* **21**, 731-741 (2021). https://doi.org/10.1038/s41568-021-00400-x

17 Paulson, T. G. *et al.* Somatic whole genome dynamics of precancer in Barrett's esophagus reveals features associated with disease progression. *Nat Commun* **13**, 2300 (2022). https://doi.org/10.1038/s41467-022-29767-7

18 Galipeau, P. C., Prevo, L. J., Sanchez, C. A., Longton, G. M. & Reid, B. J. Clonal expansion and loss of heterozygosity at chromosomes 9p and 17p in premalignant esophageal (Barrett's) tissue. *J Natl Cancer Inst* **91**, 2087-2095 (1999). https://doi.org/10.1093/jnci/91.24.2087

19 Barrett, M. T. *et al.* Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat Genet* **22**, 106-109 (1999). https://doi.org/10.1038/8816

20 Maley, C. C. *et al.* Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet* **38**, 468-473 (2006). https://doi.org/10.1038/ng1768

21 Maley, C. C. *et al.* Selectively advantageous mutations and hitchhikers in neoplasms: p16 lesions are selected in Barrett's esophagus. *Cancer Res* **64**, 3414-3427 (2004). https://doi.org/10.1158/0008-5472.CAN-03-3249
22 Nones, K. *et al.* Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nat Commun* **5**, 5224 (2014). https://doi.org/10.1038/ncomms6224
23 Stachler, M. D. *et al.* Detection of Mutations in Barrett's Esophagus Before Progression to High-Grade Dysplasia or Adenocarcinoma. *Gastroenterology* **155**, 156-167 (2018). https://doi.org/10.1053/j.gastro.2018.03.047
24 Cancer Genome Atlas Research, N. *et al.* Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169-175 (2017). https://doi.org/10.1038/nature20805
25 Frankell, A. M. *et al.* The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet* **51**, 506-516 (2019). https://doi.org/10.1038/s41588-018-0331-5
26 Noorani, A. *et al.* Genomic evidence supports a clonal diaspora model for metastases of esophageal adenocarcinoma. *Nat Genet* **52**, 74-83 (2020). https://doi.org/10.1038/s41588-019-0551-3
27 Ross-Innes, C. S. *et al.* Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat Genet* **47**, 1038-1046 (2015). https://doi.org/10.1038/ng.3357
28 Ng, A. W. T. *et al.* Rearrangement processes and structural variations show evidence of selection in oesophageal adenocarcinomas. *Commun Biol* **5**, 335 (2022). https://doi.org/10.1038/s42003-022-03238-7
29 Janjigian, Y. Y. *et al.* Genetic Predictors of Response to Systemic Therapy in Esophagogastric Cancer. *Cancer Discov* **8**, 49-58 (2018). https://doi.org/10.1158/2159-8290.CD-17-0787
30 Samstein, R. M. *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* **51**, 202-206 (2019). https://doi.org/10.1038/s41588-018-0312-8
31 Stachler, M. D. *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet* **47**, 1047-1055 (2015). https://doi.org/10.1038/ng.3343
32 Dulak, A. M. *et al.* Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. *Cancer Res* **72**, 4383-4393 (2012). https://doi.org/10.1158/0008-5472.CAN-11-3893
33 Galipeau, P. C. *et al.* NSAID use and somatic exomic mutations in Barrett's esophagus. *Genome Med* **10**, 17 (2018). https://doi.org/10.1186/s13073-018-0520-y
34 Katz-Summercorn, A. C. *et al.* Multi-omic cross-sectional cohort study of pre-malignant Barrett's esophagus reveals early structural variation and retrotransposon activity. *Nat Commun* **13**, 1407 (2022). https://doi.org/10.1038/s41467-022-28237-4
35 Killcoyne, S. *et al.* Genomic copy number predicts esophageal cancer years before transformation. *Nat Med* **26**, 1726-1732 (2020). https://doi.org/10.1038/s41591-020-1033-y
36 Reid, B. J. *et al.* Predictors of progression in Barrett's esophagus II: baseline 17p (p53) loss of heterozygosity identifies a patient subset at increased risk for neoplastic progression. *Am J Gastroenterol* **96**, 2839-2848 (2001). https://doi.org/10.1111/j.1572-0241.2001.04236.x
37 Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv*, 060012 (2021). https://doi.org/10.1101/060012
38 Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* **34**, 57-65 (2013). https://doi.org/10.1002/humu.22225
39 Hu, Q. *et al.* MTAP Deficiency-Induced Metabolic Reprogramming Creates a Vulnerability to Cotargeting De Novo Purine Synthesis and Glycolysis in Pancreatic Cancer. *Cancer Res* **81**, 4964-4980 (2021). https://doi.org/10.1158/0008-5472.CAN-20-0414
40 Shi, L. Z. & Bonner, J. A. Bridging Radiotherapy to Immunotherapy: The IFN-JAK-STAT Axis. *Int J Mol Sci* **22** (2021). https://doi.org/10.3390/ijms222212295

41 Hardy, M. P., Owczarek, C. M., Jermiin, L. S., Ejdeback, M. & Hertzog, P. J. Characterization of the type I interferon locus and identification of novel genes. *Genomics* **84**, 331-345 (2004). https://doi.org/10.1016/j.ygeno.2004.03.003
42 Sulahian, R. *et al.* SOX15 governs transcription in human stratified epithelia and a subset of esophageal adenocarcinomas. *Cell Mol Gastroenterol Hepatol* **1**, 598-609 e596 (2015). https://doi.org/10.1016/j.jcmgh.2015.07.009
43 Thompson, C. A., DeLaForest, A. & Battle, M. A. Patterning the gastrointestinal epithelium to confer regional-specific functions. *Dev Biol* **435**, 97-108 (2018). https://doi.org/10.1016/j.ydbio.2018.01.006
44 Nowicki-Osuch, K. *et al.* Molecular phenotyping reveals the identity of Barrett's esophagus and its malignant transition. *Science* **373**, 760-767 (2021). https://doi.org/10.1126/science.abd1449
45 Busslinger, G. A. *et al.* Human gastrointestinal epithelia of the esophagus, stomach, and duodenum resolved at single-cell resolution. *Cell Rep* **34**, 108819 (2021). https://doi.org/10.1016/j.celrep.2021.108819
46 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8
47 Acha-Sagredo, A., Ganguli, P. & Ciccarelli, F. D. Somatic variation in normal tissues: friend or foe of cancer early detection? *Ann. Oncol.* **33**, 1239-1249 (2022). https://doi.org/10.1016/j.annonc.2022.09.156
48 Colom, B. *et al.* Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. *Nat Genet* **52**, 604-614 (2020). https://doi.org/10.1038/s41588-020-0624-3
49 Fitzgerald, R. C. *et al.* British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* **63**, 7-42 (2014). https://doi.org/10.1136/gutjnl-2013-305372
50 Iyer, P. G. & Chak, A. Surveillance in Barrett's Esophagus: Challenges, Progress, and Possibilities. *Gastroenterology* **164**, 707-718 (2023). https://doi.org/10.1053/j.gastro.2023.01.031
51 Chen, Z. *et al.* Comprehensive Analysis Revealed that CDKN2A is a Biomarker for Immune Infiltrates in Multiple Cancers. *Front Cell Dev Biol* **9**, 808208 (2021). https://doi.org/10.3389/fcell.2021.808208
52 Cheng, T. *et al.* CDKN2A-mediated molecular subtypes characterize the hallmarks of tumor microenvironment and guide precision medicine in triple-negative breast cancer. *Front Immunol* **13**, 970950 (2022). https://doi.org/10.3389/fimmu.2022.970950
53 Chikh, A. *et al.* iASPP/p63 autoregulatory feedback loop is required for the homeostasis of stratified epithelia. *EMBO J* **30**, 4261-4273 (2011). https://doi.org/10.1038/emboj.2011.302
54 Pandolfi, S., Montagnani, V., Lapucci, A. & Stecca, B. HEDGEHOG/GLI-E2F1 axis modulates iASPP expression and function and regulates melanoma cell growth. *Cell Death Differ* **22**, 2006-2019 (2015). https://doi.org/10.1038/cdd.2015.56
55 Sethi, I. *et al.* A global analysis of the complex landscape of isoforms and regulatory networks of p63 in human cells and tissues. *BMC Genomics* **16**, 584 (2015). https://doi.org/10.1186/s12864-015-1793-9
56 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009). https://doi.org/10.1093/bioinformatics/btp324
57 Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811-1817 (2012). https://doi.org/10.1093/bioinformatics/bts271
58 Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* **107**, 16910-16915 (2010). https://doi.org/10.1073/pnas.1009843107
59 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010). https://doi.org/10.1101/gr.107524.110

60 Scheinin, I. *et al.* DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res* **24**, 2022-2032 (2014). https://doi.org/10.1101/gr.175141.114
61 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010). https://doi.org/10.1093/nar/gkq603
62 Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat* **37**, 235-241 (2016). https://doi.org/10.1002/humu.22932
63 Frankel, A. *et al.* Genome-wide analysis of esophageal adenocarcinoma yields specific copy number aberrations that correlate with prognosis. *Genes Chromosomes Cancer* **53**, 324-338 (2014). https://doi.org/10.1002/gcc.22143
64 Murugaesu, N. *et al.* Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov* **5**, 821-831 (2015). https://doi.org/10.1158/2159-8290.CD-15-0412
65 Secrier, M. *et al.* Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat Genet* **48**, 1131-1141 (2016). https://doi.org/10.1038/ng.3659
66 Heinze, G. & Schemper, M. A solution to the problem of separation in logistic regression. *Stat Med* **21**, 2409-2419 (2002). https://doi.org/10.1002/sim.1047
67 Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118 (2013). https://doi.org/10.1371/journal.pcbi.1003118
68 Yin, S. *et al.* SMIXnorm: Fast and Accurate RNA-Seq Data Normalization for Formalin-Fixed Paraffin-Embedded Samples. *Front Genet* **12**, 650795 (2021). https://doi.org/10.3389/fgene.2021.650795
69 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015). https://doi.org/10.1016/j.cels.2015.12.004
70 Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-D503 (2020). https://doi.org/10.1093/nar/gkz1031
71 Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* **2**, lqaa078 (2020). https://doi.org/10.1093/nargab/lqaa078
72 Jimenez-Sanchez, A., Cast, O. & Miller, M. L. Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. *Cancer Res* **79**, 6238-6246 (2019). https://doi.org/10.1158/0008-5472.CAN-18-3560
73 Alshetaiwi, H. *et al.* Defining the emergence of myeloid-derived suppressor cells in breast cancer using single-cell transcriptomics. *Sci Immunol* **5** (2020). https://doi.org/10.1126/sciimmunol.aay6017
74 Montorsi, L. *et al.* Unanticipated interacting features of human gut-associated lymphoid tissues link microbiota, intestinal immunity and autoimmunity. *bioRxiv*, 2023.2008. 2029.555265 (2023).
75 Bortolomeazzi, M. *et al.* A SIMPLI (Single-cell Identification from MultiPLexed Images) approach for spatially-resolved tissue phenotyping at single-cell resolution. *Nat Commun* **13**, 781 (2022). https://doi.org/10.1038/s41467-022-28470-x
76 McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol* **16**, e2005970 (2018). https://doi.org/10.1371/journal.pbio.2005970
77 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018). https://doi.org/10.1038/nbt.4096
78 Plaisier, C. L. *et al.* Causal Mechanistic Regulatory Network for Glioblastoma Deciphered Using Systems Genetics Network Analysis. *Cell Syst* **3**, 172-186 (2016). https://doi.org/10.1016/j.cels.2016.06.006

79 Reiss, D. J., Plaisier, C. L., Wu, W. J. & Baliga, N. S. cMonkey2: Automated, systematic, integrated detection of co-regulated gene modules for any organism. *Nucleic Acids Res* **43**, e87 (2015). https://doi.org/10.1093/nar/gkv300
80 Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605-D612 (2021). https://doi.org/10.1093/nar/gkaa1074
81 Lachmann, A., Giorgi, F. M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **32**, 2233-2235 (2016). https://doi.org/10.1093/bioinformatics/btw216
82 Aten, J. E., Fuller, T. F., Lusis, A. J. & Horvath, S. Using genetic markers to orient the edges in quantitative trait networks: the NEO software. *BMC Syst Biol* **2**, 34 (2008). https://doi.org/10.1186/1752-0509-2-34