

Divergent mutational processes distinguish hypoxic and normoxic tumours

Vinayak Bhandari^{1,2}, Constance H. Li^{1,2}, Robert G. Bristow^{3,4,5,§}, Paul C. Boutros^{1,2,6,7,8,9,10,11,§}, on behalf of the PCAWG Network

¹ Department of Medical Biophysics, University of Toronto, Toronto, Canada

² Informatics and Biocomputing Program, Ontario Institute for Cancer Research, Toronto, Canada

³ Division of Cancer Sciences, Faculty of Biology, Health and Medicine, University of Manchester, Manchester, UK

⁴ The Christie NHS Foundation Trust, Manchester, UK

⁵ CRUK Manchester Institute and Manchester Cancer Research Centre, Manchester, UK

⁶ Department of Pharmacology and Toxicology, University of Toronto, Toronto, Canada

⁷ Vector Institute for Artificial Intelligence, Toronto, Canada

⁸ Department of Human Genetics, University of California, Los Angeles, USA

⁹ Department of Urology, University of California, Los Angeles, USA

¹⁰ Jonsson Comprehensive Cancer Centre, University of California Los Angeles, Los Angeles, USA

¹¹ Institute for Precision Health, University of California Los Angeles, Los Angeles, USA

[§] Co-corresponding authors

Dr. Paul C. Boutros
12-109 CHS
10833 Le Conte Avenue
Los Angeles, CA, 90095
Email: pboutros@mednet.ucla.edu
Phone: 310-794-7160

Dr. Robert G. Bristow
Manchester Cancer Research Centre, University of Manchester
555 Wilmslow Road
Manchester, UK
M20 4GJ
Email: robert.bristow@manchester.ac.uk
Phone: +44(0)-161-306-3249

Abstract

Many primary tumours have low levels of molecular oxygen (hypoxia). Hypoxic tumours are more likely to metastasize to distant sites and respond poorly to multiple therapies. Surprisingly, then, the pan-cancer molecular hallmarks of tumour hypoxia remain poorly understood, with limited understanding of its associations with specific mutational processes, non-coding driver genes and evolutionary features. To fill this gap, we quantified hypoxia in 1,188 tumours spanning 27 cancer types. We show that elevated hypoxia is associated with increased mutational load across cancers, irrespective of the underlying mutational class. The proportion of mutations attributed to several mutational signatures of unknown aetiology are directly associated with the level of hypoxia, suggesting underlying mutational processes for these signatures. At the gene level, driver mutations in *TP53*, *MYC* and *PTEN* are enriched in tumours with high hypoxia, and mutations in *PTEN* interact with hypoxia to direct the evolutionary trajectory of tumours. Overall, this work demonstrates that hypoxia plays a critical role in shaping the genomic landscape of cancer.

Introduction

Approximately half of all solid tumours are characterized by low levels of molecular oxygen (hypoxia)¹⁻⁴. Sub-regions of hypoxia can result from disrupted oxygen supply: irregular and disorganized tumour vasculature can reduce oxygen availability⁵. Hypoxia can also be caused by changes in oxygen demand: altered tumour metabolism^{6,7} can increase intra-cellular demand for oxygen, potentially extending hypoxia signalling to liquid tumours. The adaptation of tumour cells to this imbalance in oxygen supply and demand is associated with poor clinical prognosis in several cancer types, attributed at least in part to hypoxia-associated genomic instability and clonal selection⁸⁻¹⁶.

Previous work has provided insight into the molecular origins and consequences of tumour hypoxia and genomic instability. Dynamic cycling of hypoxia can select for cells with *TP53* mutations and for those that are apoptosis-deficient^{17,18}. Indeed mutations in *TP53* occur at a higher frequency in hypoxic primary tumours of at least 9 types¹⁶. The abundance of proteins involved in homologous recombination (*e.g.* RAD51) and non-homologous end joining (*e.g.* Ku70) are reduced under hypoxia, and these changes can persist for two days after reoxygenation¹⁹⁻²¹. Genes central to efficient mismatch repair (*e.g.* *MLH1* and *MSH2*) are also downregulated under hypoxia^{22,23}. Further, co-presence of tumour hypoxia and high genomic instability^{14,15}, specific cellular morphologies like intraductal and cribriform carcinoma²⁴ or specific mutations like loss of *PTEN*¹⁶, synergistically predict for rapid relapse after definitive local therapy in some tumour types, particularly prostate cancer. These data underscore the relationship between hypoxia and DNA repair defects, and suggest the tumour microenvironment applies a selective pressure leading to the development of specific genomic profiles.

We previously evaluated the exomic and copy-number alteration (CNA) consequences of tumour hypoxia across 19 cancer types¹⁶. However, the influence of tumour hypoxia on pan-cancer driver alterations, mutational signatures and subclonal architectures remains unclear. To fill this gap, we calculated tumour hypoxia scores for 1,188 tumours with whole-genome and RNA-sequencing sequencing, spanning 27 cancer types. This high-quality harmonized dataset represents a powerful hypothesis-generating mechanism to suggest useful back-translational *in vitro* experiments and better define the hypoxia-associated mutator phenotype. We associated hypoxia with

key driver alterations in coding and non-coding regions of the genome, and find hypoxia is associated with specific mutational signatures of unknown aetiology. We illustrate the joint impact of *PTEN* and the tumour microenvironment in influencing the evolutionary trajectory of tumours. Overall, these data highlight the genomic changes through which hypoxia drives aggressive cancers.

Results

The pan-cancer landscape of tumour hypoxia

We compiled a cohort of 1,188 tumours from 27 cancer types in the Pan-Cancer Analysis of Whole Genomes (PCAWG) dataset with matched tumour/normal whole-genome sequencing and tumour RNA sequencing data. Whole-genome sequencing²⁵ and RNA-sequencing²⁶ analyses were systematically carried out by centralized teams with consistent bioinformatics pipelines. Normal samples had a mean whole-genome sequencing coverage of 30 reads per base-pair while coverage for tumour samples had a bimodal distribution with maxima at 38 and 60 reads per base-pair²⁵. All samples underwent an extensive and systematic quality assurance process²⁵.

We used linear mixed-effect models to associate hypoxia with features of interest across cancers while adjusting for age. Cancer type and sex were further incorporated as random effects in every model, allowing us to consider a different baseline value for the feature of interest for each cancer type and sex²⁷. As a measure of effect size, we report R_{LMM}^2 values which reflect the variance explained by the fixed and random factors in each model²⁸.

We scored tumour hypoxia in all 1,188 tumours using a trio of mRNA-based hypoxia signatures from Buffa²⁹, Winter³⁰ and Ragnum³¹ (**Figure 1a, Supplementary Figure 1a-b, Supplementary Table 1-2**). Hypoxia scores from each of these independent signatures were strongly correlated ($\rho = 0.71 - 0.88$, all $p < 2.2 \times 10^{-16}$, AS89; **Supplementary Figure 1c**) and consistently predicted squamous tumours of the lung (Lung-SCC), cervix (Cervix-SCC) and head (Head-SCC) as the most hypoxic (**Supplementary Figure 1d-e**). Comparatively, chronic lymphocytic leukemias (Lymph-CLL) and thyroid adenocarcinomas (Thy-AdenoCA) were the least hypoxic, consistent with previous¹⁶ reports ($\rho = 0.94$, $p < 2.2 \times 10^{-16}$, AS89; **Figure 1b, Supplementary Figure 1f-h**). Remarkably, subsets of patients from 23/27 cancer types have tumours with elevated hypoxia (hypoxia score > 0) and tumours

consistently have elevated hypoxia compared to normal tissues (**Supplementary Figure 2a-c**).

Considering the strong agreement between the Winter, Buffa and Ragnum hypoxia signatures (**Figure 1a, Supplementary Figure 1c-d**), we used the Buffa signature for subsequent analyses. We first assessed the degree of inter-tumoural heterogeneity in hypoxia that lies within individual cancer types rather than between them. Over 42% of the variance in hypoxia scores occurs within individual cancer types, highlighting the microenvironmental diversity between tumours arising in a single tissue. This variability in hypoxia score within cancer types was especially elevated in some tumour types, particularly biliary adenocarcinomas (interquartile range, IQR = 43.0; Biliary-AdenoCA), mature B-cell lymphomas (IQR = 36.0; Lymph-BNHL), lung adenocarcinomas (IQR = 34.0; Lung-AdenoCA) and breast adenocarcinomas (IQR = 32.0; Breast-AdenoCA). This was in contrast to chronic lymphocytic leukemias (IQR = 2.0; Lymph-CLL) and prostate adenocarcinomas (IQR = 6.0; Prost-AdenoCA) where little inter-tumoural variability in hypoxia was observed. The variability in hypoxia score was not significantly associated with the median hypoxia score within cancer types ($\rho = 0.20$, $p = 0.30$, AS89; **Supplementary Figure 2d**) or with sample size ($\rho = 0.22$, $p = 0.27$, AS89; **Supplementary Figure 1e**). Overall, extensive heterogeneity exists in hypoxia levels within and across cancer types.

The genomic correlates of tumour hypoxia

To determine whether genomic instability arising from specific mutational classes is associated with hypoxia, we looked to identify hypoxia-associated pan-cancer mutational density and summary features³². We first considered as a positive control the percentage of the genome with a copy-number aberration (PGA), an engineered feature that is a surrogate for genomic instability and is associated with hypoxia across several tumour types¹⁶ (**Supplementary Figure 2f**). Indeed, in this diverse pan-cancer cohort, hypoxic tumours have elevated genomic instability while controlling for cancer type, age and sex²⁷ ($p = 5.01 \times 10^{-8}$, $R^2_{LME} = 0.57$, linear mixed-effect model; **Figure 2a**).

We then considered the association of hypoxia scores with 14 other metrics of the mutation density of CNAs, structural variants (SVs) and single nucleotide variants (SNVs) using linear mixed-effect models (**Figure 2a, Supplementary Figure 2f, Supplementary Table 3-4**). The strongest single correlate of tumour hypoxia was the

total number of deletions, where patients with elevated hypoxia had more deletions ($p = 1.31 \times 10^{-10}$, $R^2_{LMEM} = 0.59$, linear mixed-effect model). Elevated numbers of other structural variants such as duplications ($p = 2.82 \times 10^{-4}$, $R^2_{LMEM} = 0.60$, linear mixed-effect model) and truncations ($p = 2.58 \times 10^{-3}$, $R^2_{LMEM} = 0.59$, linear mixed-effect model) were also associated with high hypoxia. Other features associated with elevated hypoxia include smaller CNAs ($p = 2.22 \times 10^{-3}$, $R^2_{LMEM} = 0.59$, linear mixed-effect model) and more SNVs/Mbp ($p = 7.60 \times 10^{-3}$, $R^2_{LMEM} = 0.60$, linear mixed-effect model). Overall, hypoxia is associated with increased numbers of most types of somatic mutations.

Considering the strong association of hypoxia with mutational density, we next looked to determine if these were only general effects or selectively affected specific genes or chromosome regions. We leveraged a catalog of 653 driver mutations³³, with CNA, SV and SNV data available for 1,096 patients. In cases where a patient had multiple mutations in the same gene (*e.g.* a CNA and an SNV) we denoted these as compound events. We again used linear mixed-effect models to associate hypoxia with each driver feature across cancers (**Figure 2b**). Adjusting for cancer type, age and sex, 10 driver events were associated with hypoxia across cancers (FDR < 0.10, linear mixed-effect models; **Supplementary Figure 2f, Supplementary Table 5**). Tumours with mutations in *BCL2* (FDR = 7.09×10^{-15} , $R^2_{LMEM} = 0.61$, linear mixed-effect model) showed lower levels of hypoxia compared to those without. All alterations of *BCL2* in this cohort were SVs, so it is important to note that this association could not be identified from previous exome-sequencing data. Similarly, mutations in the tumour suppressor *TP53* were associated with elevated hypoxia (FDR = 1.58×10^{-12} , $R^2_{LMEM} = 0.58$, linear mixed-effect model), consistent with previous descriptions of hypoxia-mediated selection of *TP53*-mutated cells¹⁷ and elevated hypoxia in breast cancer patients with *TP53* mutations¹⁶. Mutations of the oncogene *MYC* (FDR = 1.09×10^{-4} , $R^2_{LMEM} = 0.59$, linear mixed-effect model) and tumour suppressor *PTEN* (FDR = 1.80×10^{-2} , $R^2_{LMEM} = 0.59$, linear mixed-effect model) were also associated with elevated hypoxia. Thus, hypoxia is associated with both broad elevation of mutation density of most types of somatic variation, along with a consistent signature of alterations in oncogenes and tumour suppressors across cancers.

Hypoxia associated mutational signatures

Previous work has used nonnegative matrix factorization to identify distinct mutational processes in cancer cells from endogenous and exogenous agents³⁴. To identify hypoxia-associated mutational processes, we tested if hypoxia score was associated with the proportion of mutations attributed to each mutational signature using linear mixed-effect models. Of the 65 single base substitution (SBS) signatures tested, 9 showed differential activity in hypoxic tumours compared to non-hypoxic ones while controlling for cancer type, age and sex (FDR < 0.10, linear mixed-effect models; **Figure 3a, Supplementary Table 6**). Of these, six were more active and three less active in tumours with elevated hypoxia. Since previous work has shown that DNA repair is impaired under hypoxia, it was not surprising to observe that a higher proportion of mutations were attributed to SBS3 in tumours with elevated hypoxia score (FDR = 2.01×10^{-3} , $R_{LME}^2 = 0.60$, linear mixed-effect model). Further, SBS6 (FDR = 2.01×10^{-3} , $R_{LME}^2 = 0.61$, linear mixed-effect model) and SBS21 (FDR = 3.58×10^{-2} , $R_{LME}^2 = 0.60$, linear mixed-effect model), both related to defective DNA mismatch repair, had a higher proportion of attributed mutations with increasing hypoxia. A lower proportion of mutations were also attributed to SBS1, previously related to the deamination of 5-methylcytosine, with increasing hypoxia (FDR = 2.11×10^{-7} , $R_{LME}^2 = 0.61$, linear mixed-effect model).

Intriguingly, hypoxia was also associated with a number of SBS signatures with unknown aetiology (**Figure 3b**). The strongest of these was SBS5, where elevated hypoxia was associated with a significantly lower proportion of mutations attributed to the signature (FDR = 1.33×10^{-6} , $R_{LME}^2 = 0.59$ linear mixed-effect model). A significantly lower proportion of mutations were also attributed to SBS12 with increasing hypoxia score (FDR = 3.65×10^{-2} , $R_{LME}^2 = 0.60$, linear mixed-effect model). In contrast, a higher proportion of mutations were attributed to SBS17a (FDR = 3.38×10^{-3} , $R_{LME}^2 = 0.61$, linear mixed-effect model) and SBS17b (FDR = 2.01×10^{-3} , $R_{LME}^2 = 0.61$, linear mixed-effect model) with increasing hypoxia.

Analysis of small insertion and deletion (ID) signatures illustrated a similar story. Of the 17 ID signatures analyzed, the activity of 5 was associated with tumour hypoxia scores while controlling for cancer type, age and sex (FDR < 0.10, linear mixed-effect models; **Figure 3b, Supplementary Table 7**). Of these, 3 were more active in tumours with elevated hypoxia while 2 were less active in them. The defective

homologous recombination signature ID6 (FDR = 6.61×10^{-5} , $R^2_{LMEM} = 0.60$, linear mixed-effect model) and defective DNA mismatch repair signature ID2 (FDR = 4.34×10^{-3} , $R^2_{LMEM} = 0.61$, linear mixed-effect model) had a higher proportion of attributed mutations as hypoxia score increased. Several signatures with unknown aetiology were also significantly associated with hypoxia score, including ID5 (FDR = 1.26×10^{-3} , $R^2_{LMEM} = 0.59$, linear mixed-effect model) and ID9 (FDR = 4.34×10^{-3} , $R^2_{LMEM} = 0.60$, linear mixed-effect model). These data suggest that oxygen levels play a direct or indirect role in the accumulation of specific mutations in cancer cells that are reflected by these signatures.

The subclonal hallmarks of tumor hypoxia

State-of-the-art methods for subclonal reconstruction rely on whole-genome sequencing data³⁵, making the PCAWG dataset an ideal place to understand the evolutionary pressures imposed by hypoxia. Our group and others have shown that some mutations consistently occur early during tumorigenesis while others occur later and that hypoxia is associated with CNAs occurring early in localized prostate cancer^{16,36,37}. To explore if this interaction between the tumour microenvironment and mutational landscape exists more broadly in cancer, we assessed if hypoxia was related to the number of clonal or subclonal mutations across 1,188 tumours from 27 cancer types³⁷. Clonal mutations are common to all cells in a tumour, while subclonal ones are only present in a subpopulation of cells. We found that elevated hypoxia was significantly associated with an increased number of clonal alterations across cancers (Bonferroni-adjusted $p = 4.51 \times 10^{-3}$, $R^2_{LMEM} = 0.60$, linear mixed-effect model; **Figure 4a**), particularly clonal structural variants ($p = 3.19 \times 10^{-6}$, $R^2_{LMEM} = 0.59$, linear mixed-effect model). In contrast, tumour hypoxia was not significantly associated with the number of subclonal alterations (Bonferroni-adjusted $p = 0.15$, $R^2_{LMEM} = 0.60$, linear mixed-effect model; **Figure 4a**). Further, consistent with previous findings in prostate cancer¹⁶, hypoxia was not associated with the number of subclones detected (Bonferroni-adjusted $p = 0.80$, $R^2_{LMEM} = 0.59$, linear mixed-effect model; **Figure 4a**). These data suggest that hypoxia applies a selective pressure on tumours during early tumour development, prior to subclonal diversification.

Next, we assessed if the mutational background of a tumour together with its oxygenation level was linked to its evolutionary trajectory. We previously demonstrated that patients with hypoxic polyclonal prostate tumours with loss of the

tumour suppressor *PTEN* tend to have a poor prognosis¹⁶. Indeed, here we observed a significant interaction between tumor hypoxia and loss of *PTEN* in predicting subclonal architecture ($p_{\text{interaction}} = 3.05 \times 10^{-3}$, $R_{LME}^2 = 0.56$, linear mixed-effect model; **Figure 4b**). Specifically, tumours with both of these features tend to have a polyclonal architecture across cancers. The downstream impact of this interaction between the genome and the tumour microenvironment was observed in RNA data: tumours with both altered *PTEN* and elevated hypoxia had the lowest abundance of *PTEN* mRNA ($p = 5.24 \times 10^{-14}$, $R_{LME}^2 = 0.47$, linear mixed-effect model; **Figure 4c**). Thus, the evolutionary trajectory of a tumour may be driven by the presence of a mutation in a specific microenvironmental context (**Figure 5**).

Discussion

Hypoxia is a feature of many solid and liquid tumours and is associated with aggressive disease. We calculated hypoxia scores for 1,188 tumours from 27 cancer types and showed the vast heterogeneity that exists in this microenvironmental feature within and across cancer types. This reinforces previous pushes for careful patient selection in prospective trials of hypoxia-targeting agents¹⁶. Further, this work prompts the consideration of basket trials for hypoxia targeting agents to help patients with elevated hypoxia across several cancer types.

For the first time, we characterized the pan-cancer whole-genome correlates of tumour hypoxia. We show the broad influence of the hypoxia associated mutator phenotype: elevated hypoxia is associated with increased mutational load across all mutational classes (*i.e.* CNAs, SVs and SNVs). This supports previous *in vitro* work that demonstrated the contextual synthetic lethality of PARP inhibition in cells with defective DNA repair due to hypoxia³⁸. Regarding this co-occurrence of genomic instability and hypoxia, our group¹⁶ and others³⁹ have previously described this metabolic reprogramming as a series of distinct genomic alterations. This is supported by our finding that alterations in *TP53*, *MYC* and *PTEN* are more common in tumours with elevated hypoxia across cancers. Our study cannot however conclusively say whether hypoxia exerts a selective pressure that enriches for specific genomic alterations or if these genomic changes directly result in hypoxia. Experimental studies of single genes support that both effects may contribute to the associations we describe^{17,22,40-42}.

Diving into the mutational processes related to hypoxia, we confirmed that several single base substitution and small indel signatures related to impaired DNA repair were associated with hypoxia. This raises the potential confounder that because hypoxic tumours have more mutations, we have more power to detect related mutational signatures. However, we demonstrated that hypoxia is indeed strongly associated with many mutational signatures with unknown aetiology, particularly SBS5, which is found in nearly all cancer types. Modelling these associations *in vitro* is particularly difficult and these data provide a high confidence measure of the mutational signatures that may be directly or indirectly driven by tumour oxygen levels. It is difficult to disentangle the timing of these events: whether a specific driver mutation gives rise to a specific mutational signature or if these are separate processes. Better mapping of the evolutionary timing of hypoxia will be particularly important in addressing this question and the advent of hypoxia signatures may facilitate future studies in this area.

We observed a significant association between elevated hypoxia and the number of clonal mutations. This supports the idea that hypoxia is an early event in cancer, as we have suggested previously¹⁶, and other models that link hypoxia to genomic instability and downstream clonal selection^{20,41}. Previous work has also demonstrated that patients with allelic loss of *PTEN* and elevated hypoxia rapidly relapse after definitive treatment for localized prostate cancer¹⁶. Here, we showed that tumours with alterations in *PTEN* and elevated hypoxia are enriched for a polyclonal tumour architecture. This illustrates the joint influence of the tumour mutational landscape and microenvironment in guiding evolutionary trajectories across cancers. Further, these data suggest that increased subclonal diversification may be a novel route *via* which *PTEN* drives aggressive tumour phenotypes, in concert with tumour hypoxia, and this can be better defined with future back-translational *in vitro* experiments. Overall, this work shows that a hypoxic tumour microenvironment is associated with specific mutational processes and distinct somatic mutational profiles, and may direct the subclonal architecture of cancers.

Methods

Pan-cancer hypoxia scoring

Hypoxia scores were calculated for all 1,188 tumours with mRNA abundance data using mRNA-abundance based signatures of tumour hypoxia developed previously by Winter *et al.*,³⁰ Buffa *et al.*²⁹ and Ragnum *et al.*³¹, as described previously^{14,16} (**Supplementary Table 2**). Briefly, patients with the top 50% of mRNA abundance values for each gene in a signature were given a score of +1. Patients with the bottom 50% of mRNA abundance values for that gene were given a score of -1. This was repeated for every gene in the signature to generate a hypoxia score for each patient, and this process was repeated for each of the three signatures used in the study. High scores suggest that the tumour was hypoxic and low scores are indicative of normoxia.

Hypoxia Score Comparison

To compare hypoxia scores generated by the different signatures, the median hypoxia score was calculated for each of the PCAWG cancer types based on each signature. The median hypoxia scores from each signature were then scaled from +1 to -1 using the *plotrix* package (v3.7). Scaled median hypoxia values for the PCAWG cancer types were also compared to scaled median hypoxia values from previously published¹⁶ TCGA data between comparable cancer groups. The groups compared are as follows (PCAWG cancer type TCGA cancer type): Bladder-TCC and BLCA; Breast-AdenoCA and BRCA; Cervix-SCC and CESC; CNS-GBM and GBM; ColoRect-AdenoCA and COADREAD; Head-SCC and HNSC; Kidney-RCC and KIRC; Liver-HCC and LIHC; Lung-AdenoCA and LUAD; Lung-SCC and LUSC; Ovary-AdenoCA and OV; Panc-AdenoCA and PAAD; Prost-AdenoCA and PRAD; Skin-Melanoma and SKCM; Thy-AdenoCA and THCA; Uterus-Adeno and UCEC. Of the 27 cancer types in PCAWG, 11 (Cervix-AdenoCA, Stomach-AdenoCA, Eso-AdenoCA, Breast-LobularCA, SoftTissue-Leiomyo, Lymph-BNHL, SoftTissue-Liposarc, Biliary-AdenoCA, Kidney-ChRCC, CNS-Oligo and Lymph-CLL) did not have hypoxia data from comparable cancers in TCGA and were not used for the comparison¹⁶. For Spearman's correlations, p-values were calculated using algorithm AS89.

Linear Mixed-Effect Models

We used linear mixed-effect models to associate hypoxia with features of interest (*e.g.*, PGA, *TP53* mutational status, etc.) across cancers using the lme4 package (v1.1-17). For each feature of interest we compared a full model (*i.e.*, a model with the feature of interest) to a null model (*i.e.* a model without the feature of interest) using an ANOVA to determine if hypoxia was significantly associated with the feature of interest across cancers:

$$full = hypoxia \sim feature + age + (1|cancer) + (1|sex)$$

$$null = hypoxia \sim age + (1|cancer) + (1|sex)$$

All models were adjusted for patient age. Cancer type and sex were used as random effects in every model. This allows us to consider a different baseline value for the feature of interest for each cancer type and sex²⁷. For each model an R^2_{LME} value is reported, reflecting the variance explained by the fixed and random factors²⁸.

All model diagnostics were done using the DHARMA package (0.2.0) which uses a simulation-based approach to create standardized residuals⁴³. For each model, scaled residuals were generated using the simulateResiduals function. The full model was used as the input for fittedModel parameter and 1,000 simulations were run. For correctly specified models, the scaled residuals were expected to be uniformly distributed and this was verified for each full model. We also compared the standardized residuals to the rank transformed predicted values to assess deviations from uniformity for each full model.

Mutational Density Analysis

Previously published data for 15 mutational density and summary features were downloaded for 1,188 tumours³². We used linear mixed-effect models to associate each feature with hypoxia score across cancers and compared each full model with a null model. Cancer type and sex were used as random effect variables. Tumours belonging to cancer types with fewer than 15 samples were excluded from the analysis. A Bonferroni p-value adjustment was applied to the p-values from linear mixed-effect modeling since fewer than 20 tests were conducted.

Driver Mutations Analysis

Data for driver mutations was first summarized at the gene level for 1,096 tumours with previously published driver mutation data³³. For each of the 653 driver features,

we summarized if a patient had an SNV, CNA or SV. Some tumours had more than one type of event in a gene (*e.g.* a CNA and an SNV) and these events were classified as compound events. We then used linear mixed-effect models to associate the mutational status of each gene with hypoxia score and compared each full model with a null model. Cancer type and sex were used as random effect variables. The driver mutation analysis did not specifically consider the type of mutation in the gene and only considered if the gene had a mutation or was wildtype. Tumours belonging to cancer types with fewer than 15 samples were excluded from the analysis. An FDR adjustment was applied to the p-values from linear mixed-effect modeling.

Mutational Signature Analysis

Previously published data for mutations attributed to various specific signatures was downloaded for 1,188 tumours³⁴. For each tumour, we calculated the proportion of total mutations attributed to each mutational signature. The proportion of mutations attributed to each signature were calculated by dividing the number of mutations attributed to each signature by the total number of mutations in the tumour. We used linear mixed-effect models to associate the proportion of mutations attributed to each signature with hypoxia score and compared each full model with a null model. Cancer type and sex were used as random effect variables. Tumours belonging to cancer types with fewer than 15 samples were excluded from the analysis. An FDR adjustment was applied to the p-values from linear mixed-effect modeling.

Subclonality Analysis

Previously reported³⁷ subclonal reconstruction data was used to summarize the number of clonal and subclonal mutations in all 1,188 tumours. We used linear mixed-effect models to associate the number of these timed mutations with hypoxia score and compared each full model with a null model. Cancer type and sex were used as random effect variables. Tumours belonging to cancer types with fewer than 15 samples were excluded from the analysis. A Bonferroni adjustment was applied to the p-values from linear mixed-effect modeling since fewer than 20 tests were conducted.

The number of subclones was calculated for all 1,188 tumours based on the number of clusters of cells identified in each sample. A linear mixed-effects model were used to associate the number of subclones with hypoxia score and this model was compared to a null model. Cancer type was used as a random effect. Tumours belonging to cancer types with fewer than 15 samples were excluded from the analysis.

Patients with only one identified cluster of cells were defined as monoclonal and patients with more than one identified cluster of cells were defined as polyclonal³⁶. Hypoxia scores were median dichotomized to classify patients as hypoxic or normoxic. To test for an interaction between tumour hypoxia and *PTEN* mutational status in selecting for a particular subclonal architecture, we used linear mixed-effect models together with an ANOVA. A full model was first created where the relationship between the hypoxia scores and *PTEN* mutational status was modelled as an interaction. A reduced model was also created where the relationship between hypoxia scores and *PTEN* mutational status was modelled in an additive manner:

$$full = clonality \sim hypoxia * PTEN + age + (1|cancer) + (1|sex)$$

$$reduced = clonality \sim hypoxia + PTEN + age + (1|cancer) + (1|sex)$$

The two models were compared using an ANOVA to test if hypoxia scores significantly interact with *PTEN* mutational status. Tumours belonging to cancer types with fewer than 15 samples were excluded from the analysis. The full model diagnostics were carried out using the DHARMA package, as described above.

All data analysis was performed in the R statistical environment (v3.4.3). Data visualization was performed using the BPG package⁴⁴ (v5.9.1). Figures were compiled using Inkscape (v0.91).

Authors' contributions

Bioinformatics and Statistical Analysis: VB

Data Processing: VB, CHL

Data Visualization: VB

Supervised Research: RGB, PCB

Initiated the Project: VB, PCB

Wrote the First Draft of the Manuscript: VB

Approved the Manuscript: All Authors

Conflict of Interest Statement

All authors declare that they have no conflicts of interest.

Funding Sources

This study was conducted with the support of Movember funds through Prostate Cancer Canada, and with the additional support of the Ontario Institute for Cancer Research, funded by the Government of Ontario. This work was supported by Prostate Cancer Canada and is proudly funded by the Movember Foundation Team Grant T2013 and #RS2014-01. Paul C. Boutros was supported by a Terry Fox Research Institute New Investigator Award, a Prostate Cancer Canada Rising Star Fellowship, a CIHR New Investigator Award, a CIHR Project Grant, the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-125) and Canadian Cancer Society (grant #705649). This work has been funded by a fellowship from the Canadian Institutes of Health Research to Vinayak Bhandari. Constance Li was supported by an award from the Government of Ontario. The authors gratefully thank the Princess Margaret Cancer Centre Foundation and Radiation Medicine Program Academic Enrichment Fund for support (to Robert G. Bristow). This work was supported by a Terry Fox Research Institute Program Project Grant. Robert G. Bristow is a recipient of a Canadian Cancer Society Research Scientist Award.

Acknowledgements

The authors gratefully thank Dr. Marianne Koritzinsky for insightful suggestions and thank Jenna E. van Leeuwen for visualization support. The authors also thank all

members of the Boutros and Bristow labs for helpful suggestions.

References

1. Weber, C. E. & Kuo, P. C. The tumor microenvironment. *Surg. Oncol.* **21**, 172–177 (2012).
2. Blagosklonny, M. V. Antiangiogenic therapy and tumor progression. *Cancer Cell* **5**, 13–7 (2004).
3. Vaupel, P., Thews, O. & Hoeckel, M. Treatment Resistance of Solid Tumors. *Med. Oncol.* **18**, 243–260 (2001).
4. Dhani, N., Fyles, A., Hedley, D. & Milosevic, M. The Clinical Significance of Hypoxia in Human Cancers. *Semin. Nucl. Med.* **45**, 110–121 (2015).
5. Brown, J. M. & Wilson, W. R. Exploiting tumour hypoxia in cancer treatment. *Nat. Rev. Cancer* **4**, 437–47 (2004).
6. Zannella, V. E. *et al.* Reprogramming Metabolism with Metformin Improves Tumor Oxygenation and Radiotherapy Response. *Clin. Cancer Res.* **19**, 6741–6750 (2013).
7. Mucaj, V., Shay, J. E. S. & Simon, M. C. Effects of hypoxia and HIFs on cancer metabolism. *Int. J. Hematol.* **95**, 464–470 (2012).
8. Rouschop, K. M. A. *et al.* The unfolded protein response protects human tumor cells during hypoxia through regulation of the autophagy genes MAP1LC3B and ATG5. *J. Clin. Invest.* **120**, 127–141 (2010).
9. Lassen, P. *et al.* HPV-associated p16-expression and response to hypoxic modification of radiotherapy in head and neck cancer. *Radiother. Oncol.* **94**, 30–35 (2010).
10. Hoskin, P. J., Rojas, A. M., Bentzen, S. M. & Saunders, M. I. Radiotherapy With Concurrent Carbogen and Nicotinamide in Bladder Carcinoma. *J. Clin. Oncol.* **28**, 4912–4918 (2010).
11. Milosevic, M. *et al.* Tumor hypoxia predicts biochemical failure following radiotherapy for clinically localized prostate cancer. *Clin. Cancer Res.* **18**, 2108–14 (2012).
12. Kim, B. *et al.* Prognostic assessment of hypoxia and metabolic markers in cervical cancer using automated digital image analysis of

- immunohistochemistry. *J. Transl. Med.* **11**, 185 (2013).
13. Yang, L. *et al.* Development and Validation of a 28-gene Hypoxia-related Prognostic Signature for Localized Prostate Cancer. *EBioMedicine* (2018). doi:10.1016/J.EBIOM.2018.04.019
 14. Lalonde, E. *et al.* Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. *Lancet. Oncol.* **15**, 1521–32 (2014).
 15. Lalonde, E. *et al.* Translating a Prognostic DNA Genomic Classifier into the Clinic: Retrospective Validation in 563 Localized Prostate Tumors. *Eur. Urol.* **72**, 22–31 (2017).
 16. Bhandari, V. *et al.* Molecular landmarks of tumor hypoxia across cancer types. *Nat. Genet.* (2019). doi:10.1038/s41588-018-0318-2
 17. Graeber, T. G. *et al.* Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours. *Nature* **379**, 88–91 (1996).
 18. Greijer, A. E. & van der Wall, E. The role of hypoxia inducible factor 1 (HIF-1) in hypoxia induced apoptosis. *J. Clin. Pathol.* **57**, 1009–14 (2004).
 19. Bindra, R. S. *et al.* Down-Regulation of Rad51 and Decreased Homologous Recombination in Hypoxic Cancer Cells. *Mol. Cell. Biol.* **24**, 8504–8518 (2004).
 20. Bristow, R. G. & Hill, R. P. Hypoxia and metabolism: Hypoxia, DNA repair and genetic instability. *Nat. Rev. Cancer* **8**, 180–192 (2008).
 21. Meng, A. X. *et al.* Hypoxia down-regulates DNA double strand break repair gene expression in prostate cancer cells. *Radiother. Oncol.* **76**, 168–176 (2005).
 22. Mihaylova, V. T. *et al.* Decreased expression of the DNA mismatch repair gene Mlh1 under hypoxic stress in mammalian cells. *Mol. Cell. Biol.* **23**, 3265–73 (2003).
 23. Koshiji, M. *et al.* HIF-1 α Induces Genetic Instability by Transcriptionally Downregulating MutS α Expression. *Mol. Cell* **17**, 793–803 (2005).
 24. Chua, M. L. K. *et al.* A Prostate Cancer “Nimbusus”: Genomic Instability and SchLAP1 Dysregulation Underpin Aggression of Intraductal and Cribriform Subpathologies. *Eur. Urol.* **71**, 183–192 (2017).

25. Campbell, P. J. *et al.* Pan-cancer analysis of whole genomes. *bioRxiv* 162784 (2017). doi:10.1101/162784
26. PCAWG Transcriptome Core Group *et al.* Genomic basis for RNA alterations revealed by whole-genome analyses of 27 cancer types. *bioRxiv* 183889 (2018). doi:10.1101/183889
27. Li, C. H., Haider, S., Shiah, Y.-J., Thai, K. & Boutros, P. C. Sex Differences in Cancer Driver Genes and Biomarkers. *Cancer Res.* **78**, 5527–5537 (2018).
28. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **4**, 133–142 (2013).
29. Buffa, F. M., Harris, A. L., West, C. M. & Miller, C. J. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br. J. Cancer* **102**, 428–35 (2010).
30. Winter, S. C. *et al.* Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res.* **67**, 3441–9 (2007).
31. Ragnum, H. B. *et al.* The tumour hypoxia marker pimonidazole reflects a transcriptional programme associated with aggressive prostate cancer. *Br. J. Cancer* **112**, 382–90 (2015).
32. Yousif, F. *et al.* The Origins and Consequences of Localized and Global Somatic Hypermutation. *bioRxiv* 287839 (2018). doi:10.1101/287839
33. Sabarinathan, R. *et al.* The whole-genome panorama of cancer drivers. *bioRxiv* 190330 (2017). doi:10.1101/190330
34. Alexandrov, L. *et al.* The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv* 322859 (2018). doi:10.1101/322859
35. Deshwar, A. G. *et al.* PhyloWGS: Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* **16**, 35 (2015).
36. Espiritu, S. M. G. *et al.* The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. *Cell* **173**, 1003–1013 (2018).
37. Gerstung, M. *et al.* The evolutionary history of 2,658 cancers. *bioRxiv* 161562 (2017). doi:10.1101/161562

38. Chan, N. *et al.* Contextual synthetic lethality of cancer cell kill based on the tumor microenvironment. *Cancer Res.* **70**, 8045–54 (2010).
39. Haider, S. *et al.* Genomic alterations underlie a pan-cancer metabolic shift associated with tumour hypoxia. *Genome Biol.* **17**, 140 (2016).
40. Nguyen, A. D., McDonald, J. G., Bruick, R. K. & DeBose-Boyd, R. A. Hypoxia Stimulates Degradation of 3-Hydroxy-3-methylglutaryl-coenzyme A Reductase through Accumulation of Lanosterol and Hypoxia-Inducible Factor-mediated Induction of Insigs. *J. Biol. Chem.* **282**, 27436–27446 (2007).
41. Luoto, K. R., Kumareswaran, R. & Bristow, R. G. Tumor hypoxia as a driving force in genetic instability. *Genome Integr.* **4**, 5 (2013).
42. Chen, Y., Cairns, R., Papandreou, I., Koong, A. & Denko, N. C. Oxygen consumption can regulate the growth of tumors, a new perspective on the Warburg effect. *PLoS One* **4**, e7033 (2009).
43. Dunn, P. K. & Smyth, G. K. Randomized Quantile Residuals. *J. Comput. Graph. Stat.* **5**, 236 (1996).
44. P'ng, C. *et al.* BPG: Seamless, automated and interactive visualization of scientific data. *BMC Bioinformatics* **20**, 42 (2019).

Figure Legends

Figure 1 – The pan-cancer landscape of tumour hypoxia

We quantified tumour hypoxia in 1,188 tumours spanning 27 different cancer types.

a) Hypoxia scores for 27 types of cancer, based on the Buffa mRNA abundance signature. Cancer types are sorted by the median hypoxia score (horizontal black line) for each cancer type. Each dot represents one tumour. Sample sizes for each cancer type are listed near the bottom along with the percent of tumours that have elevated hypoxia (hypoxia score > 0). The variability in hypoxia within cancer types was measured by the interquartile range (IQR), shown along the bottom. The IQR was particularly high in biliary adenocarcinoma (IQR = 43.0; Biliary-AdenoCA), mature B-cell lymphomas (IQR = 36.0; Lymph-BNHL), lung adenocarcinoma (IQR = 34.0; Lung-AdenoCA) and breast adenocarcinoma (IQR = 32; Breast-AdenoCA). By contrast, chronic lymphocytic leukemia (IQR = 2.0; Lymph-CLL) and thyroid adenocarcinoma (IQR = 11.0; Thy-AdenoCA) showed less variance in hypoxia score.

b) Analysis of hypoxia between 16 comparable cancer types in PCAWG and TCGA. Dots represent the mean of the scaled median hypoxia scores from three different mRNA-based hypoxia signatures. Error bars represent the standard deviation of the scaled median hypoxia scores. Overall, the pan-cancer quantification of hypoxia between the PCAWG and TCGA datasets shows strong agreement.

Figure 2 – The genomic correlates of tumour hypoxia

We associated tumour hypoxia with mutational density/summary features (**a**) and driver mutations (**b**) across 27 cancer types using linear mixed-effect models. Hypoxia scores for all 1,188 tumours are shown along the top. **a)** Elevated tumour hypoxia was strongly associated with more deletions, elevated PGA, smaller CNAs and a higher number of SNVs per megabase ($n = 1,188$ independent tumours). Bonferroni-adjusted p-values are shown on the right. **b)** We tested if driver mutations (*e.g.* any of SNV, CNA, SV or a compound event with more than one type of mutation) were associated with hypoxia in 1,096 patients with driver mutation data. Tumours with mutations in *BCL2* showed lower levels of hypoxia while patients with mutations in *TP53* showed remarkably elevated tumour hypoxia. Other driver mutations associated with elevated hypoxia include the oncogene *MYC* and the tumour suppressor *PTEN*. FDR-adjusted p-values are shown along the right. SV,

structural variant; PGA, percentage of the genome with a copy-number aberration; CNA, copy-number aberration; SNV, single nucleotide variant; H-H, head-to-head; T-T, tail-to-tail.

Figure 3 – Hypoxia associated mutational signatures

We associated hypoxia with the proportion of mutations attributed to specific mutational signatures using linear mixed-effect models. Hypoxia scores for 1,188 tumours are shown across the top while FDR-adjusted p-values are shown on the right. **a)** Hypoxia was associated with a series of single base substitution signatures with unknown aetiology including SBS5, SBS17a, SBS17b and SBS12. Some of these mutational signatures may reflect hypoxia-dependent mutational processes. Hypoxia was also associated with a lower proportion of attributed mutations to SBS1, which reflects deamination of 5-methylcytosine, and a higher proportion of attributed mutations to SBS3, which is related to deficiencies in DNA double-strand break repair and homologous recombination. **b)** Several signatures of small insertions and deletions were also associated with hypoxia, including ID6 and ID2, which reflect defective homologous recombination and defective DNA mismatch repair, respectively. ID5, ID9 and ID4, all with unknown aetiology, were significantly associated with hypoxia score.

Figure 4 – The subclonal hallmarks of tumor hypoxia

We associated tumour hypoxia with features related to the subclonal architecture of 1,188 tumours from 27 cancer types using linear mixed-effect models. **a)** Hypoxia scores are shown along the top while Bonferroni-adjusted p-values are shown on the right. Hypoxia was not associated with the number of subclones in the tumour but elevated hypoxia was associated with a higher number of clonal mutations. **b)** We also observed a significant interaction between hypoxia and altered *PTEN* where tumours with both of these features were particularly likely to be polyclonal. **c)** The mRNA abundance of *PTEN* is modulated by both *PTEN* mutational status and tumour hypoxia. Tumours with altered *PTEN* and elevated hypoxia have the lowest abundance of *PTEN* mRNA.

Figure 5 – Altered *PTEN* and hypoxia may drive subclonal diversification.

Many primary tumours have elevated hypoxia due to increased demand or decreased supply of oxygen. Tumours with elevated hypoxia tend to have altered *PTEN*.

Elevated hypoxia and altered *PTEN* may drive subclonal diversification and poor outcomes.

Supplementary Figure Legends

Supplementary Figure 1 – Pan-cancer landscape of hypoxia in primary tumours

Tumour hypoxia scores for 1,188 tumours from 27 cancer types based on the Winter³⁰ (a) and Ragnum³¹ (b) hypoxia signatures. The black horizontal lines indicate the median hypoxia score within each cancer type. The number of tumours from each cancer type are shown along the bottom. IQR values for each cancer type are also shown along the bottom. c) Hypoxia scores based on the three independent signatures were highly correlated for 1,188 tumours from 27 cancer types. d) Comparison of median hypoxia scores within cancer types based on the Buffa²⁹, Winter³⁰ and Ragnum³¹ hypoxia signatures. Median hypoxia values were scaled from +1 to -1. e) The mean of the scaled median hypoxia scores for each cancer type are shown by the filled circle. Lines show the standard deviation of the scaled median hypoxia score. Squamous tumours of the lung (Lung-SCC), cervix (Cervix-SCC) and head (Head-SCC) are amongst the most hypoxic cancer types. f-h) Association of hypoxia scores between 16 comparable types of cancer in PCAWG and TCGA. Dots represent the mean of the scaled median hypoxia scores from the Buffa²⁹ (f), Winter³⁰ (g) and Ragnum³¹ (h) hypoxia signatures. Pan-cancer calculations of hypoxia between the PCAWG and TCGA datasets is strongly correlated based on independent hypoxia signatures.

Supplementary Figure 2 – Hypoxia in tumour vs. normal samples and the genomics of hypoxia across cancers

Tumour samples consistently have elevated hypoxia compared to normal samples from the same tissue based on the Buffa (a), Ragnum (b) and Winter (c) hypoxia signatures. The IQR values of cancer types were not associated with the median hypoxia score within that cancer type (d) or the sample size (e). f) A partial summary of the knowledge around the genomics of hypoxic tumours. Hypoxia-related associations that have been previously examined within 19 tumour types in TCGA are shown along the left based on data from Bhandari *et al.*¹⁶. CNAs and SNVs in several genes were found to be associated with elevated hypoxia within tumour types. Hypoxia was also associated with elevated PGA in 10 tumour types. The right side of the figure partially summarizes the analyses carried out in this work. Several of the previous intra-tumour type findings have been extended in this work as pan-cancer

features of hypoxia and several novel features, particularly related to structural variants, have been assessed and found to be significantly associated with hypoxia. This is in addition to the novel pan-cancer work presented around single-base substitution signatures, indel signatures and tumour subclonality.

Supplementary Table Legends

Supplementary Table 1 – PCAWG cancer type codes

Cancer type codes for the 27 PCAWG cancer types examined in this study and their descriptions.

Supplementary Table 2 – Hypoxia scores and genomic data for 1,188 tumours

Data for hypoxia scores, mutational density/summary, driver mutations, mutational signatures and subclonality for 1,188 tumours from 27 cancer types. For driver mutations, 0 represents wildtype, 1 represents an SNV, 2 represents a CNA, 3 represents an SV and 4 represents a compound event (*i.e.*, those with multiple types of alterations). Mutational signature data for single base substitution signatures and insertion and deletion (ID) signatures are provided as the proportion of alterations attributed to the signature.

Supplementary Table 3 – Mutational density data by deciles

For each of the 15 mutational features analyzed, the value corresponding to each decile is provided.

Supplementary Table 4 – Hypoxia associated mutational density features

Results from linear mixed-effect models associating hypoxia with mutational density features while controlling for cancer type, age and sex.

Supplementary Table 5 – Hypoxia associated driver mutations

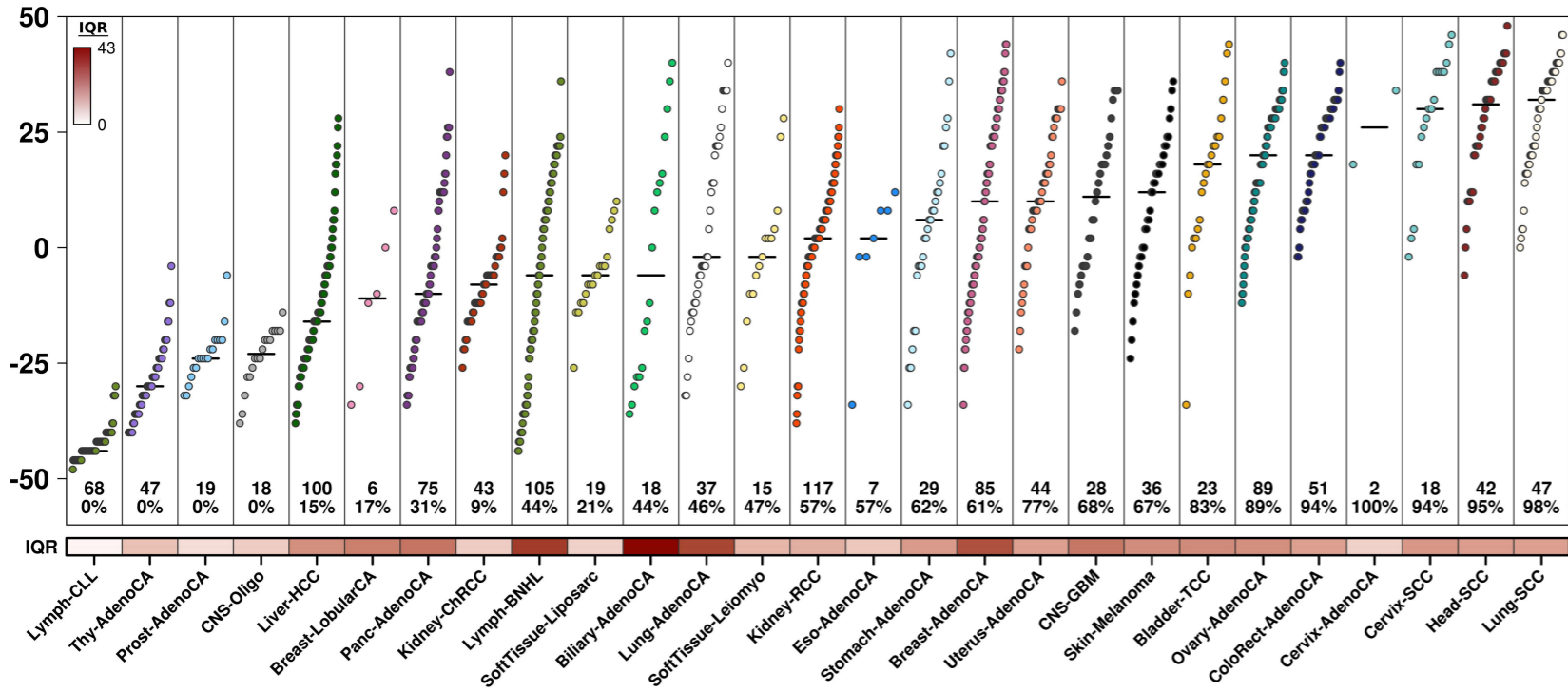
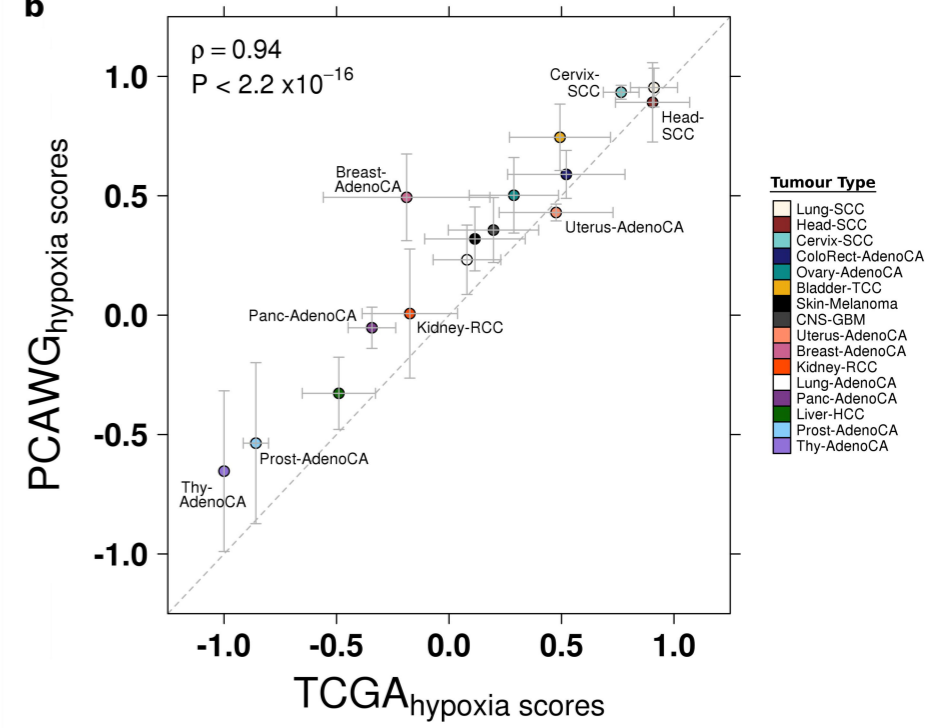
Results from linear mixed-effect models associating hypoxia with driver mutations while controlling for cancer type, age and sex.

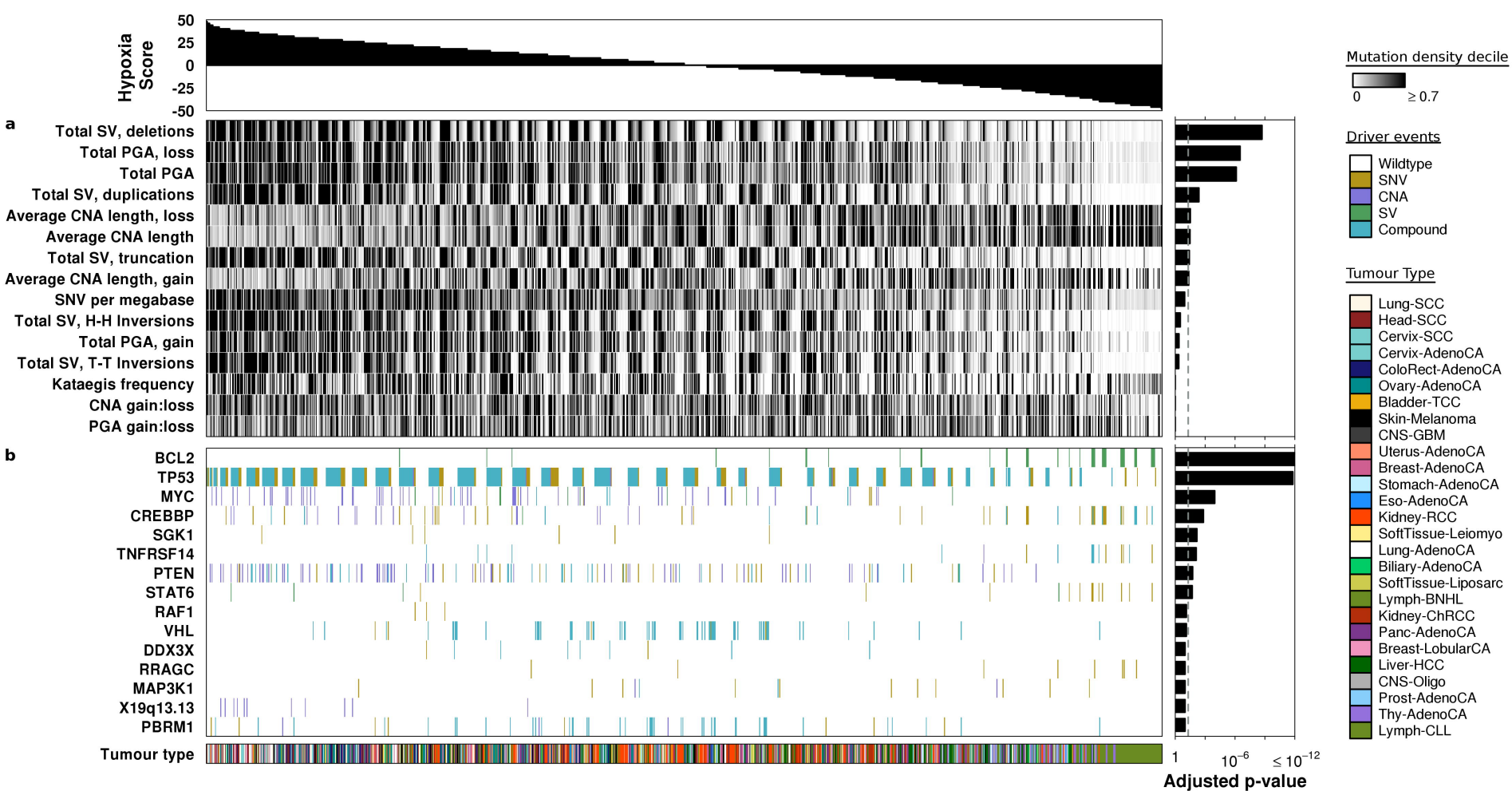
Supplementary Table 6 – Hypoxia associated single base substitution signatures

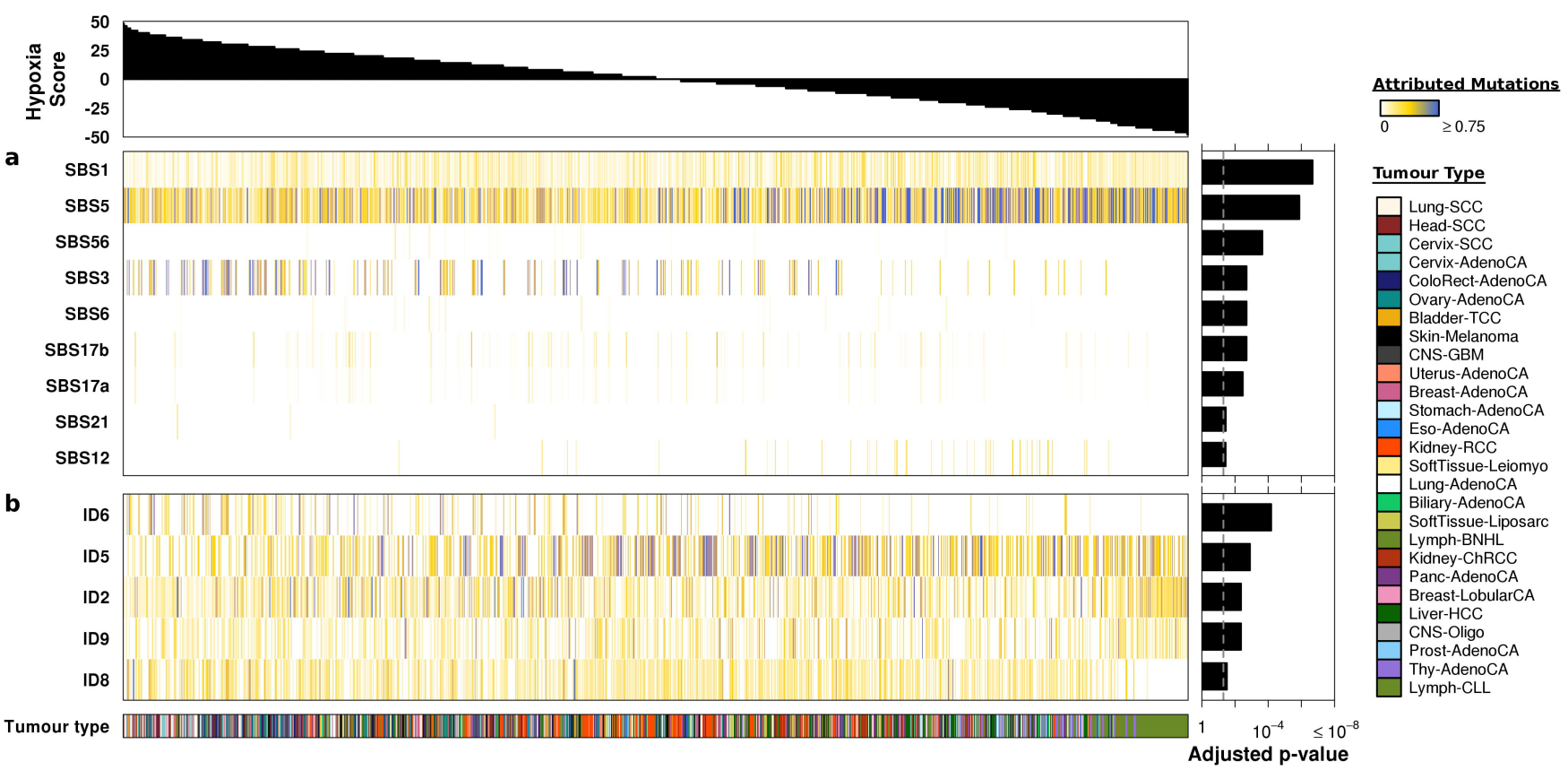
Results from linear mixed-effect models associating hypoxia with the proportion of mutations attributed to single base substitution signatures while controlling for cancer type, age and sex.

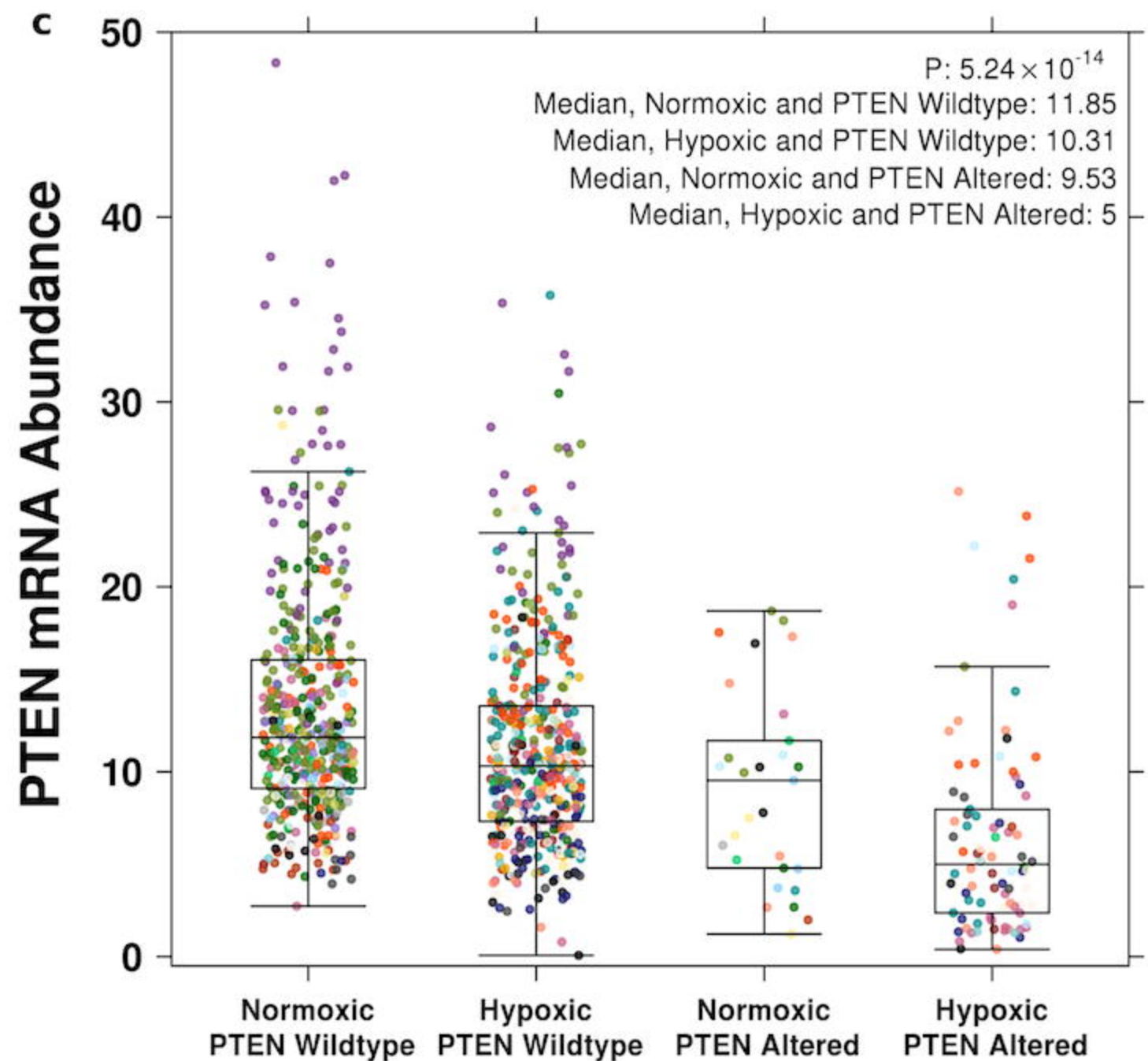
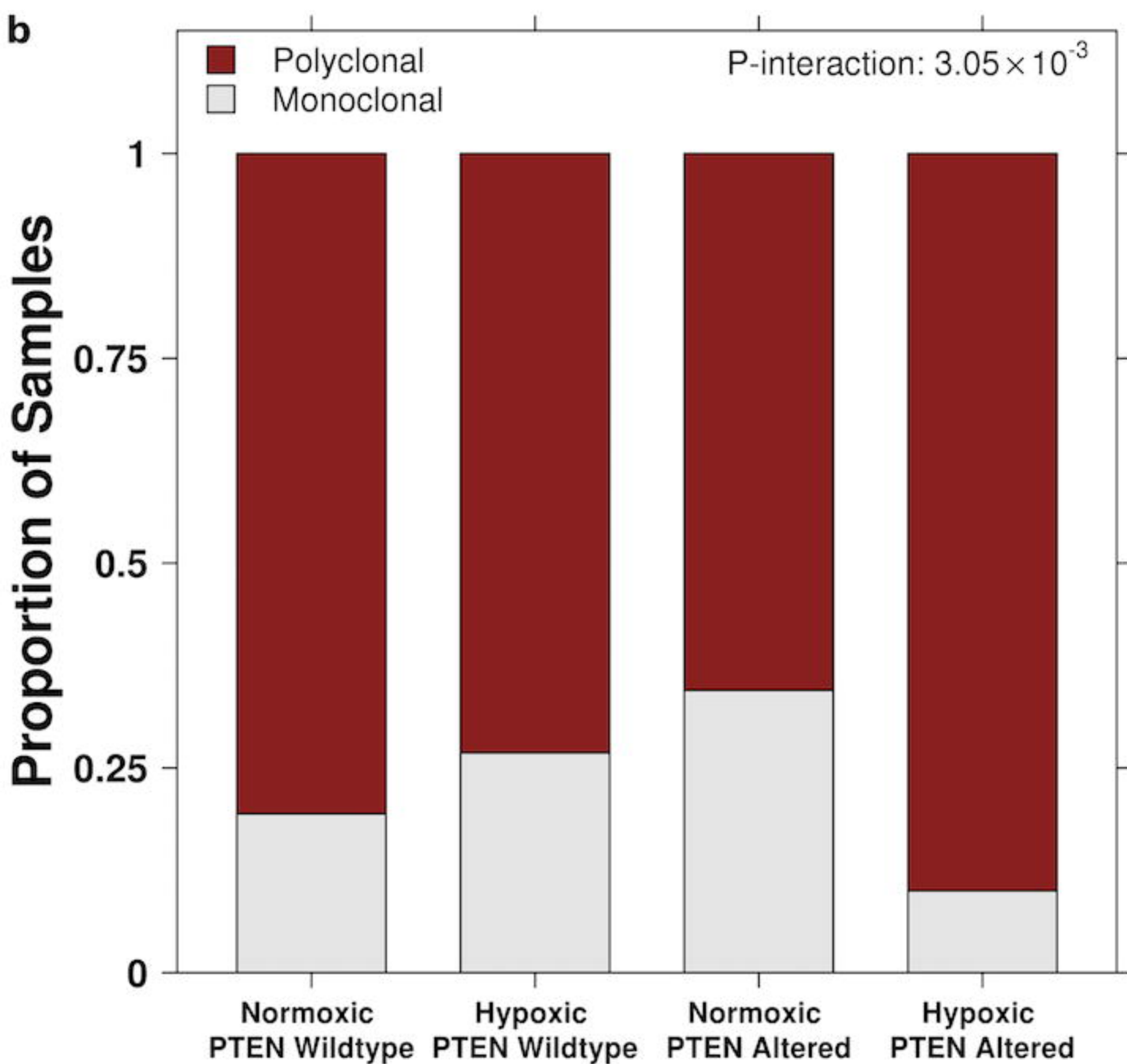
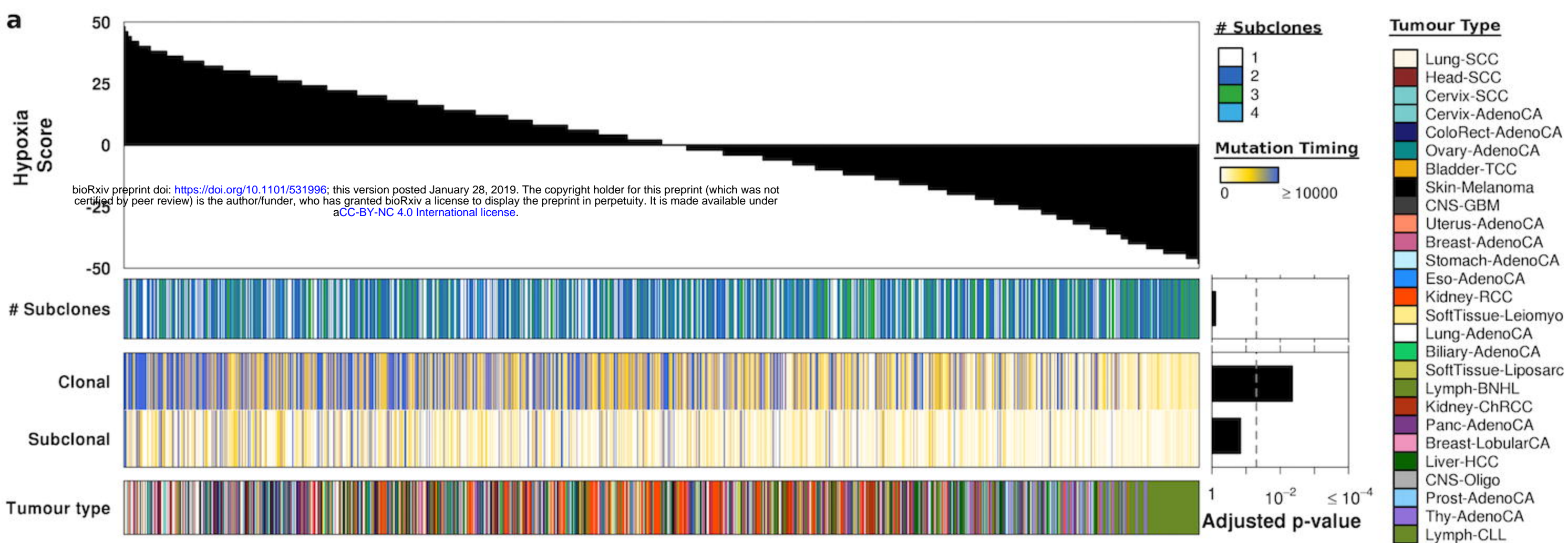
Supplementary Table 7 – Hypoxia associated small insertion and deletion signatures

Results from linear mixed-effect models associating hypoxia with the proportion of mutations attributed to small insertion and deletion signatures while controlling for cancer type, age and sex.

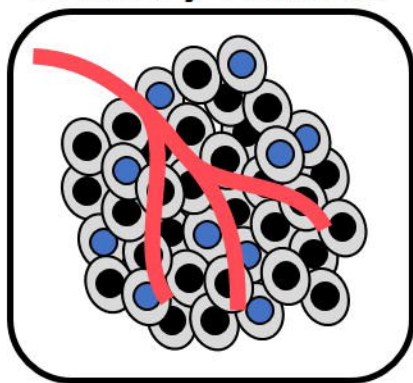
a**b**







Primary Tumour



Normoxia



Hypoxia

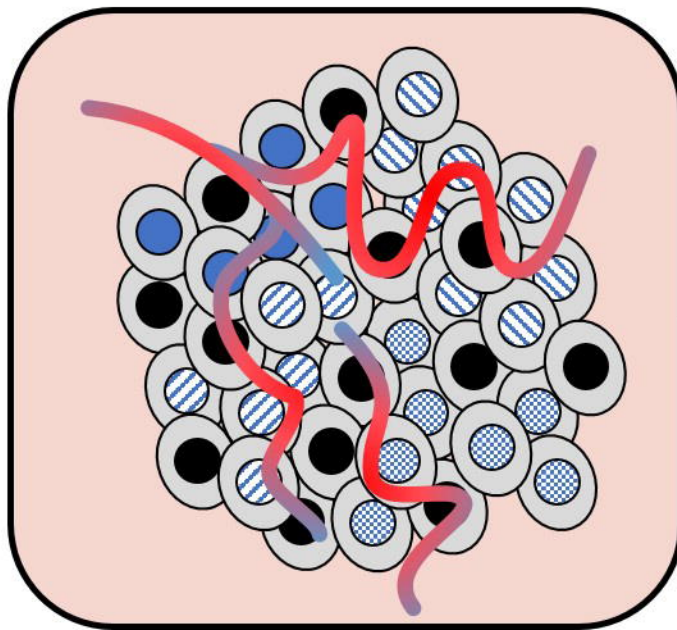
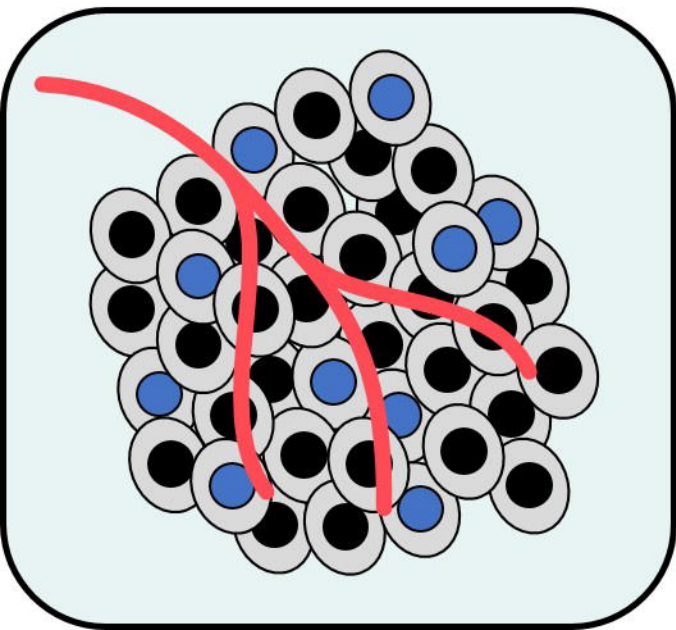







Figure Legend

-  Blood Vessel
-  PTEN^{wt} tumour cell
-    PTEN^{mut} subclones

Altered PTEN and hypoxia drive subclonal diversification