





Divergent mutational processes distinguish hypoxic and normoxic tumours

Vinayak Bhandari ^{1,2}, Constance H. Li ^{1,2,3}, Robert G. Bristow ^{4,5,6*}, Paul C. Boutros ^{1,3,7,8,9,10,11*} & PCAWG Consortium

Many primary tumours have low levels of molecular oxygen (hypoxia), and hypoxic tumours respond poorly to therapy. Pan-cancer molecular hallmarks of tumour hypoxia remain poorly understood, with limited comprehension of its associations with specific mutational processes, non-coding driver genes and evolutionary features. Here, as part of the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium, which aggregated whole genome sequencing data from 2658 cancers across 38 tumour types, we quantify hypoxia in 1188 tumours spanning 27 cancer types. Elevated hypoxia associates with increased mutational load across cancer types, irrespective of underlying mutational class. The proportion of mutations attributed to several mutational signatures of unknown aetiology directly associates 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 hypoxic tumours, and mutations in *PTEN* interact with hypoxia to direct tumour evolutionary trajectories. Overall, hypoxia plays a critical role in shaping the genomic and evolutionary landscapes of cancer.

¹Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada. ²Ontario Institute for Cancer Research, Toronto, Canada. ³Department of Human Genetics, University of California, Los Angeles, USA. ⁴Division of Cancer Sciences, Faculty of Biology, Health and Medicine, University of Manchester, Manchester, UK. ⁵The Christie NHS Foundation Trust, Manchester, UK. ⁶CRUK Manchester Institute and Centre, Manchester, UK. ⁷Department of Pharmacology and Toxicology, University of Toronto, Toronto, Canada. ⁸Vector Institute for Artificial Intelligence, Toronto, Canada. ⁹Department of Urology, University of California, Los Angeles, USA. ¹⁰Jonsson Comprehensive Cancer Centre, University of California Los Angeles, Los Angeles, USA. ¹¹Institute for Precision Health, University of California Los Angeles, Los Angeles, USA. PCAWG Consortium members and their affiliations appears online. *email: robert.bristow@manchester.ac.uk; pboutros@mednet.ucla.edu

Approximately half of all solid tumours are characterized by low levels of molecular oxygen (hypoxia)^{1–4}. Subregions 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 intracellular 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 selection^{8–16}.

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 2 days after reoxygenation^{19–21}. 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 1188 tumours with whole-genome sequencing (WGS) and RNA sequencing, spanning 27 cancer types. Genome sequencing data was aggregated by the Pan-Cancer Analysis of Whole Genomes (PCAWG) consortium and generated by the ICGC and TCGA projects. These sequencing data were re-analyzed with standardized, high-accuracy pipelines to align to the human genome (reference build hs37d5) and identify germline variants and somatic mutations, as described previously²⁵. This sequencing data together with our high-quality hypoxia quantitation represents a powerful hypothesis-generating mechanism to suggest useful back-translational *in vitro* experiments and better define the hypoxia-associated mutator phenotype across cancers. 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 1188 tumours from 27 cancer types via the PCAWG Consortium. All samples had matched tumour and reference normal WGS and tumour RNA sequencing data generated by the ICGC and TCGA projects. WGS²⁵ and RNA-sequencing²⁶ analyses were systematically carried out by centralized teams with consistent and high-accuracy bioinformatics pipelines. Normal reference samples had a mean WGS coverage of 39 reads per

base-pair while coverage for tumour samples had a bimodal distribution with modes 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 tumour purity, age, and sex^{27,28}. Cancer type was further incorporated as a random effect in every model, allowing us to consider a different baseline value for the feature of interest for each cancer type. As a measure of effect size we report conditional R^2 values, denoted as $R^2_{\text{LMEM-C}}$, which reflect the variance explained by the fixed and random factors in each model²⁹. We also report marginal R^2 values, denoted as $R^2_{\text{LMEM-M}}$, which reflect the variance explained only by the fixed factors²⁹.

We scored tumour hypoxia in all 1188 tumours using a trio of mRNA-based hypoxia signatures from Buffa³⁰, Winter³¹ and Ragnum³² (Fig. 1a, Supplementary Fig. 1a, b, Supplementary Table 1, Supplementary Data 1). 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 Fig. 1c) and consistently predicted squamous tumours of the lung (Lung-SCC), cervix (Cervix-SCC), and head (Head-SCC) as the most hypoxic (Supplementary Fig. 1d, e). Comparatively, chronic lymphocytic leukaemias (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; Fig. 1b, Supplementary Fig. 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 Fig. 2a–c).

Considering the strong agreement between the Winter, Buffa and Ragnum hypoxia signatures (Fig. 1a, Supplementary Fig. 1c, d), we used the Buffa signature for subsequent analyses. The Buffa signature has been previously used for pan-cancer analyses and shows results consistent with those from other signatures¹⁶. 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 leukaemias (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 Fig. 2d) or with sample size ($\rho = 0.22$, $p = 0.27$, AS89; Supplementary Fig. 2e). 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 features³³. As a positive control, we first considered 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 Fig. 2f). Indeed, in this diverse pan-cancer cohort,

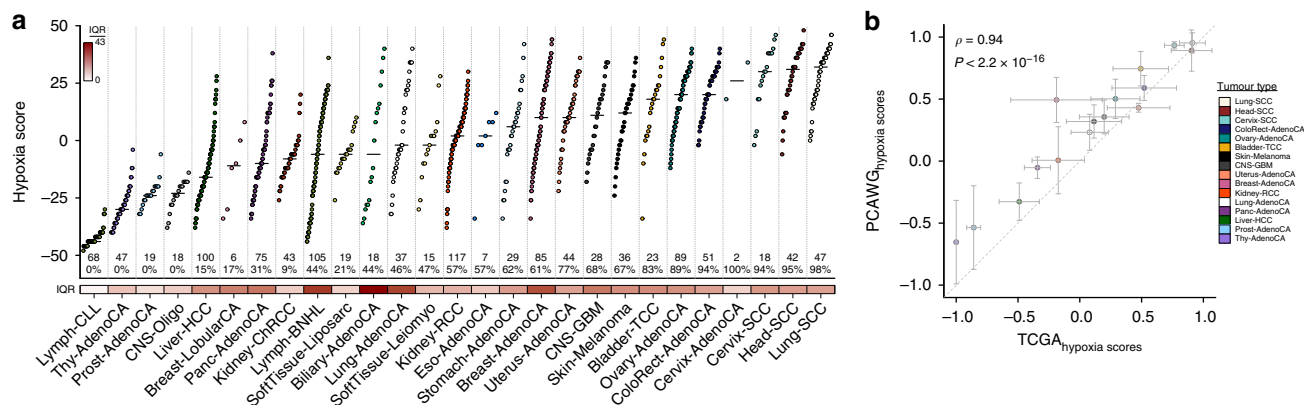


Fig. 1 The pan-cancer landscape of tumour hypoxia. We quantified tumour hypoxia in 1188 independent tumours spanning 27 different cancer types. **a** Hypoxia scores for 27 types of cancer, based on the Bufla mRNA abundance signature. Cancer types are sorted by the median hypoxia score (horizontal black 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), lymphoid B-cell non-Hodgkin's lymphomas (IQR = 36.0; Lymph-BNHL), lung adenocarcinoma (IQR = 34.0; Lung-AdenoCA) and breast adenocarcinoma (IQR = 32; Breast-AdenoCA). By contrast, chronic lymphocytic leukaemia (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 (Spearman's ρ , AS89). 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.

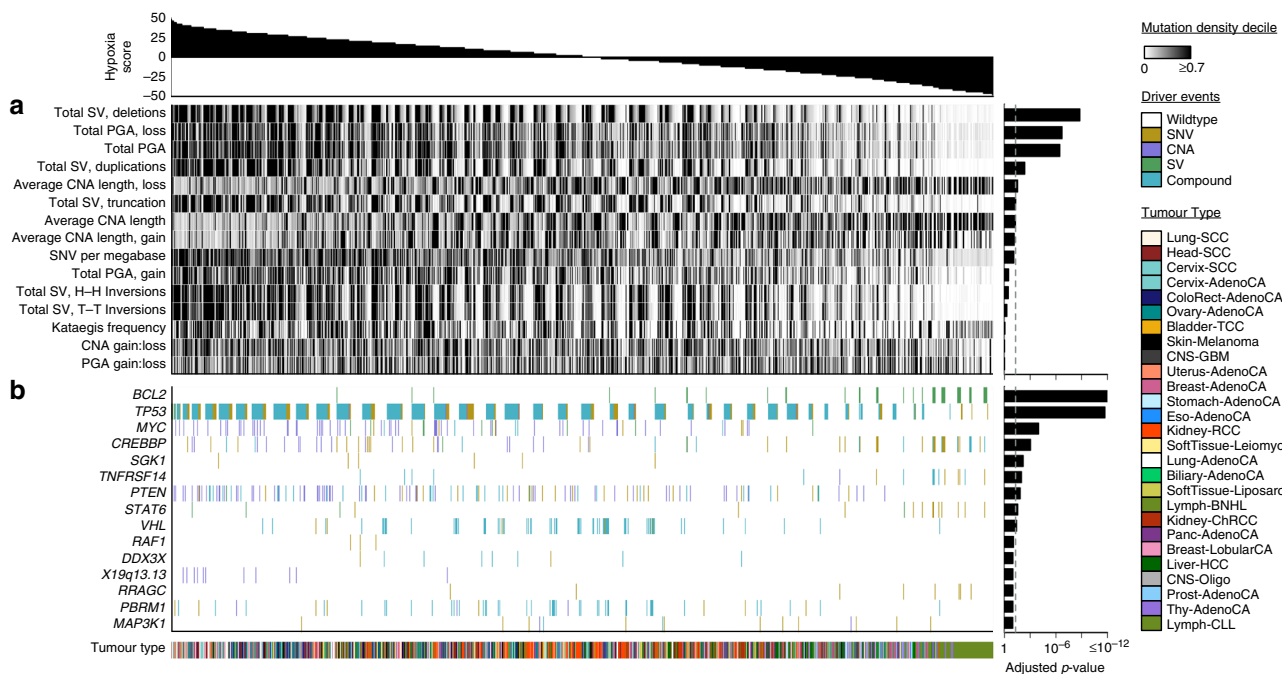


Fig. 2 The genomic correlates of tumour hypoxia. We associated tumour hypoxia with mutational density and summary features, **a**, and driver mutations, **b**, across 27 cancer types using linear mixed-effect models. Hypoxia scores for all 1188 tumours are shown along the top. **a** Elevated tumour hypoxia was strongly associated with more deletions, elevated PGA, smaller CNAs, and a higher number of SNVs per megabase ($n = 1188$ 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 1096 independent tumours 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. All associations were modelled using linear mixed-effect models while adjusting for cancer type, tumour purity, age and sex.

hypoxic tumours have elevated genomic instability while controlling for cancer type, tumour purity, age and sex²⁷ ($p = 2.41 \times 10^{-8}$, $R^2_{MEM-M} = 0.022$, $R^2_{MEM-C} = 0.57$, linear mixed-effect model; Fig. 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 (Fig. 2a, Supplementary Fig. 2f, Supplementary

Tables 2, 3). The strongest single correlate of tumour hypoxia was the total number of deletions, where patients with elevated hypoxia had more deletions ($p = 1.11 \times 10^{-10}$, $R^2_{\text{LMEM-M}} = 0.023$, $R^2_{\text{LMEM-C}} = 0.59$, linear mixed-effect model). Elevated numbers of other SVs such as duplications ($p = 2.94 \times 10^{-4}$, $R^2_{\text{LMEM-M}} = 0.0084$, $R^2_{\text{LMEM-C}} = 0.60$, linear mixed-effect model) and truncations ($p = 3.29 \times 10^{-3}$, $R^2_{\text{LMEM-M}} = 0.0062$, $R^2_{\text{LMEM-C}} = 0.60$, linear mixed-effect model) were also associated with high hypoxia, and we confirmed this within individual cancer types (Supplementary Fig. 3a). Other features associated with elevated hypoxia include smaller CNAs ($p = 3.51 \times 10^{-3}$, $R^2_{\text{LMEM-M}} = 0.0065$, $R^2_{\text{LMEM-C}} = 0.59$, linear mixed-effect model) and more SNVs/Mbp ($p = 5.55 \times 10^{-3}$, $R^2_{\text{LMEM-M}} = 0.0054$, $R^2_{\text{LMEM-C}} = 0.60$, linear mixed-effect model). Since mutational density features can be correlated, we wanted to further test if SNVs per megabase were independently associated with hypoxia after adjusting for the total number of deletions. We created a linear mixed-effect model associating hypoxia with the number of SNVs per megabase while adjusting for cancer type, age, sex, tumour purity and the number of deletions. We also created a second model which lacked our feature of interest, SNVs per megabase, and compared the two models using an ANOVA (see the “Methods” section). The p -value for this comparison was 0.011, suggesting that the number of SNVs per megabase are associated with hypoxia independent of the number of deletions (and other potential confounders included in the models). 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 catalogue of 653 driver mutations²⁵, with CNA, SV and SNV data available for 1096 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 (Fig. 2b). Adjusting for cancer type, tumour purity, age and sex, 10 driver events were associated with hypoxia across cancers (FDR < 0.10, linear-mixed effect models; Supplementary Fig. 2f, Supplementary Table 4). Tumours with mutations in *BCL2* (FDR = 7.56×10^{-15} , $R^2_{\text{LMEM-M}} = 0.045$, $R^2_{\text{LMEM-C}} = 0.62$, 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 across cancers (FDR = 1.97×10^{-12} , $R^2_{\text{LMEM-M}} = 0.043$, $R^2_{\text{LMEM-C}} = 0.59$, linear-mixed effect model), consistent with previous descriptions of hypoxia-mediated selection of *TP53*-mutated cells¹⁷ and elevated hypoxia in breast cancers with *TP53* mutations¹⁶. We also confirmed this association within individual cancer types (Supplementary Fig. 3b). Mutations of the oncogene *MYC* (FDR = 1.07×10^{-4} , $R^2_{\text{LMEM-M}} = 0.016$, $R^2_{\text{LMEM-C}} = 0.60$, linear-mixed effect model) and tumour suppressor *PTEN* (FDR = 1.50×10^{-2} , $R^2_{\text{LMEM-M}} = 0.0098$, $R^2_{\text{LMEM-C}} = 0.59$, linear mixed-effect model) were also associated with elevated hypoxia. Alterations in mitochondrial genes³⁴ were not significantly associated with tumour hypoxia (Supplementary Fig. 3c). 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 agents³⁵. 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, nine showed differential activity in hypoxic tumours compared to non-hypoxic ones, while controlling for cancer type, tumour purity, age and sex (FDR < 0.10, linear mixed-effect models; Fig. 3a, Supplementary Table 5). 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 (related to defective homologous recombination-based repair) in tumours with elevated hypoxia score (FDR = 1.98×10^{-3} , $R^2_{\text{LMEM-M}} = 0.016$, $R^2_{\text{LMEM-C}} = 0.60$, linear-mixed effect model). Further, SBS6 (FDR = 1.98×10^{-3} , $R^2_{\text{LMEM-M}} = 0.0086$, $R^2_{\text{LMEM-C}} = 0.61$, linear-mixed effect model) and SBS21 (FDR = 4.31×10^{-2} , $R^2_{\text{LMEM-M}} = 0.0051$, $R^2_{\text{LMEM-C}} = 0.61$, 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 = 8.33×10^{-8} , $R^2_{\text{LMEM-M}} = 0.033$, $R^2_{\text{LMEM-C}} = 0.61$, linear-mixed effect model).

Intriguingly, hypoxia was also associated with a number of SBS signatures with unknown aetiology (Fig. 3b). The strongest of these was SBS5, where elevated hypoxia was associated with a significantly lower proportion of mutations attributed to the signature (FDR = 1.43×10^{-6} , $R^2_{\text{LMEM-M}} = 0.022$, $R^2_{\text{LMEM-C}} = 0.59$, linear-mixed effect model). A significantly lower proportion of mutations were also attributed to SBS12 with increasing hypoxia score (FDR = 4.31×10^{-2} , $R^2_{\text{LMEM-M}} = 0.0066$, $R^2_{\text{LMEM-C}} = 0.60$, linear-mixed effect model). In contrast, a higher proportion of mutations were attributed to SBS17a (FDR = 4.80×10^{-3} , $R^2_{\text{LMEM-M}} = 0.0072$, $R^2_{\text{LMEM-C}} = 0.61$, linear mixed-effect model) and SBS17b (FDR = 2.83×10^{-3} , $R^2_{\text{LMEM-M}} = 0.079$, $R^2_{\text{LMEM-C}} = 0.61$, 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 were associated with tumour hypoxia scores while controlling for cancer type, tumour purity, age and sex (FDR < 0.10, linear mixed-effect models; Fig. 3b, Supplementary Table 6). 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 = 5.76×10^{-5} , $R^2_{\text{LMEM-M}} = 0.015$, $R^2_{\text{LMEM-C}} = 0.60$, linear mixed-effect model) and defective DNA mismatch repair signature ID2 (FDR = 7.06×10^{-3} , $R^2_{\text{LMEM-M}} = 0.011$, $R^2_{\text{LMEM-C}} = 0.61$, 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 = 1.54×10^{-3} , $R^2_{\text{LMEM-M}} = 0.016$, $R^2_{\text{LMEM-C}} = 0.60$, linear mixed-effect model) and ID9 (FDR = 7.06×10^{-3} , $R^2_{\text{LMEM-M}} = 0.0068$, $R^2_{\text{LMEM-C}} = 0.60$, 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 tumour hypoxia. State-of-the-art methods for subclonal reconstruction rely on WGS data³⁶, making the PCAWG dataset ideal for understanding the evolutionary pressures imposed by hypoxia. We and others have shown that some mutations consistently occur early during tumourigenesis while others occur later and that hypoxia is

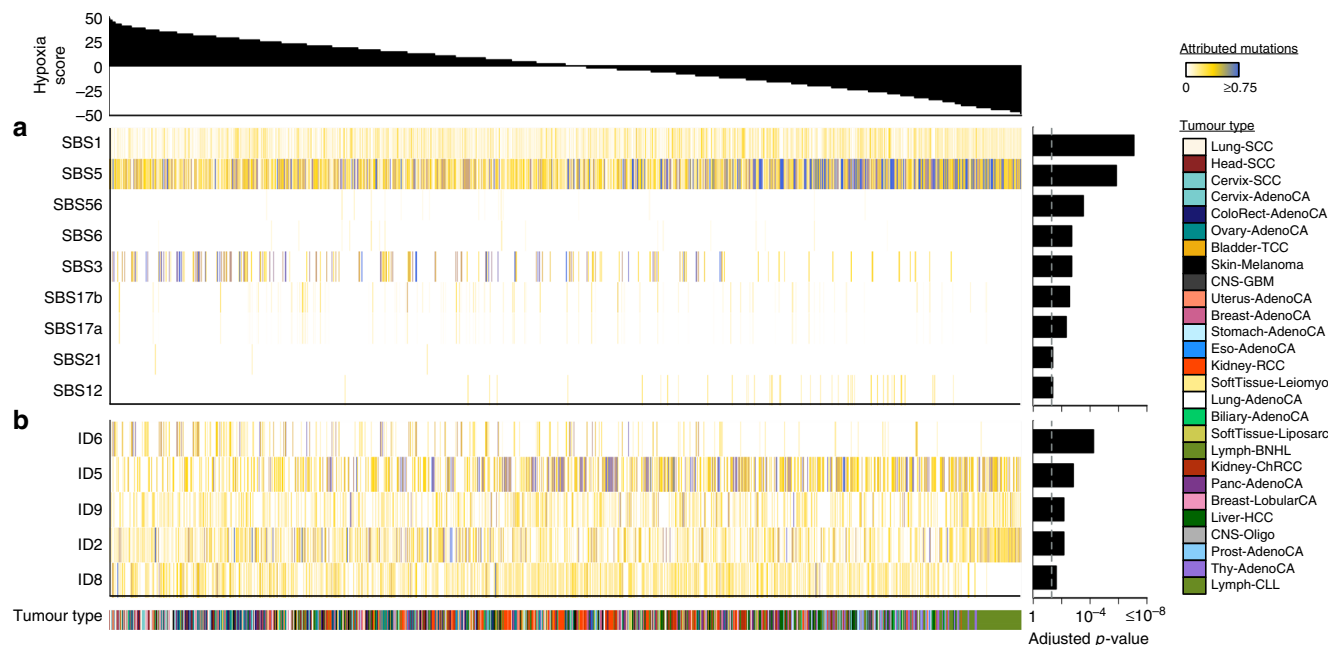


Fig. 3 Hypoxia-associated mutational signatures. We associated hypoxia with the proportion of mutations attributed to specific mutational signatures using linear-mixed effect models. Hypoxia scores for 1188 independent tumours are shown across the top while FDR-adjusted p -values are shown on the right. **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. Hypoxia was also associated with a lower proportion of attributed mutations to SBS5, SBS17a, SBS17b and SBS12, which are associated with 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. All associations were modelled using linear-mixed effect models while adjusting for cancer type, tumour purity, age and sex.

associated with CNAs occurring early in localized prostate cancer^{16,37,38}. 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 1188 tumours from 27 cancer types³⁸. 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 $p = 4.65 \times 10^{-3}$, $R^2_{\text{LMEM-M}} = 0.0074$, $R^2_{\text{LMEM-C}} = 0.60$, linear mixed-effect model; Fig. 4a, Supplementary Table 7), particularly clonal SVs ($p = 1.17 \times 10^{-5}$, $R^2_{\text{LMEM-M}} = 0.013$, $R^2_{\text{LMEM-C}} = 0.60$, linear mixed-effect model). In contrast, tumour hypoxia was not significantly associated with the number of subclonal alterations (Bonferroni-adjusted $p = 0.28$, $R^2_{\text{LMEM-M}} = 0.0039$, $R^2_{\text{LMEM-C}} = 0.60$, linear mixed-effect model; Fig. 4a). Further, consistent with findings in prostate cancer¹⁶, hypoxia was not associated with the number of subclones detected (Bonferroni-adjusted $p = 0.14$, $R^2_{\text{LMEM-M}} = 0.0051$, $R^2_{\text{LMEM-C}} = 0.60$, linear mixed-effect model; Fig. 4a). These data suggest that hypoxia applies a selective pressure on tumours during their early evolution, prior to subclonal diversification.

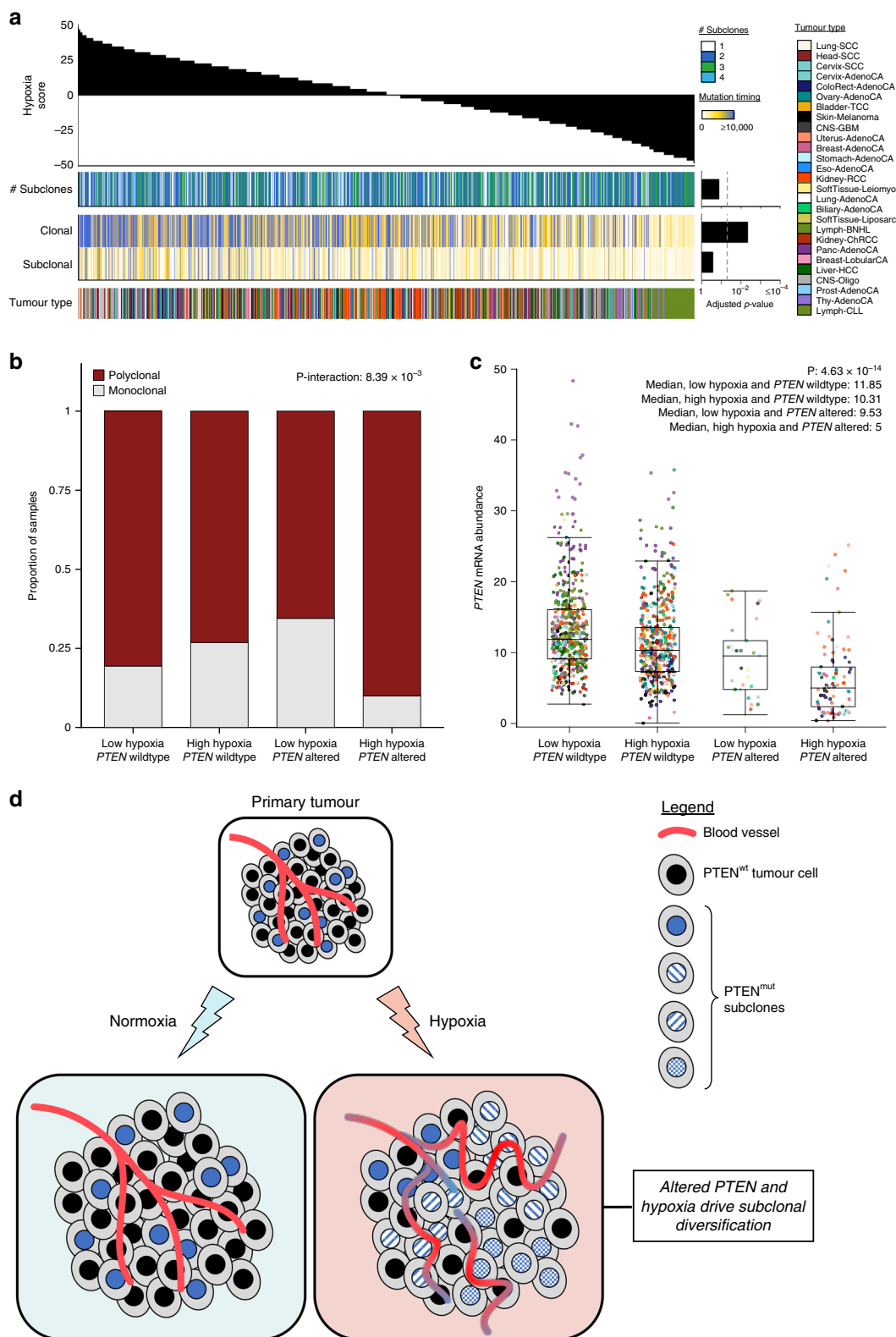
Next, we assessed if the mutational background of a tumour together with its oxygenation level was linked to its evolutionary trajectory. We previously demonstrated that patients with hypoxic polyclonal prostate tumours with loss of the tumour suppressor *PTEN* tend to have a poor prognosis¹⁶. Indeed, here we observed a significant interaction between tumour hypoxia and loss of *PTEN* in predicting subclonal architecture ($p_{\text{interaction}} = 8.39 \times 10^{-3}$, $R^2_{\text{LMEM-M}} = 0.60$, $R^2_{\text{LMEM-C}} = 0.87$, linear mixed-effect model; Fig. 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 ($p = 4.63 \times 10^{-14}$, $R^2_{\text{LMEM-M}} = 0.054$, $R^2_{\text{LMEM-C}} = 0.47$, linear mixed-effect model; Fig. 4c). Thus, the evolutionary trajectory of a tumour may be driven by the presence of a mutation in a specific microenvironmental context (Fig. 4d).

Discussion

Hypoxia is a feature of many solid and liquid tumours and is associated with aggressive disease. We calculated hypoxia scores for 1188 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.

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 hypoxia³⁹. Regarding this co-occurrence of genomic instability and hypoxia, our group¹⁶ and others⁴⁰ 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. Supporting these findings, previous *in vitro* work has



shown that heterogeneous populations of cells where a small subpopulation have mutant *TP53* can rapidly expand under cycling hypoxia to become the major subpopulation due to deficient apoptosis and selection¹⁷. We have also previously shown that tumours with *TP53* mutations have elevated hypoxia within individual breast cancer subtypes, confirming

that this association is not simply reflecting previously described molecular subtypes¹⁶.

Our study cannot 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

Fig. 4 The subclonal hallmarks of tumour hypoxia. We associated tumour hypoxia with features related to the subclonal architecture of 1188 independent 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. mRNA abundance is reported as FPKM with upper-quartile normalization. A Tukey box plot is shown. Box plots represent the median (centre line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5× the interquartile range. All associations were modelled using linear mixed-effect models while adjusting for cancer type, tumour purity, age and sex. **d** 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.

the associations we describe^{17,22,41–43}. While we have not specifically included *in vitro* experimental validation data in this report, we and others have previously validated associations first revealed by analysis of mRNA-based hypoxia signatures. For example, our group previously described microRNA-133a-3p as a hypoxia-associated miRNA in prostate cancer based on mRNA signature-based associations across multiple independent datasets¹⁶. We went on to validate that microRNA-133a-3p was indeed induced under hypoxia in multiple prostate cancer cell lines and confirmed its capacity to modulate cell proliferation and invasion. Similarly, Ye et al. applied hypoxia signatures to 10 independent datasets of cell lines and primary tumour fragments under hypoxia and normoxia⁴⁴. These 10 datasets represented seven cancer types and within each dataset samples under hypoxia showed higher hypoxia scores compared to normoxic samples. Further, they generated predictions of drugs that would be more or less potent under hypoxia and validated four drug-hypoxia interactions *in vitro*. These data illustrate that hypoxia signatures applied to large cohorts of primary tumours can generate reliable hypotheses, many of which have been validated in controlled systems. However, some aerobic cancer cells may also mimic the biological state of hypoxia (i.e., pseudohypoxia) and this may affect signature-derived hypoxia estimates. Further, while pimonidazole (which was used to develop the Ragnum hypoxia signature³²) reflects oxygen tensions below 10 mmHg (1.3% O₂), it is difficult to directly relate hypoxia signature scores with oxygen tension⁴⁵. Overall, hypoxia signalling can be distinct from microenvironmental hypoxia and this remains a critical caveat of this study.

Diving into the mutational processes related to hypoxia, we confirmed that several SBS 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^{20,42}. 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. The PCAWG dataset is the largest publicly available pan-cancer dataset to date and this limits our ability to validate our discoveries in independent datasets. Ultimately, it will be necessary to validate our findings in large, independent cohorts. Hypotheses generated here, particularly those around hypoxia and tumour evolution, will require long term, systematic *in vitro* modelling and will be the subject of future studies. 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 1188 tumours with mRNA abundance data (FPKM with upper-quartile normalization) using mRNA-abundance-based signatures of tumour hypoxia developed previously by Winter et al.³¹, Buffa et al.³⁰ and Ragnum et al.³², as described previously^{14,16} (Supplementary Data 1). 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 *plortrix* 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 and TCGA cancer type): Bladder-TCC and BLCA; Breast-AdenoCA and BRCA; Cervix-SCC and CESC; CNS-GBM and GBM; ColoRect-AdenoCA and COAD-READ; 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.

Tumour vs. normal hypoxia comparison. Previously calculated pan-cancer tumour hypoxia scores were gathered for 7791 independent tumours from 19 cancer types based on the Buffa hypoxia signature¹⁶. Hypoxia scores were then calculated for all normal tissue samples related to the 19 cancer types with hypoxia scores (*n* = 640 independent normal tissue samples). Tumour hypoxia scores were compared to normal tissue hypoxia scores for each tissue type where at least 15 normal tissue samples were available (Mann-Whitney *U*-test). A total of 5649

independent tumours and 625 independent normal tissue samples were evaluated in the comparisons.

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. A generic example of this is shown below with Eqs. (1) and (2):

$$\text{full} = \text{hypoxia} \sim \text{feature} + \text{purity} + \text{age} + \text{sex} + (1|\text{cancer}) \quad (1)$$

$$\text{null} = \text{hypoxia} \sim \text{purity} + \text{age} + \text{sex} + (1|\text{cancer}) \quad (2)$$

All models were adjusted for tumour purity, patient age and sex^{27,28}. Cancer type was incorporated as a random effect in every model. This allowed us to consider a different baseline value for the feature of interest for each cancer type. For each model a conditional R^2 value is reported ($R^2_{\text{LMEM-C}}$) which reflects the variance explained by the fixed and random factors²⁹. We also report marginal R^2 values for each model ($R^2_{\text{LMEM-M}}$) which reflect the variance explained by the fixed factors only²⁹. $R^2_{\text{LMEM-C}}$ and $R^2_{\text{LMEM-M}}$ values were calculated as described previously²⁹.

All model diagnostics were done using the DHARMA package (0.2.0) which uses a simulation-based approach to create standardized residuals⁴⁶. For each model, scaled residuals were generated using the simulateResiduals function. The full model was used as the input for fittedModel parameter and 1000 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 1188 tumours³³. 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 was incorporated as a random effect in each model while tumour purity, age and sex were incorporated as fixed effects. 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 modelling since fewer than 20 tests were conducted. All models were adjusted for tumour purity based on previously published purity data³³. The full model for evaluating PGA is shown below as an example as follows:

$$\text{full}_{\text{PGA}} = \text{hypoxia} \sim \text{PGA} + \text{purity} + \text{age} + \text{sex} + (1|\text{cancer}) \quad (3)$$

To assess if SNVs per megabase were independently associated with hypoxia after adjusting for the total number of deletions we created two linear mixed-effect models. The full model associated hypoxia with SNVs per megabase while adjusting for cancer type, age, sex, tumour purity and the number of deletions (Eq. (4)). For comparison, a null model was created without our feature of interest, SNVs per megabase (Eq. (5)). The two models were compared using an ANOVA.

$$\text{full} = \text{hypoxia} \sim \text{SNVs per megabase} + \text{purity} + \text{age} + \text{sex} + \text{total deletions} + (1|\text{cancer}) \quad (4)$$

$$\text{null} = \text{hypoxia} \sim \text{purity} + \text{age} + \text{sex} + \text{total deletions} + (1|\text{cancer}) \quad (5)$$

Driver mutations analysis. Data for driver mutations was first summarized at the gene level for 1096 tumours with previously published driver mutation data²⁵. 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 was incorporated as a random effect in each model while tumour purity, age and sex were incorporated as fixed effects. 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 modelling. The full model for evaluating PTEN is shown below as an example as follows:

$$\text{full}_{\text{PTEN}} = \text{hypoxia} \sim \text{PTEN} + \text{purity} + \text{age} + \text{sex} + (1|\text{cancer}) \quad (6)$$

Mutational signature analysis. Previously published data for mutations attributed to various specific signatures was downloaded for 1188 tumours³⁵. 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 was incorporated as a random effect in each model while purity, age and sex were incorporated as fixed effects. 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 modelling. The full model for SBS1 is shown below as an example as follows:

$$\text{full}_{\text{SBS1}} = \text{hypoxia} \sim \text{SBS1} + \text{purity} + \text{age} + \text{sex} + (1|\text{cancer}) \quad (7)$$

Subclonality analysis. Previously reported³⁸ subclonal reconstruction data was used to summarize the number of clonal and subclonal mutations in all 1188 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 was incorporated as a random effect in each model while purity, age and sex were incorporated as fixed effects. 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 modelling since fewer than 20 tests were conducted.

The number of subclones was calculated for all 1188 tumours based on the number of clusters of cells identified in each sample. A linear mixed-effects model was used to associate the number of subclones with hypoxia score and this model was compared to a null model. Cancer type was incorporated as a random effect while purity, age and sex were incorporated as fixed effects. Tumours belonging to cancer types with fewer than 15 samples were excluded from the analysis. The full model for associating the number of subclones with hypoxia score is shown below as follows:

$$\text{full}_{\text{subclones}} = \text{hypoxia} \sim \text{subclones} + \text{purity} + \text{age} + \text{sex} + (1|\text{cancer}) \quad (8)$$

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 polyclonal³⁷. 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. An interaction model was first created where the relationship between the hypoxia scores and PTEN mutational status was modelled as an interaction (Eq. (9)). An additive model was also created where the relationship between hypoxia scores and PTEN mutational status was modelled in an additive manner (Eq. (10)):

$$\text{interaction} = \text{clonality} \sim \text{hypoxia} * \text{PTEN} + \text{purity} + \text{age} + \text{sex} + (1|\text{cancer}) \quad (9)$$

$$\text{additive} = \text{clonality} \sim \text{hypoxia} + \text{PTEN} + \text{purity} + \text{age} + \text{sex} + (1|\text{cancer}) \quad (10)$$

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 package⁴⁷ (v5.9.1). Figures were compiled using Inkscape (v0.91).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Aligned sequencing data, somatic and germline variant calls, single nucleotide variants, indels, copy-number alterations, structural variants, mutational signatures, subclonal reconstructions, transcript abundances and other core data generated by the PCAWG Consortium are described here²⁵ and available for download at dcc.icgc.org/releases/PCAWG. Additional information on accessing the data, including raw read files, can be found at docs.icgc.org/pcawg/data. In accordance with the data access policies of the ICGC and TCGA projects, most molecular, clinical and specimen data are in an open tier which does not require approval for access.

To access potentially identifiable information, such as germline alleles and underlying sequencing data, researchers will need to apply to the TCGA Data Access Committee (DAC) via dbGaP (dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login) for access to the TCGA portion of the dataset, and to the ICGC Data Access Compliance Office (DACO; icgc.org/daco) for the ICGC portion of the dataset. To access somatic single nucleotide variant data derived from TCGA donors, researchers will also need to obtain dbGaP authorization.

The analyses in this paper used several PCAWG datasets that were derived from the raw sequencing data and variant calls. The individual datasets are available at Synapse (synapse.org) and are denoted with accession numbers starting with *syn*. All of these datasets are also mirrored at dcc.icgc.org. The datasets encompass: clinical data for each patient, including information about age and sex (syn10389158, dcc.icgc.org/releases/PCAWG/clinical_and_histology); histopathology data for each tumour (syn10389164,

dcc.icgc.org/releases/PCAWG/clinical_and_histology); mRNA abundance data for each tumour (syn5553991, dcc.icgc.org/releases/PCAWG/transcriptome/gene_expression); driver mutations for each tumour spanning all classes of variants (syn11639581, dcc.icgc.org/releases/PCAWG/driver_mutations); single base substitution and small indel signature exposures for each tumour (syn8366024, dcc.icgc.org/releases/PCAWG/mutational_signatures); subclonal architecture, tumour evolution and tumour purity data for each tumour (syn8272483, dcc.icgc.org/releases/PCAWG/subclonal_reconstruction).

Code availability

The core computational pipelines used by the PCAWG Consortium for alignment, quality control and variant calling are available to the public at dockstore.org/search?search=pcawg under the GNU General Public License v3.0, which allows for reuse and distribution. Code specific to this manuscript is available upon request.

Received: 2 May 2019; Accepted: 29 November 2019;

Published online: 05 February 2020

References

- Weber, C. E. & Kuo, P. C. The tumor microenvironment. *Surg. Oncol.* **21**, 172–177 (2012).
- Blagosklonny, M. V. Antiangiogenic therapy and tumor progression. *Cancer Cell* **5**, 13–17 (2004).
- Vaupel, P., Thews, O. & Hoeckel, M. Treatment resistance of solid tumors. *Med. Oncol.* **18**, 243–260 (2001).
- Dhani, N., Fyles, A., Hedley, D. & Milosevic, M. The clinical significance of hypoxia in human cancers. *Semin. Nucl. Med.* **45**, 110–121 (2015).
- Brown, J. M. & Wilson, W. R. Exploiting tumour hypoxia in cancer treatment. *Nat. Rev. Cancer* **4**, 437–447 (2004).
- Zannella, V. E. et al. Reprogramming metabolism with metformin improves tumor oxygenation and radiotherapy response. *Clin. Cancer Res.* **19**, 6741–6750 (2013).
- Mucaj, V., Shay, J. E. S. & Simon, M. C. Effects of hypoxia and HIFs on cancer metabolism. *Int. J. Hematol.* **95**, 464–470 (2012).
- 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).
- 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).
- 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).
- Milosevic, M. et al. Tumor hypoxia predicts biochemical failure following radiotherapy for clinically localized prostate cancer. *Clin. Cancer Res.* **18**, 2108–2114 (2012).
- 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).
- Yang, L. et al. Development and validation of a 28-gene hypoxia-related prognostic signature for localized prostate cancer. *EBioMedicine* **31**, 182–189 (2018).
- 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**, 1521–1532 (2014).
- 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).
- Bhandari, V. et al. Molecular landmarks of tumor hypoxia across cancer types. *Nat. Genet.* **51**, 308–318 (2019).
- Graeber, T. G. et al. Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours. *Nature* **379**, 88–91 (1996).
- Greijer, A. E. & van der Wall, E. The role of hypoxia inducible factor 1 (HIF-1) in hypoxia induced apoptosis. *J. Clin. Pathol.* **57**, 1009–1014 (2004).
- Bindra, R. S. et al. Down-regulation of Rad51 and decreased homologous recombination in hypoxic cancer cells. *Mol. Cell. Biol.* **24**, 8504–8518 (2004).
- Bristow, R. G. & Hill, R. P. Hypoxia and metabolism: hypoxia, DNA repair and genetic instability. *Nat. Rev. Cancer* **8**, 180–192 (2008).
- 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).
- Mihaylova, V. T. et al. Decreased expression of the DNA mismatch repair gene Mlh1 under hypoxic stress in mammalian cells. *Mol. Cell. Biol.* **23**, 3265–3273 (2003).
- Koshiji, M. et al. HIF-1 α induces genetic instability by transcriptionally downregulating MutSa expression. *Mol. Cell* **17**, 793–803 (2005).
- Chua, M. L. K. et al. A prostate cancer “Nimbus”: genomic instability and SchLAP1 dysregulation underpin aggression of intraductal and cribriform subpathologies. *Eur. Urol.* **71**, 183–192 (2017).
- The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* <https://doi.org/10.1038/s41586-020-1969-6> (2020).
- PCAWG Transcriptome Core Group. et al. Genomic basis for RNA alterations in cancer. *Nature* <https://doi.org/10.1038/s41586-020-1970-0> (2020).
- 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).
- Li, C. H. et al. Sex differences in oncogenic mutational processes. *Nat. Commun.* In Press (2020).
- 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).
- 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**, 428–435 (2010).
- Winter, S. C. et al. Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res.* **67**, 3441–3449 (2007).
- Ragnum, H. B. et al. The tumour hypoxia marker pimonidazole reflects a transcriptional programme associated with aggressive prostate cancer. *Br. J. Cancer* **112**, 382–390 (2015).
- Yousif, F. et al. The origins and consequences of localized and global somatic hypermutation. Preprint at <https://www.biorxiv.org/content/10.1101/287839v2> (2018).
- Yuan, Y. et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat. Genet.* <https://doi.org/10.1038/s41588-019-0557-x> (2020).
- Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* <https://doi.org/10.1038/s41586-020-1943-3> (2020).
- Deshwar, A. G. et al. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* **16**, 35 (2015).
- Espirito, S. M. G. et al. The evolutionary landscape of localized prostate cancers drives clinical aggression. *Cell* **173**, 1003–1013 (2018).
- Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* <https://doi.org/10.1038/s41586-019-1907-7> (2020).
- Chan, N. et al. Contextual synthetic lethality of cancer cell kill based on the tumor microenvironment. *Cancer Res.* **70**, 8045–8054 (2010).
- Haider, S. et al. Genomic alterations underlie a pan-cancer metabolic shift associated with tumour hypoxia. *Genome Biol.* **17**, 140 (2016).
- 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).
- Luoto, K. R., Kumareswaran, R. & Bristow, R. G. Tumor hypoxia as a driving force in genetic instability. *Genome Integr.* **4**, 5 (2013).
- 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).
- Ye, Y. et al. Characterization of hypoxia-associated molecular features to aid hypoxia-targeted therapy. *Nat. Metab.* **1**, 431–444 (2019).
- Rosenberger, C., Rosen, S., Pallege, A. & Heyman, S. N. Pimonidazole adduct immunohistochemistry in the rat kidney: detection of tissue hypoxia. *Methods Mol. Biol.* **466**, 161–174 (2009).
- Dunn, P. K. & Smyth, G. K. Randomized quantile residuals. *J. Comput. Graph. Stat.* **5**, 236 (1996).
- P'ng, C. et al. BPG: seamless, automated and interactive visualization of scientific data. *BMC Bioinforma.* **20**, 42 (2019).

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. We acknowledge the contributions of the many clinical networks across ICGC and TCGA who provided samples and data to the PCAWG Consortium, and the contributions of the Technical Working Group and the Germline Working Group of the PCAWG Consortium for collation, realignment and harmonised variant calling of the cancer genomes used in this study. We thank the patients and their families for their participation in the individual ICGC and TCGA projects. 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), Canadian Cancer Society (grant #705649) and by the University of California. 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 Programme Academic Enrichment Fund for support (to Robert G. Bristow). This work was supported by a Terry Fox Research Institute Programme Project Grant. Robert G. Bristow is a recipient of a Canadian Cancer Society Research Scientist Award. Laboratory work for R.G.B is supported by the CRUK Manchester Institute through Cancer Research UK. This work was supported by the NIH/NCI through award number P30CA016042.

Author contributions

Bioinformatics and statistical analysis: V.B. Data processing: V.B., C.H.L. Data visualization: V.B. Supervised research: R.G.B., P.C.B. Initiated the project: V.B., P.C.B. Wrote the first draft of the manuscript: V.B. Approved the manuscript: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-14052-x>.

Correspondence and requests for materials should be addressed to R.G.B. or P.C.B.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020, corrected publication 2022