

# Molecular landmarks of tumor hypoxia across cancer types

Vinayak Bhandari <sup>1,2</sup>, Christianne Hoey <sup>1,3</sup>, Lydia Y. Liu<sup>1,2</sup>, Emilie Lalonde<sup>1,2</sup>, Jessica Ray <sup>1,3</sup>, Julie Livingstone <sup>2</sup>, Robert Lesurf<sup>2</sup>, Yu-Jia Shiah <sup>2</sup>, Tina Vujcic<sup>3</sup>, Xiaoyong Huang<sup>3</sup>, Shadrielle M. G. Espiritu<sup>2</sup>, Lawrence E. Heisler<sup>2</sup>, Fouad Yousif<sup>2</sup>, Vincent Huang <sup>2</sup>, Takafumi N. Yamaguchi <sup>2</sup>, Cindy Q. Yao<sup>2</sup>, Veronica Y. Sabelnykova<sup>2</sup>, Michael Fraser<sup>2</sup>, Melvin L. K. Chua<sup>4,5</sup>, Theodorus van der Kwast<sup>6</sup>, Stanley K. Liu <sup>1,3,7</sup>, Paul C. Boutros <sup>1,2,8,9,10,11,12\*</sup> and Robert G. Bristow <sup>1,7,13,14,15,16\*</sup>

**Many primary-tumor subregions have low levels of molecular oxygen, termed hypoxia. Hypoxic tumors are at elevated risk for local failure and distant metastasis, but the molecular hallmarks of tumor hypoxia remain poorly defined. To fill this gap, we quantified hypoxia in 8,006 tumors across 19 tumor types. In ten tumor types, hypoxia was associated with elevated genomic instability. In all 19 tumor types, hypoxic tumors exhibited characteristic driver-mutation signatures. We observed widespread hypoxia-associated dysregulation of microRNAs (miRNAs) across cancers and functionally validated miR-133a-3p as a hypoxia-modulated miRNA. In localized prostate cancer, hypoxia was associated with elevated rates of chromothripsis, allelic loss of *PTEN* and shorter telomeres. These associations are particularly enriched in polyclonal tumors, representing a constellation of features resembling tumor nimbus, an aggressive cellular phenotype. Overall, this work establishes that tumor hypoxia may drive aggressive molecular features across cancers and shape the clinical trajectory of individual tumors.**

The tumor microenvironment is characterized by dynamic gradients of oxygen diffusion and consumption<sup>1–3</sup> leading to subregions of hypoxia in approximately half of all solid tumors<sup>4–7</sup>. These regions vary in their size and extent, and arise because of a decreased oxygen supply (from disorganized and irregular tumor vasculature<sup>8</sup>) or increased oxygen demand (from changes in tumor metabolism<sup>9,10</sup>). Tumor adaptation to this imbalance between oxygen supply and demand is associated with poor clinical prognosis and elevated genomic instability<sup>3,11</sup>, resistance to chemotherapy and radiotherapy<sup>12,13</sup>, immune dampening<sup>14</sup>, development of tumor stem cell-protective niches<sup>15,16</sup> and increased proclivity for distant metastasis<sup>17–20</sup>.

Cycling hypoxia can strongly select for apoptosis-deficient cells and for cells with *TP53* mutations<sup>21,22</sup>. Hypoxic cells have reduced abundance of RAD51, thus resulting in inhibited DNA repair that can persist up to 48h after reoxygenation<sup>3,23</sup>. Furthermore, the DNA-mismatch-repair genes *MLH1* and *MSH2* are downregulated under hypoxia, thus leading to increased mutagenesis<sup>24,25</sup>. Indeed, we previously reported that the copresence of tumor hypoxia (on the basis of mRNA signatures or needle electrode measurements) and genomic instability synergistically portends rapid relapse after primary treatment for prostate cancer<sup>26,27</sup>, supporting the concept that a hostile tumor microenvironment may select for or drive adaptation of a distinctive genomic profile.

However, although hypoxia is an adverse (and targetable) prognostic feature in multiple cancer types<sup>1,28,29</sup>, its genomic correlates in large cohorts of diverse primary tumors had not been described. To fill this gap, we evaluated tumor hypoxia in 8,006 tumors representing 19 distinct tumor types to create a pancancer quantification of this microenvironmental cancer hallmark. We identified somatic-mutational hallmarks of hypoxia across multiple tumor types, in addition to associations with ancestry. With a focus on localized prostate cancer, for which whole-genome-sequencing data linked to direct intratumoral oxygen measurements were available, we thoroughly characterized associations with hypoxia at the genomic, transcriptomic and subclonal levels. These data strongly suggest that hypoxia shapes tumor evolutionary landscapes, applying a strong selective pressure for specific molecular aberrations. Furthermore, they suggest a framework to guide the development of biomarkers that may inform treatment planning and potential avenues for new therapies to combat this hallmark of aggressive cancers.

## Results

**The pancancer landscape of tumor hypoxia.** We first quantified tumor hypoxia in 8,006 tumors from 19 tumor types in The Cancer Genome Atlas (TCGA) and the Canadian Prostate Cancer Genome Network (CPC-GENE) by using mRNA-based hypoxia

<sup>1</sup>Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. <sup>2</sup>Informatics and Biocomputing Program, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. <sup>3</sup>Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada. <sup>4</sup>Division of Radiation Oncology, National Cancer Centre Singapore, Singapore, Singapore. <sup>5</sup>Duke-NUS Graduate Medical School, Singapore, Singapore. <sup>6</sup>Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada. <sup>7</sup>Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada. <sup>8</sup>Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada. <sup>9</sup>Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA. <sup>10</sup>Department of Urology, University of California, Los Angeles, Los Angeles, CA, USA. <sup>11</sup>Jonsson Comprehensive Cancer Centre, University of California, Los Angeles, Los Angeles, CA, USA. <sup>12</sup>Institute for Precision Health, University of California, Los Angeles, Los Angeles, CA, USA. <sup>13</sup>Princess Margaret Hospital, University Health Network, Toronto, Ontario, Canada. <sup>14</sup>Division of Cancer Sciences, Faculty of Biology, Health and Medicine, University of Manchester, Manchester, UK. <sup>15</sup>The Christie NHS Foundation Trust, Manchester, UK. <sup>16</sup>CRUK Manchester Institute and Manchester Cancer Research Centre, Manchester, UK. \*e-mail: [pboutros@mednet.ucla.edu](mailto:pboutros@mednet.ucla.edu); [robert.bristow@manchester.ac.uk](mailto:robert.bristow@manchester.ac.uk)

signatures from Buffa<sup>30</sup>, Winter<sup>31</sup>, Ragnum<sup>32</sup>, West<sup>33</sup>, Sorensen<sup>34</sup>, Elvidge<sup>35</sup>, Hu<sup>36</sup> and Seigneuric<sup>37</sup> (Fig. 1a, Supplementary Fig. 1a,b and Supplementary Tables 1 and 2). Pancancer hypoxia scores from the eight independent signatures were strongly correlated ( $\rho=0.42\pm 0.21$ , mean  $\pm$  s.d., algorithm AS89; Supplementary Fig. 1c) and were highly consistent in their predictions of hypoxia for primary tumors (Supplementary Fig. 1d,e).

Overall, squamous cell tumors of the head and neck (HNSC), cervix (CESC) and lung (LUSC) were the most hypoxic, whereas adenocarcinomas of the thyroid (THCA) and prostate (PRAD) were the least hypoxic (Fig. 1a). These results reflect and extend previous descriptions of hypoxic microenvironments across tumor types, and in some cases reflect known benefits from hypoxia-targeting interventions<sup>29,38</sup>. However, these bulk results hide a notable degree of intertumoral heterogeneity in the levels of hypoxia within each tumor type; 42% of the variance in hypoxia scores lies within individual tumor types rather than between them. Intertumoral variability in hypoxia was particularly pronounced in adenocarcinomas of the lung (LUAD; interquartile range (IQR)=38) and pancreas (PAAD; IQR=32) and in breast tumors (BRCA; IQR=32). These contrasted with other tumor types in which hypoxia was highly consistent among tumors, such as thyroid cancers (THCA; IQR=8). Variability in hypoxia within a tumor type was not associated with the median level of hypoxia within that tumor type ( $P=0.29$ ,  $\rho=0.26$ , AS89).

Hypoxia scores generated by using protein abundance data were significantly correlated with mRNA-abundance-based hypoxia scores for all three tumor types when data were available (BRCA:  $P=2.40\times 10^{-8}$ ,  $\rho=0.58$ ; OV, ovarian serous cystadenocarcinoma:  $P=1.41\times 10^{-5}$ ,  $\rho=0.42$ ; COADREAD, colon and rectum adenocarcinoma:  $P=3.02\times 10^{-3}$ ,  $\rho=0.32$ , AS89; Supplementary Fig. 2a–c). Across these tumor types, the abundance of several proteins was significantly correlated with protein-based hypoxia scores (false discovery rate (FDR) <0.05 in all cancer types; Supplementary Fig. 2d), including LDHA, GPI, PFKF and ERO1L, all of which have key roles in hypoxia-associated metabolic changes. These data reinforce the idea that these hypoxia signatures are informative about tumor oxygen levels in primary tumors.

Although somatic-mutation profiles differ with sex and age at diagnosis<sup>39–41</sup>, tumor hypoxia did not vary with either of these features (Supplementary Fig. 3a–d). In most tumor types, we did not have sufficient statistical power to consider the relationship of hypoxia with ancestry; however, among breast tumors, for which the sample size was twice that of any other cancer type, we observed a strong association with subject-reported ancestry: tumors arising in subjects of Caucasian ancestry had less hypoxia than tumors in subjects with either Asian or African ancestry (Bonferroni-adjusted  $P=4.08\times 10^{-13}$ , Kruskal–Wallis test; Fig. 1b and Supplementary Fig. 3e–g). This may explain the observations that evofosfamide, a hypoxia-targeting agent, showed higher efficacy in subjects of Asian descent in the phase 3 MAESTRO trial.

**The genomic hallmarks of tumor hypoxia.** We next sought to identify somatic genomic changes that characterize hypoxic tumors. We first considered the percentage of the genome altered by copy-number aberrations (PGA), a surrogate of genomic instability that is correlated with aggressivity in several tumor types<sup>42,43</sup>. Tumor hypoxia was associated with significantly elevated genomic instability in 10 of 19 tumor types, and in no case was tumor hypoxia associated with decreased genomic instability (Bonferroni-adjusted  $P<0.05$ , AS89; Fig. 1c). These ten tumor types included prostate cancer, in which PGA and hypoxia synergistically predict relapse after definitive local therapy<sup>26</sup>.

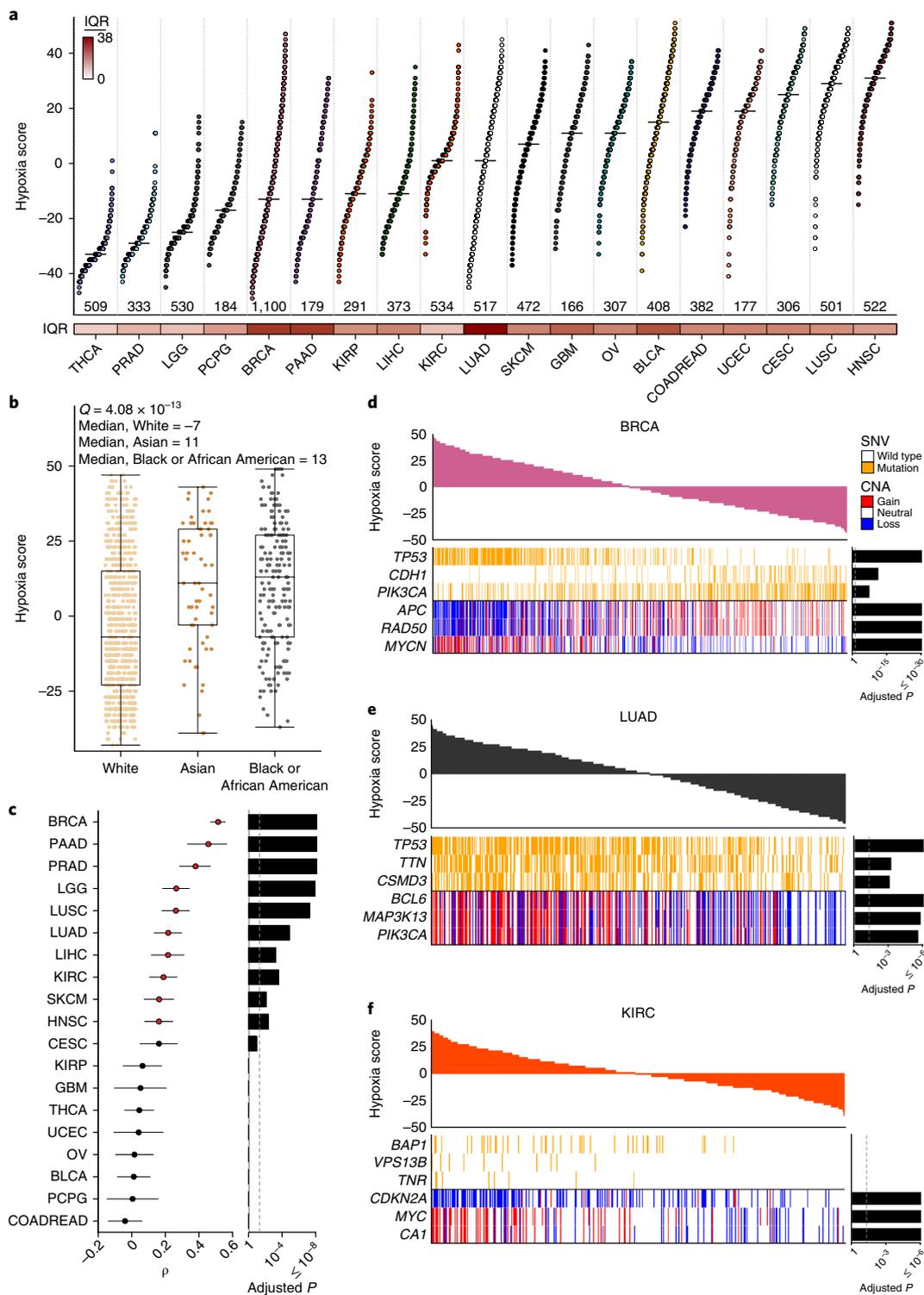
To specifically assess whether hypoxia was associated simply with an elevated mutation rate or with specific gene-level mutational features, we focused on three tumor types with large sample

sizes that showed a significant correlation between hypoxia and PGA: breast cancers, lung adenocarcinomas and renal clear cell carcinomas (Supplementary Fig. 3h). In each of these tumor types, we identified specific mutational events associated with tumor hypoxia, including both somatic copy-number aberrations (CNAs) and somatic single-nucleotide variants (SNVs). In breast cancers (Fig. 1d and Supplementary Table 3), hypoxic tumors were more likely to harbor loss of *APC* (Bonferroni-adjusted  $P=1.25\times 10^{-42}$ , Mann–Whitney *U* test) and gain of *MYCN* (Bonferroni-adjusted  $P=2.75\times 10^{-32}$ , Mann–Whitney *U* test). Hypoxic breast tumors also showed an elevated rate of *TP53* point mutations (Bonferroni-adjusted  $P=4.38\times 10^{-61}$ , Mann–Whitney *U* test), in agreement with the selection of *TP53*-mutant cells under hypoxic stress<sup>21</sup>. These CNA and SNV profiles of hypoxia were validated in an independent 1,859-subject cohort<sup>44,45</sup> (Supplementary Fig. 4a). The somatic events recurrently associated with hypoxia reflect in part differences in tumor hypoxia among breast cancer subtypes ( $P<2.20\times 10^{-50}$ , Kruskal–Wallis test; Supplementary Fig. 4b,c), with basal-like tumors being the most hypoxic, as we expected<sup>46</sup>. This finding was confirmed by using proteinabundance-based hypoxia scores ( $P=1.50\times 10^{-3}$ , Kruskal–Wallis test; Supplementary Fig. 4d). However, these hypoxia-mutation associations bridged across subtypes: tumors with SNVs in *TP53* had higher hypoxia scores than tumors with wild-type *TP53* in multiple subtypes and in both cohorts (Supplementary Fig. 4f–o). Furthermore, *TP53* protein abundance was positively correlated with protein-based hypoxia scores, as we expected ( $P=1.96\times 10^{-2}$ ,  $\rho=0.28$ , AS89; Supplementary Fig. 4e).

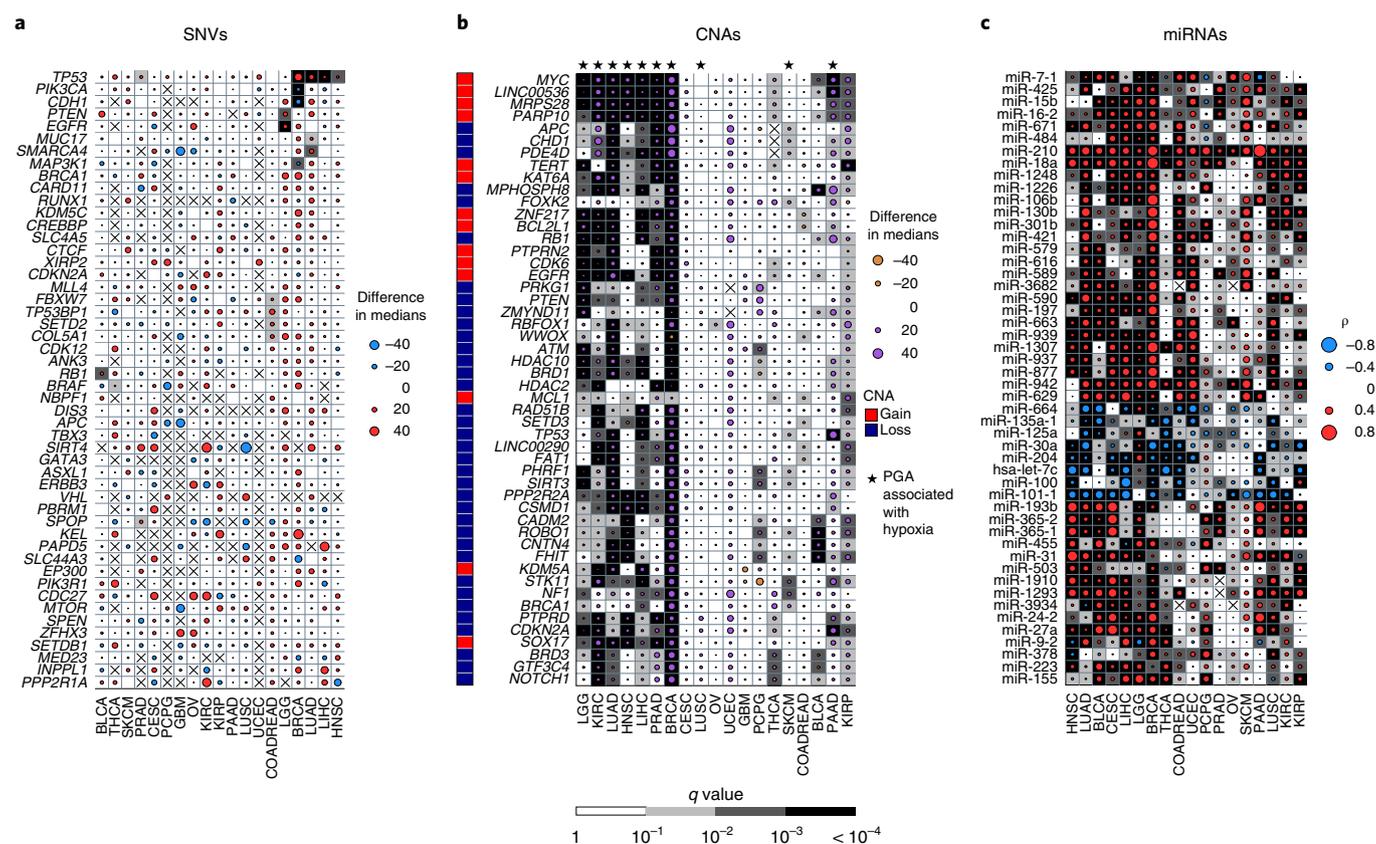
This strong association between *TP53* SNVs and hypoxia was also observed in adenocarcinomas of the lung (Fig. 1e and Supplementary Table 3). Hypoxic lung adenocarcinomas were additionally associated with gain of *BCL6* (Bonferroni-adjusted  $P=3.24\times 10^{-7}$ , Mann–Whitney *U* test) and *PIK3CA* (Bonferroni-adjusted  $P=2.93\times 10^{-6}$ , Mann–Whitney *U* test). Similarly, in renal clear cell carcinoma tumors (Fig. 1f and Supplementary Table 3), hypoxia was associated with a loss of *CDKN2A* (Bonferroni-adjusted  $P=1.40\times 10^{-9}$ , Mann–Whitney *U* test) and gain of *MYC* (Bonferroni-adjusted  $P=3.71\times 10^{-8}$ , Mann–Whitney *U* test). Alterations in several other genes were also associated with hypoxia, many of which have central roles in key signaling and DNA-repair pathways, thereby creating testable hypotheses for future in vitro modeling studies (Fig. 1d–f and Supplementary Table 3).

Next we focused on 253 cancer driver genes recurrently altered by somatic SNVs<sup>47</sup> and identified 15 as being associated with hypoxia (FDR <0.10; Fig. 2a, ordered on the basis of consensus clustering; Methods). SNVs in *TP53* were associated with hypoxia in five tumor types: breast carcinomas (FDR =  $6.95\times 10^{-63}$ ; Mann–Whitney *U* test), lung adenocarcinomas (FDR =  $1.83\times 10^{-12}$ ; Mann–Whitney *U* test), liver cancers (FDR =  $1.64\times 10^{-6}$ ; Mann–Whitney *U* test), head and neck squamous cell carcinomas (FDR =  $2.26\times 10^{-3}$ ; Mann–Whitney *U* test) and localized prostate adenocarcinomas (FDR =  $8.58\times 10^{-2}$ ; Mann–Whitney *U* test). Notably, most hypoxia-SNV associations were detected in only one tumor type, at least given the statistical power of the existing cohorts, thus highlighting strong intertumor type diversity in the SNV mutational signature of hypoxia.

In contrast, analysis of 112 cancer driver genes altered by CNAs<sup>48</sup> identified multiple oncogenes and tumor suppressors recurrently associated with hypoxia in multiple cancer types (Fig. 2b, ordered on the basis of consensus clustering; Methods). For example, gain of the *MYC* oncogene was associated with elevated hypoxia in 11 separate tumor types, whereas loss of the tumor-suppressor gene *PTEN* was associated with elevated hypoxia in seven tumor types. Loss of *TP53* was associated with elevated hypoxia in many of the same tumor types in which SNVs in *TP53* were associated with hypoxia, such as breast carcinoma (FDR =  $1.35\times 10^{-22}$ ; Mann–Whitney *U* test),



**Fig. 1 | The hallmarks of hypoxia in cancer. a**, Hypoxia scores based on the Buffa mRNA abundance signature for 19 tumor types, sorted by the median hypoxia score (horizontal black line) for each tumor type. Sample sizes are shown along bottom. Intertumoral variability in hypoxia measured by IQR was particularly high in adenocarcinomas of the lung (IQR=38.0; LUAD), pancreas (IQR=32.0; PAAD) and breast tumors (IQR=32.0; BRCA). **b**, Differences in tumor hypoxia on the basis of ancestry for subjects with breast invasive carcinoma ( $n=997$  independent tumors, Kruskal–Wallis test). A Tukey box plot is shown. Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within  $1.5 \times$  the IQR. **c**, Ten of 19 tumor types showed a significant correlation between PGA and hypoxia scores ( $n=1,082, 178, 333, 527, 498, 514, 366, 526, 470, 516, 294, 288, 160, 504, 177, 304, 404, 166$  and  $379$  independent tumors from top to bottom). Center values represent Spearman's  $\rho$ , and error bars represent the 95% CI for the correlation.  $P$  values were calculated by using the algorithm AS89 and adjusted for multiple testing (Bonferroni), shown on the right. **d–f**, Notable associations of SNVs and CNAs with tumor hypoxia in subjects with breast invasive carcinoma (BRCA,  $n=960$  independent tumors; **d**), lung adenocarcinoma (LUAD,  $n=475$  independent tumors; **e**) and renal clear cell carcinoma (KIRC,  $n=431$  independent tumors; **f**). Bonferroni-adjusted  $P$  values are on the right (Mann–Whitney  $U$  test). Tumor type codes are defined in Supplementary Table 2, and associations for all genes are summarized in Supplementary Table 3. All tests were two sided.



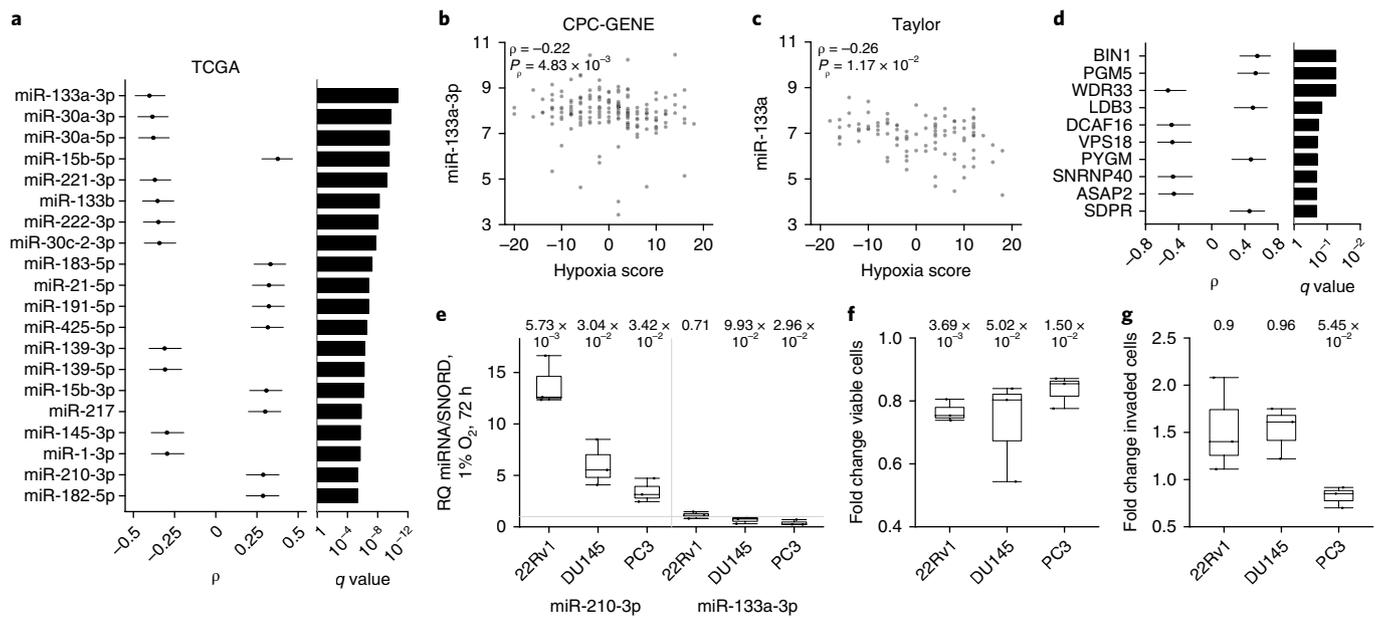
**Fig. 2 | Hypoxia associations with driver aberrations and miRNAs.** **a**, Association of hypoxia with SNVs in driver genes (Mann–Whitney *U* test). Tumors with mutations in *TP53* have higher hypoxia scores than tumors without mutations in *TP53* across multiple tumor types. Dot size indicates the difference in median hypoxia score between tumors with an SNV and those without an SNV. **b**, Association of hypoxia with CNAs in oncogenes and tumor suppressor genes (Mann–Whitney *U* test). Copy-number gain of the oncogene *MYC* and loss of the tumor-suppressor gene *PTEN* were associated with elevated hypoxia in 11 and 7 tumor types, respectively. Dot size indicates the difference in median hypoxia score between tumors with a CNA (gain for oncogene, loss for tumor suppressor gene) and those without a CNA. Stars identify tumor types in which PGA is significantly correlated with hypoxia. **c**, Hypoxia-associated miRNAs across tumor types (Spearman’s  $\rho$ , AS89). Increased hypoxia was associated with induction of miR-210 across every tumor type. Dot size indicates  $\rho$ . Background shading indicates the *q* value for **a–c**. **na** and **nc** independent tumors = 388 and 405, bladder urothelial carcinoma (BLCA); 960 and 753, BRCA; 190 and 304, CESC; 259 and 295, COADREAD; 137, glioblastoma multiforme (GBM); 497 and 478, HNSC; 431 and 254, KIRC; 280 and 290, kidney renal papillary cell carcinoma (KIRP); 513 and 512, lower-grade glioma (LGG); 360 and 367, liver hepatocellular carcinoma (LIHC); 475 and 447, LUAD; 178 and 342, LUSC; 246 and 288, OV; 119 and 178, PAAD; 162 and 179, pheochromocytoma and paraganglioma (PCPG); 333 and 330, PRAD; 290 and 97, skin cutaneous melanoma (SKCM); 486 and 500, THCA; 8 and 174, uterine corpus endometrial carcinoma (UCEC). Tumor type codes are defined in Supplementary Table 2.

lung adenocarcinomas (FDR = 3.60 × 10<sup>-5</sup>; Mann–Whitney *U* test) and liver cancers (FDR = 3.57 × 10<sup>-4</sup>; Mann–Whitney *U* test). Other tumor types in which loss of *TP53* was associated with hypoxia include renal clear cell carcinoma (FDR = 3.85 × 10<sup>-3</sup>; Mann–Whitney *U* test), renal papillary cell carcinoma (FDR = 6.76 × 10<sup>-2</sup>; Mann–Whitney *U* test) and pancreatic ductal adenocarcinoma (FDR = 3.92 × 10<sup>-6</sup>; Mann–Whitney *U* test). Thus, hypoxia seemed to be associated with a consistent signature of gene gains and losses, but individual tumor types had distinctive hypoxia–SNV relationships, and these relationships were independent of the hypoxia signature used (Supplementary Fig. 4p,q).

Tumor hypoxia was also associated with unique microRNA (miRNA, denoted with the miR prefix) abundance profiles (Fig. 2c, ordered on the basis of consensus clustering<sup>7</sup> Methods and Supplementary Fig. 4r). Fully 84% (658 of 784) of measured miRNAs showed correlations with tumor hypoxia (FDR < 0.01, AS89) in at least one tumor type, including some especially strong associations. For example, miR-210 abundance was associated with elevated hypoxia score across all 18 tumor types (Spearman’s  $\rho$  range = 0.20–0.66). This finding is consistent with previous

descriptions of nearly universal induction of miR-210 under hypoxia in six tumor types<sup>49</sup>. miR-210 abundance is positively correlated with the protein abundance of LDHA in breast cancer ( $\rho = 0.72$ , FDR = 5.66 × 10<sup>-2</sup>, AS89; Supplementary Fig. 4s) and ovarian cancer ( $\rho = 0.42$ , FDR = 6.21 × 10<sup>-4</sup>, AS89; Supplementary Fig. 4t). Most miRNAs had a consistent directionality in their association with hypoxia across tumor types, but a subset had divergent patterns, such as miR-9-2, whose abundance was associated with elevated hypoxia in seven tumor types and with diminished hypoxia in three (FDR < 0.05, AS89). These data identify a strong dysregulation of miRNA abundances associated with tumor hypoxia and demonstrate that tumor hypoxia is associated with highly specific genomic and transcriptomic signatures.

**Evaluation of hypoxia-associated miRNAs in prostate cancer.** On the basis of the extensive hypoxia-associated pancancer dysregulation of miRNAs that we observed, we sought to characterize these relationships in prostate cancer, for which several large miRNA data sets are available. In the TCGA data, miR-133a-3p was the strongest hypoxia-associated miRNA (FDR = 2.08 × 10<sup>-11</sup>,  $\rho = -0.40$ ,



**Fig. 3 | Assessment and validation of hypoxia-associated miRNAs in prostate cancer.** **a**, The abundance of several miRNAs is associated with hypoxia score in TCGA. Results are shown for the top 20 miRNAs, according to  $q$  values (Spearman's  $\rho$ , AS89). The strongest association was seen for miR-133a-3p ( $n = 330$  independent tumors). **b, c**, We confirmed that the abundance of miR-133a-3p was also significantly associated with hypoxia in the CPC-GENE cohort (Spearman's  $\rho$ , AS89;  $n = 170$  independent tumors; **b**) and a third independent cohort<sup>50</sup> (Spearman's  $\rho$ , AS89;  $n = 97$  independent tumors; **c**). **d**, The abundance of miR-133a-3p was associated with several proteins in CPC-GENE (Spearman's  $\rho$ , AS89;  $n = 57$  independent tumors). **e**, We validated the hypoxia-associated nature of miR-133a in vitro by examining three prostate cancer cell lines. The abundance of miR-210-3p, a known hypoxia-induced miRNA, was significantly elevated in all three cell lines after exposure to 1% oxygen for 72 h. Under the same conditions, the abundance of miR-133a was significantly decreased in DU145 cells and PC3 cells. SNORD, SNORD61; RQ, relative quantity. **f**, Introducing a miR-133a-3p mimic led to significantly diminished cell viability across all three cell lines. **g**, Introducing a miR-133a-3p mimic significantly decreased the invasive ability of PC3 cells. Center values in **a** and **d** represent Spearman's  $\rho$ , and error bars represent the 95% CI for the correlation. For **e–g**, numbers at top represent  $P$  values (one-sided paired Student's  $t$  test),  $n = 3$  biologically independent experiments, degrees of freedom = 2, Tukey box plots are shown. Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5x the IQR.

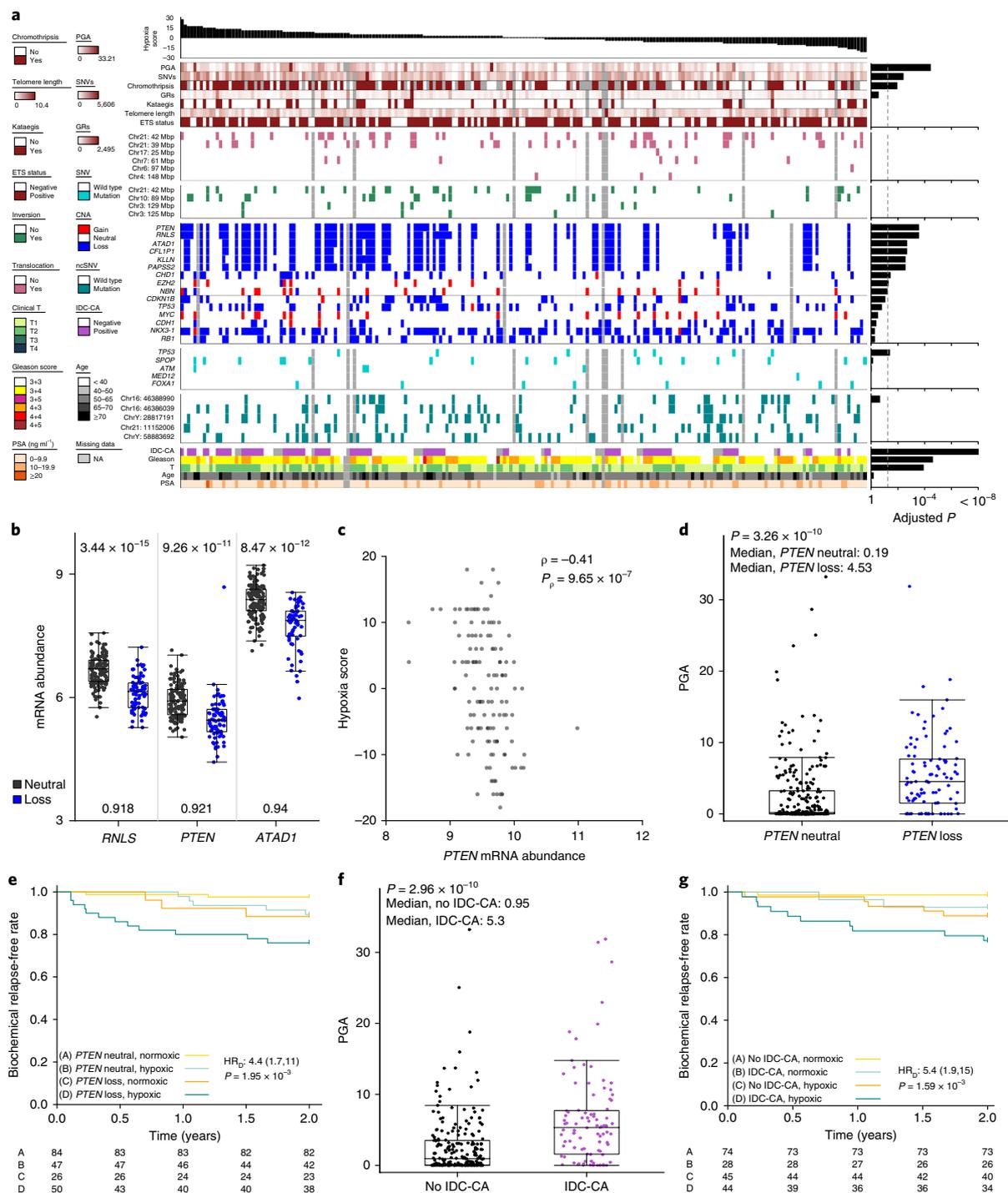
AS89; Fig. 3a). We confirmed that the abundance of miR-133a-3p was significantly associated with hypoxia in the CPC-GENE cohort (FDR =  $4.83 \times 10^{-3}$ ,  $\rho = -0.22$ , AS89; Fig. 3b) and a third independent cohort<sup>50</sup> with 97 subjects (FDR =  $1.17 \times 10^{-2}$ ,  $\rho = -0.26$ , AS89; Fig. 3c). The abundance of miR-133a-3p is associated with that of several proteins, including the tumor suppressor BIN1 (ref. <sup>51</sup>) (FDR =  $5.47 \times 10^{-2}$ ,  $\rho = 0.55$ , AS89; Fig. 3d) and the phosphotransferase PGM5 (ref. <sup>52</sup>) (FDR =  $5.47 \times 10^{-2}$ ,  $\rho = 0.53$ , AS89; Fig. 3d).

To validate the association of miR-133a-3p with hypoxia, we turned to controlled, in vitro models of prostate cancer. We first confirmed that the abundance of miR-210-3p, a known hypoxia-induced miRNA, was elevated after exposure to 1% O<sub>2</sub> for 72 h in 22Rv1 ( $P = 5.73 \times 10^{-3}$ ,  $t = -9.26$ , Student's  $t$  test), DU145 ( $P = 3.04 \times 10^{-2}$ ,  $t = -3.87$ , Student's  $t$  test) and PC3 cells ( $P = 3.42 \times 10^{-2}$ ,  $t = -3.62$ , Student's  $t$  test, Fig. 3e). Next, we confirmed that the abundance of miR-133a-3p was diminished in DU145 ( $P = 9.93 \times 10^{-2}$ ,  $t = 1.89$ , Student's  $t$ -test) and PC3 cells ( $P = 2.96 \times 10^{-2}$ ,  $t = 3.92$ , Student's  $t$  test, Fig. 3e) after exposure to 1% O<sub>2</sub> for 72 h, corroborating our prior observations from three independent primary-tumor cohorts. Indeed, introduction of a miR-133a-3p mimic significantly decreased cell viability in all the cell lines ( $P_{22Rv1} = 3.69 \times 10^{-3}$ ,  $t_{22Rv1} = 11.57$ ,  $P_{DU145} = 5.02 \times 10^{-2}$ ,  $t_{DU145} = 2.91$ ,  $P_{PC3} = 1.50 \times 10^{-2}$ ,  $t_{PC3} = 5.65$ , Student's  $t$  test; Fig. 3f) and decreased the invasive ability of PC3 cells ( $P = 5.45 \times 10^{-2}$ ,  $t = 2.78$ , Student's  $t$  test; Fig. 3g). Overall, these data demonstrate that miR-133a-3p is a hypoxia-associated miRNA with functional consequences in prostate cancer.

**The mutational landscape of hypoxia in prostate cancer.** We continued to focus on localized prostate cancer, given the strong

associations of hypoxia with clinical progression of prostate tumors<sup>26,53</sup> and the availability of large cohorts of matched whole-genome sequencing with subclonal architecture, RNA abundance and clinical-outcome data<sup>54,55</sup>. We compiled a cohort of 548 subjects with hormone-naïve, surgically managed localized prostate cancer: 149 samples with published whole-genome-sequencing data, 371 samples with published whole-exome-sequencing data and an additional 28 tumor-normal whole-genome pairs that were sequenced. Because the Ragnum hypoxia signature used in our pancancer analysis was developed from localized tumors analyzed with pimonidazole, a validated extrinsic marker of tumor hypoxia<sup>32</sup>, and the hypoxia signatures that we examined were highly correlated (Supplementary Fig. 1c–e), we used the Ragnum signature for these analyses. All samples in this cohort were surgical specimens subjected to centralized pathology review by either International Cancer Genome Consortium (ICGC) or TCGA with macrodissection of tumor regions, and were analyzed with a consistent bioinformatics pipeline (Supplementary Table 4).

Higher hypoxia scores were significantly associated with more advanced tumor extent (that is, T category; Bonferroni-adjusted  $P = 1.20 \times 10^{-4}$ , Kruskal–Wallis test; Fig. 4a and Supplementary Fig. 5a) and higher Gleason grade (Bonferroni-adjusted  $P = 2.44 \times 10^{-5}$ , Kruskal–Wallis test; Fig. 4a and Supplementary Fig. 5b). In contrast, neither pretreatment prostate-specific antigen (PSA) levels (Bonferroni-adjusted  $P = 0.89$ ,  $\rho = 0.07$ , AS89; Fig. 4a) nor age at diagnosis (Bonferroni-adjusted  $P = 0.60$ ,  $\rho = 0.07$ , AS89; Fig. 4a) was associated with hypoxia in this cohort. Many recurrent somatic-mutational features of prostate tumors were not associated with hypoxia (Supplementary Note).



**Fig. 4 | The landscape of hypoxia in prostate cancer.** **a**, We associated tumor hypoxia with an array of variables ( $n$  independent tumors = 360 for PGA, CNAs; 155 for translocations, inversions, noncoding (nc) SNVs, SNV count, genomic rearrangement (GR) count, kataegis and telomere length; 115 for chromothripsis; 354 for SNVs; 499 for IDC-CA; 479 for Gleason score and age; 405 for clinical T; 360 for PSA; Spearman's  $\rho$  and AS89 was used for all continuous features; Fisher's exact test for CNAs, translocations and inversions; Mann-Whitney  $U$  test for chromothripsis, kataegis, SNVs, ncSNVs and IDC-CA; Kruskal-Wallis test for Gleason and clinical T). Hypoxia scores are shown across the top,  $P$  values (adjusted for multiple testing) are at right, and missing data are in gray. **b**, mRNA abundance differences based on CNA status for hypoxia-associated genes.  $q$  values are shown across the top, and fold changes are shown along the bottom ( $n = 210$  independent tumors; Mann-Whitney  $U$  test). **c**, *PTEN* mRNA abundance is negatively associated with hypoxia scores in an independent cohort<sup>50</sup> ( $n = 130$  independent tumors; Spearman's  $\rho$ , AS89). **d**, Tumors with loss of *PTEN* had elevated genomic instability ( $n = 345$  independent tumors; Mann-Whitney  $U$  test). **e**, Subjects who were hypoxic and had a loss of *PTEN* had significantly poorer 2-year prognosis compared with all other subjects ( $n = 207$  independent tumors; Wald test). **f**, Tumors with IDC-CA had elevated genomic instability ( $n = 318$  independent tumors; Mann-Whitney  $U$ -test). **g**, Subjects who were hypoxic and displayed IDC-CA had significantly poorer 2-year prognosis compared with all other subjects ( $n = 191$  independent tumors; Wald test). HR<sub>D</sub> refers to the hazard ratio for subjects in group D compared with all other subjects. Tukey box plots are shown. Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5× the IQR. All tests were two sided.

Conversely, and as previously identified in TCGA data alone (Fig. 1c), elevated PGA was associated with higher levels of tumor hypoxia (Bonferroni-adjusted  $P=3.55 \times 10^{-5}$ ,  $\rho=0.24$ , AS89; Fig. 4a). Additionally, catastrophic chromothriptic events (Bonferroni-adjusted  $P=2.69 \times 10^{-2}$ , Mann–Whitney  $U$  test) and the total burden of somatic SNVs (Bonferroni-adjusted  $P=2.52 \times 10^{-2}$ ,  $\rho=0.26$ , AS89) were elevated in subjects with higher levels of tumor hypoxia. *SPOP*, the gene most frequently altered by somatic SNVs in localized prostate cancers<sup>54</sup>, was not associated with hypoxia, but as in other tumor types, SNVs in *TP53* were significantly associated with elevated hypoxia (Bonferroni-adjusted  $P=3.33 \times 10^{-2}$ , Mann–Whitney  $U$  test). Tumors with mitochondrial genome mutations previously associated with poor prognosis<sup>56</sup> also had elevated hypoxia ( $P=0.048$ , Kruskal–Wallis test; Supplementary Fig. 5c).

Because prostate cancer is a C-class tumor (that is, preferentially driven by CNAs rather than SNVs)<sup>54</sup>, we next performed a genome-wide analysis to identify CNAs associated with tumor hypoxia (Supplementary Table 5). We identified 1,189 genes with CNA–hypoxia associations (FDR < 0.10; Fisher's exact test), and these were preferentially located in chromosome 7 (Bonferroni-adjusted  $P=7.62 \times 10^{-224}$ , hypergeometric test) and chromosome 10 (Bonferroni-adjusted  $P=1.41 \times 10^{-32}$ , hypergeometric test; Supplementary Fig. 6a). To identify those genes most likely to be functionally associated with hypoxia, we considered their mRNA levels and identified 20 genes whose mRNA abundance was both correlated with hypoxia and changed by CNAs in both the TCGA (Supplementary Fig. 6b) and CPC-GENE cohorts (Supplementary Fig. 6c). This group included *RNLS*, *ATAD1*, *OTUD6B*, *OSGIN2*, *NBN*, *MRPL15*, *E2F5*, *MCM4*, *LACTB2*, *GGCT*, *MCM7*, *EZH2*, *LSM5*, *PRKDC*, *PTEN*, *CAST*, *EPB41L4A*, *TWISTNB*, *SNX10* and *ZNF768*. As a group, hypoxia-associated genes were associated with ATP-dependent DNA helicase activity, the nuclear chromosome (telomeric region) and chromosome (telomeric region) pathways (Supplementary Fig. 6d). CNAs in other driver genes commonly altered in prostate cancer (for example, *CDKN1B*, *TP53*, *RBI*, *MYC*, *NKX3-1* and *CDH1*) were not associated with higher hypoxia scores. There is thus a specific CNA signature associated with hypoxic prostate cancers.

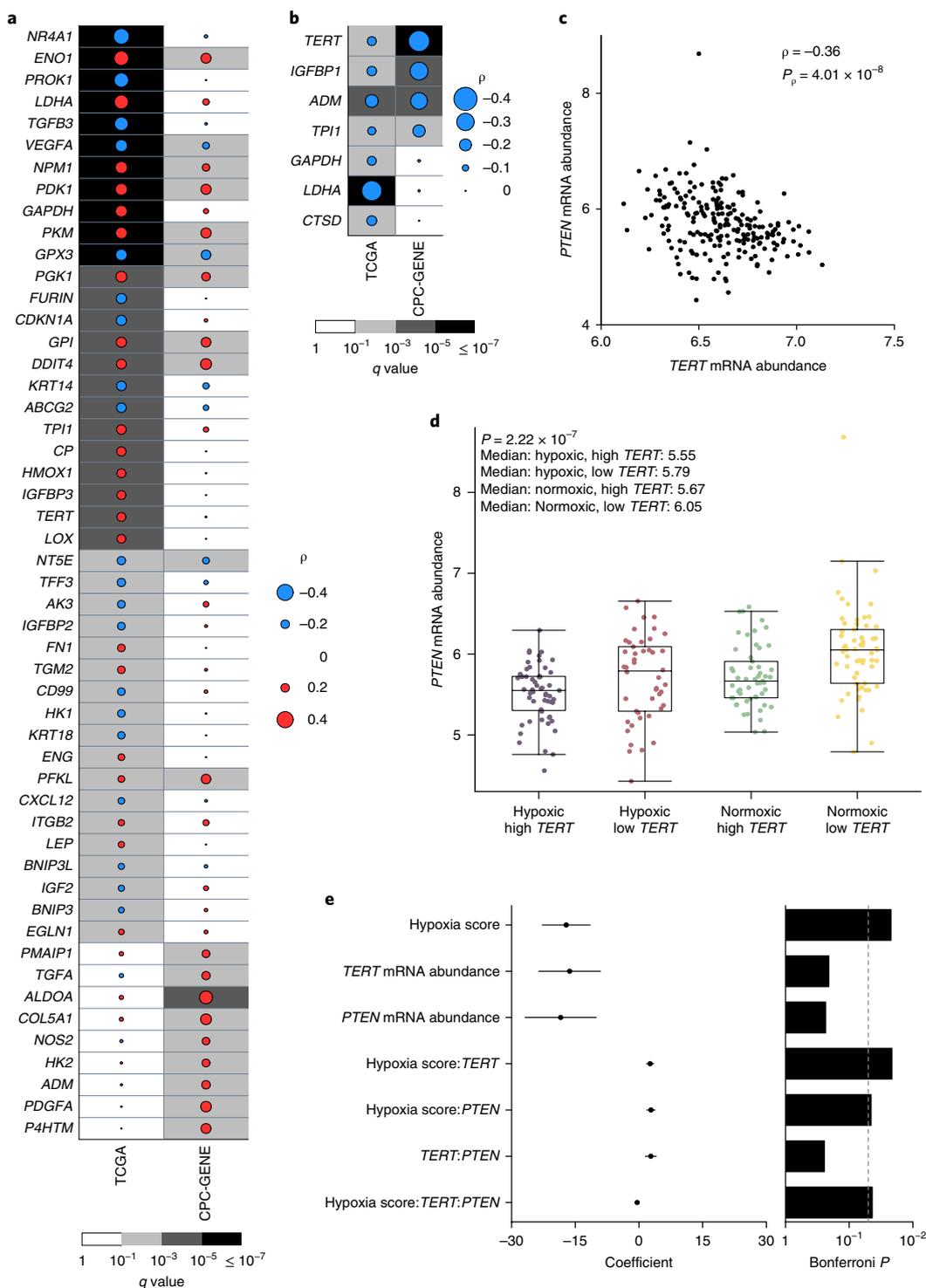
One of the strongest gene–hypoxia associations was allelic loss of the tumor suppressor *PTEN*, which was correlated with significantly elevated tumor hypoxia (FDR =  $2.69 \times 10^{-4}$ , odds ratio (OR)<sub>loss/neutral</sub> = 3.50, 95% confidence interval (CI), 2.14–5.79, Fisher's exact test; Fig. 4a). This loss was part of a focal deletion including *RNLS* and *ATAD1*, all of which were associated with reduced mRNA abundance in both the CPC-GENE (FDR =  $9.26 \times 10^{-11}$ , Mann–Whitney  $U$  test; Fig. 4b) and TCGA (FDR =  $1.26 \times 10^{-7}$ , Mann–Whitney  $U$  test; Supplementary Fig. 6b) cohorts. In another finding further validating this observation, tumors with higher hypoxia had lower *PTEN* mRNA abundance in a third cohort of 130 prostate tumors<sup>50</sup> ( $P=9.65 \times 10^{-7}$ ,  $\rho=-0.41$ , AS89; Fig. 4c). Tumors with allelic *PTEN* loss also exhibited elevated PGA ( $P=3.26 \times 10^{-10}$ , Mann–Whitney  $U$  test; Fig. 4d). Subjects whose prostate tumors had both loss of *PTEN* and high hypoxia were at significantly higher risk of biochemical relapse within 2 years, which is strongly associated with prostate cancer–specific mortality<sup>57</sup>, even after controlling for T category, Gleason score and pretreatment PSA (hazard ratio = 4.4, 95% CI = 1.7–11.0,  $P=1.95 \times 10^{-3}$ , Wald test; Fig. 4e). Thus, *PTEN* loss may preferentially occur in hypoxic tumors, as either an adaptive or a selective effect, and this is associated with elevated genomic instability and aggressive disease.

The strongest single clinicomolecular correlate of tumor hypoxia was the aggressive subpathology of intraductal and cribriform carcinoma (IDC-CA). We previously proposed the concept of a 'nimbo-sus' (Latin for 'gathering of stormy clouds') in prostate cancer related to the subpathologies of IDC-CA<sup>58,59</sup>. Nimbo-sus refers to a series of unfavorable indices such as IDC-CA and tumor hypoxia

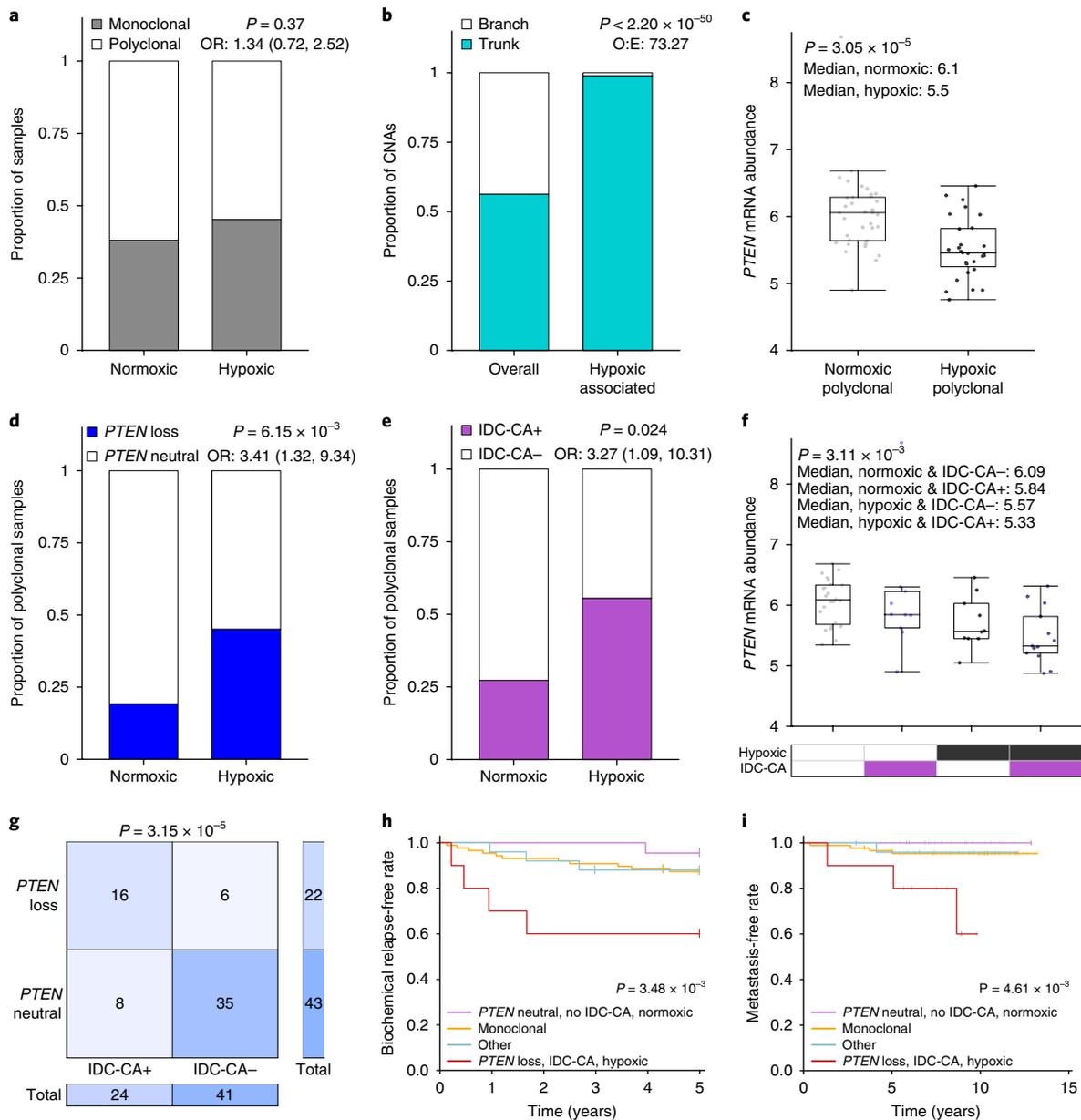
that tend to co-occur, resembling a field defect, and that are associated with poor prognosis. Here, IDC-CA was associated with elevated PGA ( $P=2.96 \times 10^{-10}$ , Mann–Whitney  $U$  test; Fig. 4f; ref. <sup>58</sup>) and with high hypoxia ( $P_{\text{CPC-GENE}}=3.46 \times 10^{-4}$ ;  $P_{\text{TCGA}}=1.13 \times 10^{-11}$ , Mann–Whitney  $U$  test; Fig. 4a and Supplementary Fig. 5d,e). Importantly, subjects with tumors harboring both a high hypoxia score and IDC-CA had a significantly elevated risk of relapse 2 years after primary therapy, even after controlling for T category, Gleason score and pretreatment PSA (hazard ratio = 5.4, 95% CI = 1.9–15,  $P=1.59 \times 10^{-3}$ , Wald test; Fig. 4g). Collectively, these analyses suggest that tumor hypoxia potentially underpins several aggressive genomic and pathological features in localized prostate cancer.

**Interactions of hypoxia in modulating telomere length.** To explore how hypoxia may influence *PTEN* mRNA abundance, we examined known downstream targets<sup>60</sup> of hypoxia-inducible factor 1 $\alpha$  (HIF1A). First, we associated hypoxia scores with the mRNA abundance of HIF1A-activated genes and observed a strong correlation for 51 of 85 HIF1A targets in the CPC-GENE or TCGA cohorts (AS89; Fig. 5a). To identify candidate *PTEN*-regulated genes, we assessed the correlation of each of the 51 hypoxia-associated HIF1A targets with *PTEN* mRNA abundance, and identified seven genes with strong negative associations (AS89; Fig. 5b). These genes included *TERT*, whose mRNA abundance was negatively correlated with that of *PTEN* in both the CPC-GENE ( $\rho=-0.36$ ,  $p=4.01 \times 10^{-8}$ , AS89; Fig. 5c) and TCGA cohorts ( $\rho=-0.15$ ,  $P=7.21 \times 10^{-3}$ , AS89; Supplementary Fig. 5f). Together, hypoxia and *TERT* mRNA abundance additively predicted *PTEN* mRNA abundance in both cohorts (Fig. 5d and Supplementary Fig. 5g). The lowest *PTEN* mRNA abundance was observed under high hypoxia and high *TERT* mRNA abundance. Given that hypoxia-associated CNAs alter telomere-related pathways (Supplementary Fig. 6d), we estimated telomere length in each sample by using TelSeq<sup>61</sup>. Telomere length was related to hypoxia, *PTEN* and *TERT* mRNA abundance ( $P_{\text{model}}=2.17 \times 10^{-3}$ , linear model; Fig. 5e). Notably, although there was no direct relationship between hypoxia and telomere length (Fig. 4a), a model incorporating hypoxia, *PTEN* and *TERT* mRNA abundance demonstrated a significant interaction between these features in modulating telomere length (Bonferroni-adjusted  $P_{\text{interaction}}=4.34 \times 10^{-2}$ , linear model, Fig. 5e, bottom). Thus the effect of hypoxia score on telomere length was different for different values of *TERT* and *PTEN* mRNA abundance, and there was some additional effect with respect to telomere length that was not captured by considering only the individual effects of these features. Collectively, our data show that hypoxic prostate tumors are associated with dysregulated *PTEN*, which is strongly correlated with decreased *TERT* expression and shortening of telomeres.

**Hypoxia and tumor subclonal evolution.** Given that hypoxic prostate tumors are both genomically unstable and enriched for specific mutations, we next sought to understand whether the evolutionary life histories of prostate tumors were also influenced by hypoxia. We evaluated 191 localized prostate tumors that had both hypoxia scores and subclonal architecture reconstructions available<sup>35</sup>. Hypoxia was not directly associated with the number of cancer cell populations detected in the index lesion (Fig. 6a), although, in agreement with hypoxia being an early selective event, monoclonal tumors with high hypoxia scores showed elevated *TERT* mRNA abundance (Supplementary Fig. 5h). Furthermore, we previously described a number of genes in which CNAs show biased evolutionary timing<sup>55</sup>. That is, CNAs in some genes tend to occur early during tumor development (that is, in the trunk of the tumor) whereas CNAs in other genes tend to occur later (that is, in the branch of the tumor). Notably, 99% (660 of 667) of the CNAs associated with hypoxia, including *PTEN*, that showed biased evolutionary timing



**Fig. 5 | Interactions among hypoxia, HIF1A targets and PTEN in modulating telomere length.** We examined how HIF1A targets influence *PTEN* mRNA abundance and how these features interact to influence telomere length. **a**, Correlations of hypoxia score and mRNA abundance for HIF1A targets in CPC-GENE and TCGA. Tumor hypoxia was significantly associated with the mRNA abundance of 51 HIF1A targets in either the TCGA or CPC-GENE cohorts. Dot size indicates the magnitude of  $\rho$ . Background color indicates  $q$  value. **b**, mRNA abundance correlations between HIF1A targets and *PTEN* for genes negatively correlated with *PTEN* mRNA abundance. Dot size indicates magnitude of  $\rho$ . Background color indicates  $q$  value. **c**, *PTEN* mRNA abundance is negatively correlated with mRNA abundance of *TERT*, a HIF1A target, in CPC-GENE. **d**, *PTEN* mRNA abundance levels differ depending on hypoxia status and *TERT* mRNA abundance in CPC-GENE (Kruskal–Wallis test). A Tukey box plot is shown. Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5 $\times$  the IQR. **e**, Relationship of telomere length, hypoxia score and mRNA abundance of *TERT* and *PTEN* (linear model). These features were modeled in one model, and each row corresponds to a separate coefficient (center values) from the model. Lines around each dot represent standard error. Bonferroni-adjusted  $P$  values are at right. Hypoxia, *TERT* and *PTEN* significantly interact to influence telomere length.  $n_{\text{TCGA}} = 333$  independent tumors and  $n_{\text{CPC-GENE}} = 215$  independent tumors for **a–d**;  $n = 155$  independent tumors for **e**. Correlations in **a–c** represent Spearman’s  $\rho$ , and  $P$  values were calculated by using algorithm AS89.



**Fig. 6 | Subclonal evaluation of hypoxia.** **a**, Hypoxia is not associated with the number of cancer cell populations detected in an index lesion ( $n = 187$  independent tumors). **b**, Of the CNAs associated with hypoxia that showed biased evolutionary timing, 99% preferentially occurred early during tumor evolution, thus underscoring hypoxia as an early event in prostate tumor development ( $n = 330$  independent tumors). O:E, the observed ratio of hypoxia associated CNAs in the trunk and branch relative to the expected ratio of hypoxia associated CNAs in the trunk and branch. **c**, Hypoxic polyclonal tumors<sup>55</sup> have lower *PTEN* mRNA abundance than do normoxic polyclonal tumors ( $n = 65$  independent tumors; Mann-Whitney *U* test). **d**, Polyclonal tumors that are hypoxic are enriched for allelic loss of *PTEN* ( $n = 103$  independent tumors). **e**, Hypoxic polyclonal tumors are enriched for IDC-CA ( $n = 71$  independent tumors). **f**, IDC-CA status together with hypoxia is associated with *PTEN* mRNA abundance in polyclonal tumors ( $n = 57$  independent tumors; Kruskal-Wallis test). **g**, Allelic loss of *PTEN* and IDC-CA often co-occur ( $n = 65$  independent tumors). **h,i**, Examining 5-year biochemical relapse-free survival on the basis of hypoxia, subclonal architecture, IDC-CA and *PTEN* copy-number status identified groups of subjects with distinct outcomes after controlling for PSA concentration, Gleason score and T category. Subjects with hypoxic polyclonal tumors with IDC-CA and a loss of *PTEN* displayed poor prognosis, whereas those with normoxic tumors with neutral *PTEN* and no IDC-CA had a favorable prognosis. These results were recapitulated in an examination of metastasis-free survival ( $n = 144$  independent tumors; Wald test). Fisher's exact test was used for **a,b,d,e,g**. Tukey box plots are shown in **c,f**. Box plots represent the median (center line) and upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within  $1.5 \times$  the IQR. All tests were two sided.

preferentially occurred early during tumor evolution (observed/expected = 73,  $P = 6.71 \times 10^{-249}$ , hypergeometric test; Fig. 6b).

Given that polyclonal tumors show very poor outcomes relative to monoclonal ones<sup>55</sup> we next focused on understanding the role of hypoxia in polyclonal samples. In 108 tumors with polyclonal

architecture, elevated hypoxia was associated with lower *PTEN* mRNA abundance ( $P = 3.05 \times 10^{-5}$ , Mann-Whitney *U* test; Fig. 6c) and allelic loss of *PTEN* (OR = 3.41, 95% CI = 1.32–9.34,  $P = 6.15 \times 10^{-3}$ , Fisher's exact test; Fig. 6d). The functionality of *PTEN* loss was confirmed by examining hypoxia and *PTEN* loss

together, because the hypoxic polyclonal samples with a *PTEN* loss had the lowest *PTEN* mRNA abundance ( $P=1.86\times 10^{-5}$ , Kruskal–Wallis test; Supplementary Fig. 5i). In agreement with clinicohypoxia correlations (Fig. 3a), polyclonal samples with high hypoxia were also significantly enriched for IDC-CA (OR=3.27, 95% CI=1.09–10.31,  $P=0.024$ , Fisher's exact test; Fig. 6e). Indeed, *PTEN* mRNA abundance was additively associated with both hypoxia and IDC-CA ( $P=3.11\times 10^{-3}$ , Kruskal–Wallis test; Fig. 6f) and *PTEN*-deleted tumors were significantly more likely to harbor IDC-CA (OR=11.10, 95% CI=3.02–47.27,  $P=3.15\times 10^{-5}$ , Fisher's exact test; Fig. 6g).

Although monoclonal tumors showed many of the same hypoxia-associated patterns as polyclonal tumors (Supplementary Fig. 6e–g), these data in conjunction with previous work suggest that a constellation of co-occurring molecular features (nimbus) are associated with aggressive disease, including hypoxia, tumor polyclonality, loss of *PTEN* function and IDC-CA. Subjects with all these adverse prognostic features, which reflect tumor evolution trajectories, genomics, pathological morphology and the microenvironment, were at the highest risk of rapid biochemical failure (Fig. 6h) and metastasis (Fig. 6i), even after controlling for T category, Gleason score and pretreatment PSA.

## Discussion

Tissue hypoxia is an important differentiating metabolic characteristic between normal and malignant tissues, and it leads to aggressive tumor cell phenotypes<sup>19,62,63</sup>. According to our calculations of hypoxia scores in 8,006 tumors spanning 19 tumor types, there is extensive intertumor- and intratumor-type heterogeneity in tumor hypoxia. This suggests that subsets of patients with a range of solid tumors with genetic instability may benefit from hypoxia-targeting therapy. Clinical trials of hypoxia-targeting agents should therefore focus on solid tumors with both elevated hypoxia and associated genetic instability, similarly to how trials of other targeted therapies have focused on subjects with a specific mutation. This push for careful subject selection may be further modified if our discovery of ancestry-specific variations in tumor hypoxia is validated.

The metabolic reprogramming of cells via hypoxia<sup>64</sup> can now be associated with a series of distinct genomic alterations. Some molecular aberrations show intertumor-type variations in their associations with hypoxia, whereas others, such as SNVs and CNAs in *TP53*, show a strong consistency as a pancancer correlate of hypoxic tumors. *TP53* mutations are enriched in hypoxic tumors within each breast cancer subtype, thus supporting the idea that they are a genomic consequence of tumor hypoxia<sup>21</sup>. This highlights the need to further examine the somatic-mutational architecture of other cancer types, such as pediatric tumors, in relation to hypoxia in large, well-powered data sets. Notably, hypoxia is much more tightly associated with CNA profiles than SNV profiles, and this may relate to the differential timing of these events in different tumor types<sup>65</sup>. We also observed vast hypoxia-associated dysregulation of miRNAs across cancers, and additional work should characterize the role of these miRNAs in canonical hypoxia-response pathways.

We also characterized the whole-genome correlates of hypoxia in prostate cancer. Our results suggest that hypoxic prostate cancers are associated with nimbus<sup>58</sup>, genomic instability, mutant *TP53*, allelic loss of *PTEN*, chromothripsis and shorter telomeres. Previous work suggests that loss of *PTEN* mediates HIF1A stabilization<sup>66</sup>; careful in vitro modeling is required to further delineate the relationship between these features. In considering these data, a model of prostate cancer aggression under hypoxia emerges (Supplementary Fig. 6h). Hypoxia applies a selective pressure to the tumor. Surviving subclones with aggressive features, such as loss of *PTEN* and mutant *TP53*, rapidly expand after normal oxygen levels are reestablished, culminating in poor prognosis for these localized tumors. This is supported by a previous longitudinal analysis of

primary and metastatic prostate cancer in which *TP53* mutations were detected in low-frequency subclones in primary tumors and enriched in matched metastatic samples<sup>67</sup>. Microenvironmental pressures may therefore be required to drive aggressive prostate cancer, especially given its long natural life history. Overall, these data show that the tumor microenvironment can leave distinctive somatic-mutational signatures, influencing tumor evolution, aggressivity and response to therapy.

**URLs.** CDF files, [http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/CDF\\_download.asp](http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/CDF_download.asp).

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41588-018-0318-2>.

Received: 21 April 2018; Accepted: 20 November 2018;  
Published online: 14 January 2019

## References

- Wilson, W. R. & Hay, M. P. Targeting hypoxia in cancer therapy. *Nat. Rev. Cancer* **11**, 393–410 (2011).
- Harris, A. L. Hypoxia: a key regulatory factor in tumour growth. *Nat. Rev. Cancer* **2**, 38–47 (2002).
- Bristow, R. G. & Hill, R. P. Hypoxia and metabolism: hypoxia, DNA repair and genetic instability. *Nat. Rev. Cancer* **8**, 180–192 (2008).
- 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).
- Zanella, 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).
- Luoto, K. R., Kumareswaran, R. & Bristow, R. G. Tumor hypoxia as a driving force in genetic instability. *Genome Integr.* **4**, 5 (2013).
- Brizel, D. M., Dodge, R. K., Clough, R. W. & Dewhirst, M. W. Oxygenation of head and neck cancer: changes during radiotherapy and impact on treatment outcome. *Radiother. Oncol.* **53**, 113–117 (1999).
- Nordsmark, M. & Overgaard, J. Tumor hypoxia is independent of hemoglobin and prognostic for loco-regional tumor control after primary radiotherapy in advanced head and neck cancer. *Acta Oncol.* **43**, 396–403 (2004).
- Noman, M. Z. et al. Hypoxia: a key player in antitumor immune response. A review in the theme: cellular responses to hypoxia. *Am. J. Physiol. Cell Physiol.* **309**, C569–C579 (2015).
- Mohyeldin, A., Garzón-Muvdi, T. & Quiñones-Hinojosa, A. Oxygen in stem cell biology: a critical component of the stem cell niche. *Cell Stem Cell* **7**, 150–161 (2010).
- Eliasson, P. & Jönsson, J.-I. The hematopoietic stem cell niche: low in oxygen but a nice place to be. *J. Cell Physiol.* **222**, 17–22 (2010).
- Lin, P.-Y. et al. Expression of hypoxia-inducible factor-1 $\alpha$  is significantly associated with the progression and prognosis of oral squamous cell carcinomas in Taiwan. *J. Oral Pathol. Med.* **37**, 18–25 (2007).
- Rankin, E. B. & Giaccia, A. J. Hypoxic control of metastasis. *Science* **352**, 175–180 (2016).
- Gilkes, D. M. & Semenza, G. L. Role of hypoxia-inducible factors in breast cancer metastasis. *Future Oncol.* **9**, 1623–1636 (2013).
- Zhong, H. et al. Overexpression of hypoxia-inducible factor 1 $\alpha$  in common human cancers and their metastases. *Cancer Res.* **59**, 5830–5835 (1999).
- 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).

23. Bindra, R. S. et al. Down-regulation of rad51 and decreased homologous recombination in hypoxic cancer cells. *Mol. Cell Biol.* **24**, 8504–8518 (2004).
24. 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).
25. Koshiji, M. et al. Hif-1 $\alpha$  induces genetic instability by transcriptionally downregulating muts $\alpha$  expression. *Mol. Cell* **17**, 793–803 (2005).
26. 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).
27. 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).
28. Janssens, G. O. et al. Accelerated radiotherapy with carbogen and nicotinamide for laryngeal cancer: results of a phase III randomized trial. *J. Clin. Oncol.* **30**, 1777–1783 (2012).
29. 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).
30. 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).
31. 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).
32. 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).
33. Eustace, A. et al. A 26-gene hypoxia signature predicts benefit from hypoxia-modifying therapy in laryngeal cancer but not bladder cancer. *Clin. Cancer Res.* **19**, 4879–4888 (2013).
34. Sorensen, B. S., Toustrup, K., Horsman, M. R., Overgaard, J. & Alsner, J. Identifying pH independent hypoxia induced genes in human squamous cell carcinomas *in vitro*. *Acta Oncol. (Madr.)* **49**, 895–905 (2010).
35. Elvidge, G. P. et al. Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition. *J. Biol. Chem.* **281**, 15215–15226 (2006).
36. Hu, Z. et al. A compact VEGF signature associated with distant metastases and poor outcomes. *BMC Med.* **7**, 9 (2009).
37. Seigneuric, R. et al. Impact of supervised gene signatures of early hypoxia on patient survival. *Radiother. Oncol.* **83**, 374–382 (2007).
38. Hoskin, P. J. et al. Carbogen and nicotinamide in the treatment of bladder cancer with radical radiotherapy. *Br. J. Cancer* **76**, 260–263 (1997).
39. Milholland, B., Auton, A., Suh, Y. & Vijg, J. Age-related somatic mutations in the cancer genome. *Oncotarget* **6**, 24627–24635 (2015).
40. Yuan, Y. et al. Comprehensive characterization of molecular differences in cancer between male and female patients. *Cancer Cell* **29**, 711–722 (2016).
41. 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).
42. Hieronymus, H. et al. Copy number alteration burden predicts prostate cancer relapse. *Proc. Natl Acad. Sci. USA* **111**, 11139–11144 (2014).
43. Volland, H. K. M. et al. A tumor DNA complex aberration index is an independent predictor of survival in breast and ovarian cancer. *Mol. Oncol.* **9**, 115–127 (2015).
44. Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
45. Pereira, B. et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat. Commun.* **7**, 11479 (2016).
46. Tan, E. Y. et al. The key hypoxia regulated gene CAIX is upregulated in basal-like breast tumours and is associated with resistance to chemotherapy. *Br. J. Cancer* **100**, 405–411 (2009).
47. Lawrence, M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
48. Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
49. Huang, X. et al. Hypoxia-inducible mir-210 regulates normoxic gene expression involved in tumor initiation. *Mol. Cell* **35**, 856–867 (2009).
50. Taylor, B. S. et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
51. Sakamuro, D., Elliott, K. J., Wechsler-Reya, R. & Prendergast, G. C. BIN1 is a novel MYC-interacting protein with features of a tumour suppressor. *Nat. Genet.* **14**, 69–77 (1996).
52. Edwards, Y. H., Putt, W., Fox, M. & Ives, J. H. A novel human phosphoglucomutase (pgm5) maps to the centromeric region of chromosome 9. *Genomics* **30**, 350–353 (1995).
53. Milosevic, M. et al. Tumor hypoxia predicts biochemical failure following radiotherapy for clinically localized prostate cancer. *Clin. Cancer Res.* **18**, 2108–2114 (2012).
54. Fraser, M. et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359–364 (2017).
55. Espiritu, S. M. G. et al. The evolutionary landscape of localized prostate cancers drives clinical aggression. *Cell* **173**, 1003–1013 (2018).
56. Hopkins, J. F. et al. Mitochondrial mutations drive prostate cancer aggression. *Nat. Commun.* **8**, 656 (2017).
57. Jackson, W. C. et al. Intermediate endpoints after postprostatectomy radiotherapy: 5-year distant metastasis to predict overall survival. *Eur. Urol.* <https://doi.org/10.1016/j.eururo.2017.12.023> (2018).
58. 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).
59. Chua, M. L. K., van der Kwast, T. H. & Bristow, R. G. Intraductal carcinoma of the prostate: anonymous to ominous. *Eur. Urol.* **67**, 496–503 (2017).
60. Benita, Y. et al. An integrative genomics approach identifies hypoxia inducible factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Res.* **37**, 4587–4602 (2009).
61. Ding, Z. et al. Estimating telomere length from whole genome sequence data. *Nucleic Acids Res.* **42**, e75 (2014).
62. Semenza, G. L. Hypoxia, clonal selection, and the role of hif-1 in tumor progression. *Crit. Rev. Biochem. Mol. Biol.* **35**, 71–103 (2000).
63. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
64. Haider, S. et al. Genomic alterations underlie a pan-cancer metabolic shift associated with tumour hypoxia. *Genome Biol.* **17**, 140 (2016).
65. Gerstung, M. et al. The evolutionary history of 2,658 cancers. Preprint at <https://www.biorxiv.org/content/early/2017/08/30/161562> (2017).
66. Zundel, W. et al. Loss of PTEN facilitates HIF-1-mediated gene expression. *Genes Dev.* **14**, 391–396 (2000).
67. Hong, M. K. H. et al. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. *Nat. Commun.* **6**, 6605 (2015).

## Acknowledgements

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 funded by the Movember Foundation (grant RS2014-01). P.C.B. was supported by a Terry Fox Research Institute New Investigator Award, a Prostate Cancer Canada Rising Star Fellowship, and a Canadian Institutes of Health Research (CIHR) New Investigator Award. This work has been funded by Fellowships from the CIHR and the Ontario government to V.B. and E.L. S.K.L. is supported as a Movember Rising Star award recipient funded by the Movember Foundation (grants RS2014-03, D2015-12 and D2017-1811), the Telus Motorcycle Ride For Dad (Huronia Branch) and a Ministry of Research and Innovation Early Researcher Award. The authors thank the Princess Margaret Cancer Centre Foundation and Radiation Medicine Program Academic Enrichment Fund for support (to R.G.B.). This work was supported by a Terry Fox Research Institute Program Project Grant. R.G.B. is supported as 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. The authors thank all members of the Boutros and Bristow laboratories for helpful suggestions.

## Author contributions

V.B. conducted bioinformatics and statistical analysis; V.B., L.Y.L., E.L., J.L., R.L., Y.J.S., S.M.G.E., L.E.H., F.Y., V.H., T.N.Y., C.Q.Y. and V.Y.S. performed data processing; C.H., J.R., T.V. and X.H. performed *in vitro* experiments; V.B. performed data visualization; M.F., M.L.K.C., T.v.d.K., S.K.L., P.C.B. and R.G.B. supervised research; V.B., P.C.B. and R.G.B. initiated the project; V.B. wrote the first draft of the manuscript; all authors approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41588-018-0318-2>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

## Methods

**Pancancer hypoxia scoring.** Level 3 mRNA abundance data were downloaded for each of the 19 tumor types (28 January 2016 data release). Hypoxia scores were calculated by using mRNA-based signatures developed by Buffa<sup>40</sup>, Winter<sup>31</sup>, Ragnum<sup>32</sup>, West<sup>33</sup>, Sorensen<sup>34</sup>, Elvidge<sup>35</sup>, Hu<sup>36</sup> and Seigneuric<sup>37</sup>, as previously described<sup>26</sup>. mRNA abundance data for all genes in a hypoxia signature were extracted from each of the tumor types. Signature-specific mRNA abundance data from all 19 tumor types were joined and scored as one cohort ( $n = 7,791$  tumors) to compare hypoxia across tumor types (Supplementary Table 1). Tumors with the top 50% of mRNA abundance values for each gene in a signature were given a score of +1, and tumors with the bottom 50% of mRNA abundance values for that gene were given a score of -1. This procedure was repeated for every gene in the signature to generate a hypoxia score for each subject by using each signature. High scores suggested that the tumor was hypoxic, whereas low scores were indicative of normoxia.

We assessed the correlation of hypoxia scores across all eight independent hypoxia signatures. For each pair of signatures, the Spearman's  $\rho$  for hypoxia scores was calculated for each of the 19 tumor types. This was done to avoid bias from Simpson's paradox. To compare hypoxia rankings across signatures with different ranges, the median hypoxia scores for the 19 tumor types were scaled from -1 to 1 by using the *plotrix* package (v3.7). To assess the association of hypoxia with age, sex and ancestry within each tumor type, hypoxia scores were generated by using only the subjects from that tumor type (Supplementary Table 1).

Hypoxia scores were also generated by using protein abundance data for three data sets where proteomic data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) were available. Data were available for breast invasive carcinoma<sup>68</sup>, ovarian serous cystadenocarcinoma<sup>69</sup>, and colon adenocarcinoma and rectum adenocarcinoma<sup>70</sup>. Hypoxia was scored independently in each data set by using the Buffa<sup>40</sup> hypoxia signature and compared to mRNA-abundance-based hypoxia scores also generated by using only subjects from that tumor type.

**Pancancer associations of sex, ancestry and hypoxia.** Differences in tumor hypoxia according to age and sex were assessed in 18 and 14 tumor types, respectively, by using the Buffa<sup>40</sup>, Winter<sup>31</sup> and Ragnum<sup>32</sup> hypoxia signatures. Hypoxia was associated with age for all tumor types for which age data were available (Spearman's  $\rho$ , AS89). Age data were not available for subjects in the pheochromocytoma and paraganglioma (PCPG) cohort. The association of hypoxia and sex was tested in all 14 tumor types for which data from males and females were available (Mann-Whitney *U* test).

**Molecular features associated with hypoxia in breast, renal and lung cancer.** The association of hypoxia with CNAs and nonsynonymous SNVs was tested in subjects with breast invasive carcinoma, lung adenocarcinoma and renal clear cell carcinoma, by using all subjects who had both CNA and SNV data ( $n_{BRCA} = 960$ ,  $n_{LUAD} = 475$ ,  $n_{KIRC} = 431$ ). CNA and SNV biases were assessed for all genes for which data were available within these tumor types. CNA biases were tested for 24,776 genes in each of the tumor types. SNV biases were tested for 15,899, 17,211 and 11,194 genes in the breast invasive carcinoma, lung adenocarcinoma and renal clear cell carcinoma cohorts, respectively. Copy-number changes were associated with hypoxia by a comparison of hypoxia scores between tumors that were copy-number neutral to those with a copy-number gain (and separately, to those with a loss; Mann-Whitney *U* test). A Bonferroni *P*-value adjustment was applied within each tumor-type cohort. For each gene with SNV data, hypoxia scores were compared between tumors with an SNV and those without an SNV (Mann-Whitney *U* test). A Bonferroni *P*-value adjustment was applied within each tumor-type cohort.

**Pancancer associations of driver SNVs and hypoxia.** Broader pancancer driver SNV associations were assessed within all 19 tumor types for genes previously defined as drivers<sup>47</sup>. Differences in hypoxia scores were compared between tumors with or without a nonsynonymous SNV in each driver gene within each tumor type (Mann-Whitney *U* test). An FDR *P*-value adjustment was applied within each cohort. The top 50 driver SNVs were plotted according to the number of tumor types in which each driver event had  $P < 0.1$ . Consensus clustering was applied to the genes and tumor types for visualization purposes by using the ConsensusClusterPlus<sup>71</sup> package (v1.24.0) (genes: distance = Euclidean,  $pItem = 0.75$ ,  $finalLinkage = Ward.D2$ ,  $innerLinkage = Ward.D2$ ; tumor types: distance = Euclidean,  $pItem = 0.75$ ,  $finalLinkage = Ward.D2$ ,  $innerLinkage = Ward.D2$ ).

**Pancancer associations of driver CNAs and hypoxia.** Driver CNA associations were assessed within all 19 tumor types for previously described oncogenes and tumor-suppressor genes<sup>48</sup>. For each oncogene, differences in hypoxia scores were compared between tumors that were copy-number neutral and those with a copy-number gain (Mann-Whitney *U* test). For each tumor-suppressor gene, differences in hypoxia scores were compared between tumors that were copy-number neutral and those with a copy-number loss (Mann-Whitney *U* test). An FDR *P*-value adjustment was applied within each cohort. The top 50 driver CNAs were plotted according to the number of tumor types in which each driver event had  $P < 0.1$ .

Consensus clustering was applied to the genes and tumor types for visualization purposes (genes: distance = Euclidean,  $pItem = 0.75$ ,  $finalLinkage = Ward.D2$ ,  $innerLinkage = Ward.D2$ ; tumor types: distance = Euclidean,  $pItem = 0.50$ ,  $finalLinkage = Ward.D2$ ,  $innerLinkage = Ward.D2$ ).

**Pancancer associations of miRNAs and hypoxia.** The association of hypoxia with miRNA abundance was assessed in 18 of 19 tumor types of interest. The association of hypoxia with miRNAs was not assessed in the GBM cohort because none of the GBM tumors with hypoxia data had miRNA data. Within each of the other tumor types, hypoxia score was tested for an association with each miRNA that had an abundance  $> 0$  in at least 10% of the tumors with miRNA and hypoxia data (Spearman's  $\rho$ , AS89). An FDR *P*-value adjustment was applied within each tumor type cohort. The top 50 miRNAs were plotted according to the number of tumor types in which each miRNA had an FDR-adjusted  $P < 0.1$ . Consensus clustering was applied to the miRNAs and tumor types for visualization purposes (miRNAs: distance = Jaccard,  $pItem = 0.75$ ,  $finalLinkage = Ward.D2$ ,  $innerLinkage = Ward.D2$ ; tumor types: distance = Euclidean,  $pItem = 0.75$ ,  $finalLinkage = Ward.D2$ ,  $innerLinkage = Ward.D2$ ).

**Prostate cancer cell culture and hypoxia.** Human prostate adenocarcinoma (22Rv1, PC3 and DU145) cells were purchased from the American Type Culture Collection (Life Sciences Reporting Summary). Cells were cultured as follows: 22Rv1 cell lines in RPMI-1640 medium; PC3 and DU145 cell lines in DMEM. Both RPMI-1640 and DMEM culture media were supplemented with 10% FBS (Invitrogen) and penicillin (100 U ml<sup>-1</sup>)-streptomycin (100  $\mu$ g ml<sup>-1</sup>) (Invitrogen). Cells were incubated in 5% CO<sub>2</sub> at 37°C. Cells were passaged when they reached 70–90% confluence and were tested regularly to verify the absence of mycoplasma contamination using the MycoAlert kit (Lonza). For hypoxia, cells were incubated at 5% CO<sub>2</sub>, 1% O<sub>2</sub> and 94% N<sub>2</sub> in a HERAccl 150i CO<sub>2</sub> Incubator (Thermo Fisher Scientific) at 37°C. The influence of hypoxia on miR-210-3p and miR-133a-3p abundance was assessed with a one-sided paired Student's *t* test.

**Transfection of miR-133a-3p mimics.** Cells were seeded in six-well plates at a density of  $2 \times 10^5$  cells per well for PC3 and DU145, and  $5 \times 10^5$  cells per well for 22Rv1. miR-133a-3p mimics (5  $\mu$ M; Shanghai GenePharma) were transfected into cells 24 h after seeding, by using Lipofectamine 2000 Reagent (Invitrogen) and Opti-MEM I (1 $\times$ ) reduced serum medium (Invitrogen), according to the manufacturer's recommendations. Cells were collected 24 h after transfection for subsequent experiments. miRNA mimic sequences are in Supplementary Table 6.

**Quantitative real-time PCR cDNA synthesis.** Quantitative real-time PCR was performed with a miScript SYBR Green PCR Kit for miRNA transcripts (Qiagen). miRNA expression was calculated through the comparative Ct method with an Applied Biosystems StepOnePlus Real-Time PCR System (Thermo Fisher Scientific).

For miR-133a-3p expression under hypoxia, transfected 22Rv1, DU145 and PC3 cells were incubated at 1% O<sub>2</sub>. After 72 h, miRNA was extracted, cDNA was synthesized and quantitative rtPCR was performed as described above. miR-210-3p, a known hypoxia-induced miRNA<sup>49</sup>, was used as a positive control. Primer sequences are available in Supplementary Table 6 (Invitrogen).

**Proliferation assays.** To assess the effect of miR-133a-3p on cell proliferation, transfected cells were seeded in triplicate at a density of  $5 \times 10^4$  cells per well of a six-well plate for PC3 and DU145, and at  $1 \times 10^5$  cells per well for 22Rv1. Cells were trypsinized and collected 4 d later for all cell lines. Viable cells were measured with a Countess automated cell counter (Life Technologies). The influence of miR-133a-3p on proliferation was assessed with a one-sided paired Student's *t* test.

**Invasion assays.** To evaluate the invasive potential of miR-133a-3p, we seeded cells transfected with control or miR-133a-3p mimics (PC3,  $1 \times 10^5$  cells; DU145,  $1 \times 10^5$  cells; 22Rv1,  $5 \times 10^5$  cells) in low-serum medium into Matrigel-coated Transwell inserts (BD Biosciences). Below the insert, 10% DMEM was used as a chemoattractant. Cells invading through the Matrigel-coated inserts were fixed and stained with DiffQuik (modified Giemsa stain; ThermoFisher Scientific) at 24 h for PC3, 20 h for DU145, and 48 h for 22Rv1. Invading cells were imaged and counted with a Leica DM LB2 microscope (Leica Microsystems). The influence of miR-133a-3p on invasion was assessed with a one-sided paired Student's *t* test.

**CPC-GENE subject cohort.** Subject selection, tissue collection and sample processing has been described in detail<sup>54</sup>. Briefly, informed consent, consistent with the guidelines of the local Research Ethics Board and ICGC, was obtained at the time of clinical follow-up. Previously collected tumor tissues were used, following University Health Network Research Ethics Board-approved study protocols (UHN 06-0822-CE, UHN 11-0024-CE, CHUQ 2012-913:H12-03-192). Tumor cellularity and Gleason grades were evaluated independently by two genitourinary pathologists on scanned haematoxylin- and eosin-stained slides. Serum PSA concentrations (ng ml<sup>-1</sup>) are reported according to the reading at the time of diagnosis. Additional information can be found in the Life Sciences Reporting Summary.

**SNP microarray data generation and analysis.** Microarray data generation and analysis have been described in detail<sup>54</sup>. Briefly, SNP microarrays were performed with 200 ng of DNA on Affymetrix OncoScan FFPE Express 2.0 and 3.0 arrays. Analysis of the probe assays was performed with OSCHP files generated by OncoScan Console (v1.1) using a custom reference<sup>54</sup>. BioDiscovery's Nexus Express for OncoScan 3 Software was used to call copy-number aberrations by using the SNP-FASST2 algorithm. Gene-level copy-number aberrations for each tumor were identified by overlapping copy-number segments, with RefGene (2014-07-15) annotation, by using BEDTools (v2.17.0). PGA was calculated for each sample by dividing the number of base pairs involved in a copy-number change by the total length of the genome.

**mRNA abundance data generation and analysis.** Generation and analysis of mRNA abundance data have been previously described in detail<sup>54</sup>. Briefly, total RNA was extracted with a mirVana miRNA Isolation Kit (Life Technologies), according to the manufacturer's instructions. Total RNA (100–150 ng) was assayed on the Human Gene 2.0 ST array or the Affymetrix Human Transcriptome Array 2.0. Background correction, normalization and annotation were carried out with the oligo (v1.34.2) package. The robust multichip average algorithm was applied to the raw intensity data<sup>72</sup>. Probes were mapped to Entrez gene IDs by using custom CDF files (v20) for HTA 2.0 and HuGene 2.0 ST array (see URLs).

**Whole-genome sequencing and data analysis.** Whole-genome sequencing was conducted as described<sup>54</sup>. Briefly, sequencing libraries were prepared with 50 ng of genomic DNA and enzymatic reagents from KAPA Library Preparation Kits (KAPA Biosystems, KK8201) according to protocols described for end repair, A-tailing and adapter ligation<sup>73</sup>. Sequencing was carried out on the Illumina HiSeq 2000 platform (coverage =  $74 \times \pm 23 \times$ ; Illumina).

Analysis of whole-genome-sequencing data has been described in detail<sup>54</sup>. Briefly, raw sequencing reads were aligned to the human reference genome, GRCh37, by using bwa-mem (v0.7.12+) <sup>74</sup> at the lane level. Picard (v1.92) was used to merge the lane-level binary alignment map (BAM) files from the same library and mark duplicates. Library-level BAMs from each sample were also merged without marking duplicates using Picard. Local realignment and base quality recalibration was carried out on tumor and normal pairs together using the Genome Analysis Toolkit (GATK v3.4.0+) <sup>75</sup>. Tumor and normal sample level BAMs were extracted, headers were corrected using samtools<sup>76</sup> (v0.1.9), and files were indexed with Picard (v1.92).

Identification of somatic SNVs, genomic rearrangements and fusion events was conducted as described<sup>54</sup>. Briefly, SomaticSniper (v1.0.5) was used to call SNVs on callable bases. Coding versus noncoding SNVs were determined with Annovar<sup>77</sup>. Genomic rearrangements were identified with Delly<sup>78</sup> (v0.7.2-0.7.6). Gene fusion events involving *ERG* or E twenty-six variant (ETV) genes were collectively referred to as ETS events. Genomic-rearrangement calls were examined to determine whether breakpoints led to a *TMPRSS2-ERG* fusion or if breakpoints were found in both 1-Mbp bins surrounding the following gene pairs: *ERG-SLC45A3*, *ERG-NDRG1*, *ETV1-TMPRSS2*, *ETV4-TMPRSS2*, *ETV1-SLC45A3*, *ETV4-SLC45A3*, *ETV1-NDRG1* and *ETV4-NDRG1*. *ERG* immunohistochemistry and deletion calls between *TMPRSS2* and *ERG* loci in OncoScan SNP array data provided further support for these fusions.

**Associating hypoxia with features of localized prostate cancer.** Hypoxia scores for subjects with localized prostate adenocarcinoma from TCGA and CPC-GENE were generated with the Ragnum<sup>72</sup> signature. Subjects from TCGA ( $n=333$ ) and CPC-GENE ( $n=215$ ) were scored independently (Supplementary Table 1 and 4). High scores suggested that the tumor was hypoxic, whereas low scores indicated normoxia. In addition to the continuous hypoxia scores from each cohort, the tumors with the top 50% of hypoxia scores were classified as hypoxic, whereas the remaining tumors were classified as normoxic.

We assessed whether hypoxia was associated with ETS fusion status, kataegis, chromothripsis, the total number of nonsynonymous SNVs, the total number of genomic rearrangements and PGA. Kataegis was predicted with the SeqKat (v0.0.1) R package. Chromothripsis scores were generated with ShatterProof<sup>79</sup> (v0.14) with default settings. Hypoxia was associated with a previously described set of recurrent translocations, inversions and coding nonsynonymous SNVs<sup>54</sup>. The association of hypoxia with noncoding SNVs was assessed for the five most recurrent noncoding SNVs in the CPC-GENE cohort. All tests were two sided, unless otherwise stated. For Spearman's correlations,  $P$  values were calculated by using algorithm AS89.

Telomere length was estimated with TelSeq<sup>81</sup> (v0.0.1) on BAM files generated using (bwa-mem v 0.7.12+<sup>74</sup>, GATK<sup>75</sup>). As a quality measure, TelSeq estimates for each sample were generated per sequencing lane. Reads from lanes that contained too few reads to calculate an estimate and outlier lanes, as identified by Grubbs' test, were removed from input BAMs by using BAMQL<sup>80</sup>. After outliers were removed, TelSeq was run again, ignoring read groups with the -u parameter. Samples with telomere estimates <1 were removed from that analysis.

**Hypoxia associated CNAs in localized prostate cancer.** CNA data were available for 360 subjects from TCGA and CPC-GENE for 23,741 genes. Genes with CNAs

in <5% of the cohort were discarded. The remaining 6,121 genes were associated with hypoxia status (Fisher's exact test), and 1,189 genes with an FDR <0.1 were carried forward for confirmation using mRNA abundance data, which was available for 909 of 1,189 genes. mRNA abundance values for each of the 909 genes were correlated with hypoxia scores independently in the TCGA and CPC-GENE cohorts (Spearman's  $\rho$ , AS89). Of the 909 genes, 101 genes had a  $\rho \geq 0.15$  in both cohorts and an FDR <0.05 in at least one cohort. mRNA abundance values were compared independently in the TCGA and CPC-GENE cohorts between tumors that were copy-number neutral and those that had a CNA for these 101 genes. Of the genes tested, 20 genes had significantly different mRNA abundance values between tumors that were copy-number neutral and those that had a CNA (FDR <0.05).

**Survival analyses.** For subjects who underwent surgery, biochemical recurrence was defined as two consecutive post-surgery PSA measurements >0.2 ng ml<sup>-1</sup> (backdated to the date of the first increase), as long as the subject had a post-surgery PSA <0.2 ng ml<sup>-1</sup>; otherwise, this was considered a primary treatment failure rather than a biochemical recurrence. If a subject had successful salvage radiation therapy, the subject was not considered to have a biochemical recurrence. If PSA continued to rise after salvage radiation therapy, biochemical recurrence was backdated to the first PSA measurement >0.2 ng ml<sup>-1</sup>. If a subject received other salvage treatment (such as hormonal therapy or chemotherapy), this was considered a biochemical recurrence. For subjects who underwent precision image-guided radiotherapy, biochemical recurrence was defined as a PSA rise of  $\geq 2.0$  ng ml<sup>-1</sup> above the nadir PSA. The proportional hazards assumption was verified by examining Schoenfeld residuals.

**Subclonal assessment of hypoxia in localized prostate cancer.** The subclonal architecture of 293 tumors with localized prostate cancer has been characterized by using TITAN<sup>81</sup> and PhyloWGS<sup>82</sup>, as described before<sup>55</sup>. Briefly, bwa-aln (v0.5.7) was used to align raw sequencing reads to the reference genome, and TITAN was used to predict CNA segments and tumor cellularities. SomaticSniper was used to call SNVs on callable bases. Using default parameters, PhyloWGS was used to generate possible tree structures for each sample. All possible trees were parsed to determine the best tree on the basis of the largest log likelihood value. A subset of the tumors had only clonal mutations (termed monoclonal), whereas others had multiple tumor populations that originated from one ancestral clone (termed polyclonal). To associate subclonal architecture with tumor hypoxia, we considered subjects with Ragnum hypoxia scores as well as subjects who underwent image-guided radiotherapy and had intraprostatic  $p_{O_2}$  measurements.

Intraprostatic measurements of  $p_{O_2}$  were measured before radiotherapy by using an ultrasound-guided transectal needle piezoelectrode, as described before<sup>83</sup>. Briefly, between 40 and 80 individual oxygen measurements were obtained along two to four measurement tracks from regions of the prostate gland expected to contain tumor, on the basis of previous diagnostic biopsies.  $HP_{10}$  values (the percentage of  $p_{O_2}$  measurements <10 mm Hg) were used for the hypoxia analyses.  $p_{O_2}$  measurements ranged from 0 to 1. Ragnum hypoxia scores were scaled from 0 to 1 with the formula  $y = (x - \min(x)) / (\max(x) - \min(x))$ , where  $x$  is the Ragnum hypoxia score, and  $y$  is the scaled hypoxia score. Tumors with an  $HP_{10}$  value  $\geq 0.5$  or a scaled Ragnum hypoxia score  $\geq 0.5$  were considered hypoxic, whereas the remaining tumors were considered normoxic. Scaled hypoxia values and subclonality information were available for 191 tumors. Four samples classified as polytumor were excluded from subclonality analyses<sup>55</sup>. Subclonality-related analyses involving mRNA abundance data examined only CPC-GENE subjects.

All data analysis was performed in the R statistical environment (v3.3.1). Data visualization was performed with the lattice (v0.20-34), latticeExtra (v0.6-28) and BPG<sup>84</sup> (v5.6.21) packages. Box plots represent the median (center line), upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5 $\times$  the interquartile range (Tukey box plots). Figures were compiled by using Inkscape (v0.91). Pathway analysis was completed by using g:Profiler (v0.6.7)<sup>85</sup> and visualized by using Cytoscape (v3.4.0).

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

## Data availability

The raw sequencing data have been deposited in the European Genome-phenome Archive under accession code EGAS00001000900. Processed variant calls are available through the ICGC Data Portal under the project PRAD-CA. TCGA data are available at <https://portal.gdc.cancer.gov/projects/TCGA-PRAD>. Previously published CPC-GENE data are available at the European Genome-phenome Archive under accession code EGAS00001000900. Previously published CPC-GENE mRNA abundance data are available at the Gene Expression Omnibus under accession code GSE84043.

## References

68. Mertins, P. et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).

69. Zhang, H. et al. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell* **166**, 755–765 (2016).
70. Zhang, B. et al. Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382–387 (2014).
71. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
72. Irizarry, R. A. et al. Summaries of affymetrix genechip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
73. Fisher, S. et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* **12**, R1 (2011).
74. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
75. McKenna, A. et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
76. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
77. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
78. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
79. Govind, S. K. et al. ShatterProof: operational detection and quantification of chromothripsis. *BMC Bioinformatics* **5**, 78 (2014).
80. Masella, A. P. et al. BAMQL: a query language for extracting reads from BAM files. *BMC Bioinformatics* **17**, 305 (2016).
81. Ha, G. et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* **24**, 1881–1893 (2014).
82. Deshwar, A. G. et al. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* **16**, 35 (2015).
83. Parker, C. et al. Polarographic electrode study of tumor oxygenation in clinically localized prostate cancer. *Int. J. Radiat. Oncol.* **58**, 750–757 (2004).
84. P'ng, C. et al. BPG: seamless, automated and interactive visualization of scientific data. Preprint at <https://www.biorxiv.org/content/early/2017/06/26/156067> (2017).
85. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler: a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated
- Clearly defined error bars  
*State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on [statistics for biologists](#) may be useful.

### Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used to collect data.

Data analysis

R statistical environment (v3.3.1), ConsensusClusterPlus (v1.24.0), OncoScan Console (v1.1), BEDTools (v2.17.0), oligo (v1.34.2), bwamem (v0.7.12+), Picard (v1.92), samtools (v0.1.9), SomaticSniper (v1.0.5), Delly (v0.7.2-0.7.6), SeqKat (v0.0.1), ShatterProof (v0.14), TelSeq (v0.0.1), lattice (v0.20-34), latticeExtra (v0.6-28), plotrix (v3.7).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The raw sequencing data have been deposited in the European Genome-phenome Archive under accession EGAS00001000900 (<https://www.ebi.ac.uk/ega/studies/>)

EGAS00001000900). Processed variant calls are available through the ICGC Data Portal under the project PRAD-CA (<https://dcc.icgc.org/projects/PRAD-CA>). TCGA data is available at <https://portal.gdc.cancer.gov/projects/TCGA-PRAD>. Previously published CPC-GENE data is available at the European Genome-phenome Archive under the accession: EGAS00001000900. Previously published CPC-GENE mRNA abundance data is available at the Gene Expression Omnibus database under GEO:GSE84043.

## Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	In our pan-cancer analysis we analyzed all tumours with mRNA abundance data from 19 tumour types where data was available (n = 7,791). In the CPC-GENE cohort, we considered all patients with mRNA abundance (n = 215) or needle hypoxia data (n = 116 for subclonality related analyses). For our pan-cancer analyses describing hypoxia associated CNAs and SNVs in three tumour types, we calculated statistical power for all tumour types and chose three that were well powered with hypoxia-PGA correlations (n-BRCA = 960, n-LUAD = 475, n-KIRC = 431).
Data exclusions	Four samples classified as polytumour were excluded from the subclonality analyses. This was done because subclonal reconstructions of polytumour samples are challenging (please see Espiritu et al., Cell 2018 for more details). This exclusion criteria was established before the analysis was carried out.
Replication	We reproduced our findings related to breast cancer (TCGA) in an independent 1,859 patient cohort (METABRIC). We replicated many of our findings from the CPC-GENE cohort in the TCGA prostate dataset, where data was available (n=333). We also replicated a key finding in prostate cancer in an independent 130 patient cohort (Taylor et al., Cell 2010). Finally, we validated miR-133a as a hypoxia-associated miRNA in vitro to confirm our findings from primary prostate tumours.
Randomization	Samples were allocated to groups based on mutational status or hypoxia levels, if necessary. The mutational profiles and hypoxia levels of individuals were not known prior to inclusion in the study.
Blinding	All participants were recruited into the study based only on clinical features. Therefore, mutational profiles and hypoxia levels were not known upon inclusion.

## Reporting for specific materials, systems and methods

### Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

### Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	All cells were purchased from the American Type Culture Collection (ATCC, USA). The cell lines used in this study were: 22Rv1, DU145, PC3.
Authentication	ATCC ensures cell line authentication prior to distribution using morphology, karyotyping, PCR-based approaches, COI analyses and STR profiling.
Mycoplasma contamination	Cells were tested regularly to verify the absence of mycoplasma contamination using the MycoAlert kit (Lonza).
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	None of the cell lines used are listed in the ICLAC database.

## Human research participants

---

Policy information about [studies involving human research participants](#)

### Population characteristics

Prostate cancer patients ranged in age from 42 – 76 years. All prostate cancer patients were diagnosed with localized disease and were primarily classified as having intermediate risk prostate cancer. All prostate cancer patients were male.

### Recruitment

We obtained Informed consent, consistent with the guidelines of the local Research Ethics Board (REB) and International Cancer Genome Consortium at the time of clinical follow-up. Previously collected tumour tissues were used, following University Health Network REB-approved study protocols (UHN 06-0822-CE, UHN 11-0024-CE, CHUQ 2012-913:H12-03-192). The CPC-GENE study focused on patients with localized prostate cancer. All prostate cancer samples were obtained from the University Health Network (UHN) Pathology BioBank or from the Genito-Urinary BioBank of the Centre Hospitalier Universitaire de Québec (CHUQ). Most patients enrolled into CPC-GENE were those with intermediate risk prostate cancer. While the TCGA prostate cancer patients span the different risk categories for prostate cancer, some of the results presented may be biased to reflect the biology of intermediate risk prostate cancer.