

Integrated Genomic Analysis of Hypoxia Genes across Cancer Types Identifies Significant Associations with Cancer Hallmarks

Lingjian Yang¹, Laura Forker¹, Christina S. Fjeldbo², Robert G. Bristow¹, Heidi Lyng²,
Catharine M. L. West^{1,3*}

Author affiliations: ¹Division of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Christie Hospital, Manchester, M20 4BX, UK. ²Department of Radiation Biology, Norwegian Radium Hospital, Oslo University Hospital, Pb 4950 Nydalen, Oslo 0424, Norway. ³NIHR Manchester Biomedical Research Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, UK.

***Corresponding author:** Email: Catharine.west@manchester.ac.uk

Key words: hypoxia, pan cancer analysis, integrative analysis, omics data, RNA sequencing, microRNA

ABSTRACT

Hypoxia is a generic micro-environmental factor in most solid tumours. While most published literature focused on *in vitro* or single tumour type investigations, we carried out the first multi-omics pan cancer analysis of hypoxia with the aim of gaining a comprehensive understanding of its implication in tumour biology. A core set of 52 mRNAs were curated based on experimentally validated hypoxia gene sets from multiple cancer types. The 52 mRNAs collectively stratified high- and low-hypoxia tumours from The Cancer Genome Atlas (TCGA) database (9698 primary tumours) in each of the 32 cancer types available. High-hypoxia tumours had high expression of not only mRNA but also protein and microRNA markers of hypoxia. In a pan cancer transcriptomic analysis, $\geq 70\%$ of the known cancer hallmark pathways were enriched in high-hypoxia tumours, most notably epithelial mesenchymal transition potential, proliferation (G2M checkpoint, E2F targets, MYC targets) and immunology response. In a multi-omics analysis, gene expression-determined high-hypoxia tumours had a higher non-silent mutation rate, DNA damage repair deficiency and leukocyte infiltration. The associations largely remained significant after correcting for confounding factors, showing a profound impact of hypoxia in tumour evolution across cancer types. High-hypoxia tumours determined using the core gene set had a poor prognosis in 16/32 cancer types, with statistical significances remaining in five after adjusting for tumour stage and omics biomarkers. In summary, this first comprehensive *in vivo* map of hypoxia in cancers highlights the importance of this micro-environmental factor in driving tumour progression.

INTRODUCTION

Hypoxia, low oxygen tension, is a characteristic feature of solid tumours and a hallmark of cancer. Cancer cells adapt to hypoxia via processes including HIF activation and unfolded protein response signalling which alter the expression of genes involved in multiple pathways such as angiogenesis, metabolism, invasion, and epithelial-to-mesenchymal transition ^{1,2}. Hypoxia also increases DNA hypermethylation and genome instability ^{3,4}. The transcriptional reprogramming generates a pro-survival advantage to hypoxic cancer cells and results in adverse phenotypes, including resistance to therapy and a higher potential to metastasise ^{5,6}.

Given its crucial role in tumour progression and resistance to therapy, hypoxia has been explored as a therapeutic target. Approaches for targeting hypoxia have been successfully developed and include increasing oxygen delivery, sensitising hypoxic cells to radiation or using bio-reductive agents that are selectively cytotoxic to hypoxic tumour cells ⁷⁻⁹. The effect of hypoxia modification has been supported by high level evidence from clinical trials ^{6,10}, making it arguably the most validated target yet to be translated into the clinic ⁷.

Despite increasing efforts to study omics response to hypoxia ¹¹⁻¹⁵, most of the literature focused on *in vitro* investigation and limited to single or a small number of tumour types. The *in vivo* impact of hypoxia on tumour genetics and its relationship with other cancer hallmarks are still largely unexplored. To date, the largest *in vivo* multiple cancer types hypoxia gene expression studies were performed by Chi et al., ¹⁶ and Buffa et al. ¹⁷. However, both studies were limited to three cancer types and transcriptome data only. Here we evaluated a set of core hypoxia genes in a large-scale pan-cancer multi-omics analysis. We first defined a core gene set with a high confidence of hypoxia inducibility in multiple cell types. In The Cancer Genome Atlas (TCGA) programme containing mRNA expression of 9698 primary tumours from 32 cancer types, tumours with consistently high expression of core hypoxia genes were identified in each cancer type. High-hypoxia tumours had high expression of other well-known hypoxia markers, including NDRG1 and GAPDH proteins and

miR-210 and miR-21 microRNAs. High-hypoxia tumours were highly enriched with the majority of cancer hallmark pathways, including most notably epithelial mesenchymal transition, proliferation and immunology response. By analysing paired clinical and multi-omics data, we also identified significant associations between hypoxia and tumour stage, homologous recombination deficiency, high DNA mutational load and immune cell infiltrate in multiple cancer types. High-hypoxia tumours were associated with a poor prognosis in 16 cancer types. Overall, our work generates the first comprehensive molecular landscape of hypoxia in human cancers, which highlights the important and independent associations with various cancer hallmarks and also profound impact on patient survival.

RESULTS

Definition of 52 Core Hypoxia Genes

To curate genes that reliably reflected transcriptional response to hypoxia, we examined hypoxia gene expression experimental studies collated in two recent review papers^{14,15}. The study inclusion criteria were: (1) genome-wide identification of hypoxia responsive genes, and (2) use of multiple cell lines (or tissues) to increase the confidence of finding. Similar studies were combined or filtered out to minimise redundancy. In total ten gene sets^{16,18–28} were included (more details in Supplementary Methods). Most of the above literature reported genes both induced or suppressed by hypoxia, and we selected only the hypoxia inducible genes, as defined per the original publications. Furthermore, a hallmark hypoxia gene set was obtained from the Molecular Signatures Database^{29,30}, which was derived from four cell line datasets independent from the above ten studies. Another expert curated hypoxia gene set was taken from a review¹. The 12 founder gene sets represented experimentally validated transcriptional response genes to hypoxia from nine cancer types, *i.e.* head and neck, cervix, sarcoma, breast, neuroblastoma, glioblastoma, prostate, non-small cell lung cancer, and melanoma. The 12 gene sets also included hypoxia response genes in the following normal cells: B cell, epithelium, astrocytes, monocytes, embryonic cell (Supplementary

Table S1). The core hypoxia genes were defined from the founder gene sets as those consistently inducible by hypoxia in multiple ($\geq M$) studies. Increasing the value of M increases the hypoxia specificity of the core gene set, at the cost of decreasing the coverage of the captured hypoxia transcriptional programme. The 52 genes present in at least four studies ($M=4$) made the final core hypoxia gene set and were used in the subsequent data analysis. Different values of M were also attempted with broadly similar results.

The 52 core hypoxia genes overlapped well with most of the 12 founder gene sets (Figure 1A), with the highest overlaps observed with hypoxia genes derived for cervix (79.0%)²¹, head and neck (74.1%)¹⁸, neuroblastoma (68.8%)²⁴ and sarcoma (67.6%)²⁰. Interestingly, all the four gene sets were identified from single cancer types with a large number of cell lines (4 to 11), suggesting that consistently hypoxia inducible genes identified across multiple cell lines in one cancer type are well, albeit not perfectly, preserved. The core hypoxia genes loosely overlapped (12.9%) with the prostate cancer gene set correlating with pimonidazole staining²³, which might either reflect a sharply different hypoxia transcriptional programme in profoundly hypoxic prostate carcinoma or a different type of hypoxia measured by pimonidazole^{31,32}. Details of the 52 core hypoxia genes are provided in Supplementary Table S2.

Core Hypoxia Genes Identified Tumours with High-hypoxia Phenotype

The expression pattern of the 52 core hypoxia inducible genes were used to collectively define the extent of hypoxia in 9698 human primary cancerous tissues from 32 cancer types in the TCGA, a widely used methodology^{16,17,20,33}. In each cancer type, unsupervised clustering (K -means, $K=2$) of the 52 core genes stratified tumours into two groups. In each cancer type, almost all core hypoxia inducible genes were simultaneously elevated in one group of tumours, therefore termed as high-hypoxia gene expression phenotype (Supplementary Figure S1).

We asked whether the high-hypoxia phenotype based on core hypoxia genes captured genome-wide hypoxia transcriptional programme reflected by the founder gene sets in individual cancer types. In each cancer type, genome-wide differential expression analysis between high- and low-hypoxia tumours was performed using LIMMA, where a t statistics was produced for each gene quantifying its relative difference between high-hypoxia and low-hypoxia tumours. Positive t statistics indicate up-regulation in high-hypoxia tumours while negative values indicate down-regulation. Among the 12 founder gene sets, eight were derived from samples (cell lines or primary tumours) whose cancer types were among the 32 available in TCGA. For each of those eight gene sets, we compared the distribution of member gene t statistics with null distribution of genome-wide t statistics in the corresponding cancer type(s). For seven founder gene sets, member genes had significantly higher (t test, $P < 0.001$) t statistics compared to the null distributions (Figure 1B), indicating their significant up-regulation in high-hypoxia tumours. Therefore, the results suggest that the identified high-hypoxia phenotype faithfully reflected the transcriptional response to hypoxia in individual cancer types.

We then asked whether an *in vivo* hypoxia transcriptional program was preserved across cancer types. Similarity of the global hypoxia transcriptional program between two cancer types was computed as the Spearman correlation of the genome-wide t statistics between high- and low-hypoxia tumours. Using an arbitrary cut-off of 0.3, we identified 62 similar pairs of cancer types (Figure 1C). Our results therefore suggest that molecular responses to hypoxia were conserved across certain cancer types.

High-hypoxia Phenotype Associated with Cancer Hallmarks

We next sought to identify which biological processes were consistently associated with an *in vivo* high-hypoxia phenotype in a pan cancer setting. Mixed effects linear regression model was applied for each gene, explaining (Z standardised) gene expression as a function of hypoxia phenotypes (high- or low-), with cancer type included as a random effect variable. The ~20000 mRNAs were

ranked according to their association strengths with hypoxia (coefficients in the regression models), where gene set enrichment analysis (GSEA)²⁹ was employed to identify gene sets whose expressions were enriched or depleted in high-hypoxia tumours.

Among 50 cancer hallmarks, 36 were significantly enriched in high-hypoxia tumours (false discovery rate (FDR)<0.01, Supplementary Table S3), suggesting a large effect of hypoxia on other cancer hallmarks. The top hallmarks enriched in high-hypoxia tumours included not only the well expected hypoxia, glycolysis, angiogenesis and unfolded protein response gene sets, but also epithelial mesenchymal transition (EMT), tumour necrosis factor α (TNF- α) signalling via nuclear factor- κ B (NF- κ B), cell cycle and proliferation (G2M checkpoint, E2F targets, MYC targets) and immunology response (inflammatory response, interferon gamma/alpha response, allograft rejection, Figure 2). The emergence of EMT as the most enriched process suggests an important role of hypoxia in stimulating tumour invasion and metastasis, consistent with previous literature⁵. The identified cell cycle and proliferation pathways were recently suggested in the development of high copy number alteration and mutation phenotypes^{34,35}. Our results indicated that some of the genome instability associated with dysregulated cell cycle and proliferation pathways might be due to hypoxia. P53 pathway genes were also significantly up-regulated in the high-hypoxia tumours, in line with a recent report on a p53-dependent pro-apoptotic pathway responding to severe hypoxia³⁶. Moreover, TNF- α and NF- κ B signalling control the pro-inflammatory response and have been implicated in tumour cell invasion and metastasis^{37,38}. Our data suggests that hypoxia contributes to tumour progression via activating TNF- α and NF- κ B signalling and manipulating inflammatory response. Whether the TNF- α inhibitors could be successfully re-purposed to modify hypoxia in primary cancerous tissues will be of interest³⁹.

As tumour purity might be a confounding factor in the genomic analysis⁴⁰, we re-ran the above mixed linear regression model with purity corrected estimates. Tumour purity was previously

inferred using ABSOLUTE method on copy number alteration data ³⁴, and was benchmarked histologically ⁴¹. Correcting for tumour purity only marginally affected the GSEA analysis results.

Proteins and MicroRNAs Associated with High-hypoxia Phenotype

To gain more biological understanding of the high-hypoxia phenotype, we analysed functional proteomic and microRNA (miRNA) abundance data for their differential expression between high-hypoxia and low-hypoxia tumours. Abundance of 198 cancer-related proteins and phosphoproteins was measured for TCGA tumours using reverse-phase protein arrays (RPPA) ⁴². To identify the proteins significantly and consistently associated with hypoxia in a pan cancer setting, we again applied a mixed effects linear regression model for each protein, predicting (Z standardised) its abundance as a function of hypoxia phenotype, with cancer type included as a random effect variable. Thanks to the robust statistical framework employed here, we identified 42 and 85 proteins that were respectively up-regulated and down-regulated in high- vs. low-hypoxia tumours (FDR<0.01, Supplementary Table S4). The large number of down-regulated proteins in high-hypoxia tumours is consistent with a global decrease of protein synthesis under hypoxia via reduced mRNA translational efficiency ⁴³.

Among the top proteins enriched in high-hypoxia tumours, N-Myc Downstream Regulated 1 and Glyceraldehyde-3-Phosphate Dehydrogenase mRNAs (*NDRG1* and *GAPDH*) were included in the core hypoxia gene set. In addition, mRNA of plasminogen activator inhibitor-1 protein (*SERPINE1*) is stress regulated and was shown to be inducible by hypoxia in a murine macrophage cell line ⁴⁴. Accumulation of epidermal growth factor receptor (EGFR) protein under hypoxia and by over-expressing hypoxia-inducible factor 2- α (HIF2- α) was also reported previously in glioma, breast, prostate cancer cell lines ^{45,46}. Fibronectin binds to extracellular matrix, and has significant functions in cell adhesion, migration and growth. Hypoxia was shown to induce fibronectin in mouse embryonic stem cells, promoting their proliferation and migration ⁴⁷. Moreover, annexin 1

(ANNEXIN1), involved in anti-inflammatory effects, was found to be hypoxia inducible in a breast cancer cell line and highly expressed in triple-negative tumours⁴⁸.

Among the proteins most down-regulated in high-hypoxia tumours, B-cell lymphoma 2 (BCL2) is apoptosis-regulating and resides in the mitochondrial membrane. A previous study suggested that hypoxia induced autophagy by up-regulating BCL2 interacting protein 3 like (*BNIP3L*) and BCL2 interacting protein 3 (*BNIP3*) and disrupting the BCL-2-Becclin1 complex, which generates a survival advantage under hypoxia⁴⁹. Another study showed that down-regulation of BCL2 might be mediated through miR-210, a well-known hypoxia inducible miRNA⁵⁰. Another protein highly down-regulated in the high-hypoxia gene expression phenotype was estrogen receptor- α (ER- α), which regulates hormone responsiveness and was shown to be inhibited by hypoxia in breast cancer ER-positive cell lines^{51–53}. Our result was also in line with the clinical observation of an anti-correlation between ER- α and HIF-1 α ^{53,54}. In a phase II trial, HIF-1 α was implicated in the resistance to endocrine treatment for ER- α -positive breast cancer patients⁵⁵. Our results therefore suggest a possibly important association between hypoxia and ER- α in cancer types beyond breast carcinoma, and the potential benefit of adding inhibitors targeting HIF signalling to mediate endocrine sensitivity. Taken together, our pan cancer protein analysis suggests that certain known hypoxia-mediated cellular processes observed previously in individual cancer types could be generalised to multiple cancer types.

MiRNAs are short non-coding RNAs that have important function in mRNA post-transcriptional silencing and are frequently dysregulated in cancers⁵⁶. Abundance of more than 700 miRNAs were profiled previously for TCGA tumours⁵⁷. We applied a mixed effects linear regression model to identify the miRNAs consistently differentially expressed between high- and low-hypoxia tumours in a pan cancer setting. We identified 331 miRNAs induced by hypoxia and another 161 miRNAs suppressed by hypoxia (FDR<0.01, Supplementary Table S5). Among the top hypoxia inducible miRNAs identified by the model, hsa-miR-210 is arguably the most well-known hypoxia miRNA biomarker in the literature^{11,12,58}. In addition, hsa-miR-21 was found to be induced by hypoxia in

both normal smooth muscle cells and a pancreatic cancer cell line^{59,60}. Also, four of seven miRNAs shown to be hypoxia inducible in two bladder cancer cell lines¹² ranked highly in our pan cancer analysis.

Among the miRNAs most suppressed in hypoxia gene expression phenotypes, hsa-miR-139 expression was reported to be lost in invasive breast cancer tumours, while its *in vitro* over-expression suppressed invasion and migration⁶¹. Our analysis suggested that the effect of has-miR-139 might be mediated through hypoxia-related signalling. A direct regulatory relationship was also reported between hsa-miR-30a and autophagy-related BCL2 and Beclin-1 in small cell lung cancer and head and neck cancers^{62,63}. Hsa-miR-101 is a well-documented tumour suppressor that is repressed by the pro-inflammatory cytokine IL-1 β via the COX2-HIF-1 α pathway⁶⁴. Our analyses thus identify some key miRNAs that are probably universally important in mediating hypoxia responses across cancer types and are potential therapeutic targets.

High-hypoxia Phenotype Associated Independently with High Tumour Stage, Genome Instability, DNA Damage Repair Deficiency, and Immune Infiltrate

While a role of hypoxia in promoting tumour aggressiveness, metastasis and genome instability was identified from studies on single cancer types, we highlight in this work its importance across cancer types. Our pan-cancer evaluation aimed to quantify associations between a high-hypoxia phenotype and multiple aggressive features characterised in previous studies^{34,65–68}: tumour stage, mutational load, homologous recombination deficiency, leukocyte infiltrate and aneuploidy score. Briefly, mutations were called from whole-exome sequencing data and non-silent mutations per million base pairs were used for mutation rate. Mutation rate was log₂ transformed to normalise the skewed distribution. Homologous recombination deficiency score was summarised from three types of homologous recombination deficiency or genome scarring, *i.e.* loss of heterozygosity, large-scale state transitions and the number of telomeric allelic imbalances⁶⁸. Leukocyte score reflected the

total level of tumour immune infiltrate and was derived from methylation data. Aneuploidy score was the number of chromosome arms with arm-level copy number alterations.

American joint committee on cancer pathology stages were used in most cancer types. For CESC, DLBC, OV, UCEC, UCS and THYM, only clinical stages were available and therefore used instead. In eight of the 27 cancer types where tumour stages were available, high-hypoxia tumours were associated with higher tumour stage (Fisher's exact test, $FDR < 0.1$, Figure 3A). One interesting question is whether hypoxia was associated with higher tumour stage independently of other confounding factors. We performed multiple regression analysis predicting tumour stage (modelled as a discrete variable from 1 to 4) as a function of hypoxia phenotype, mutation load, homologous recombination deficiency score, aneuploidy score and leukocyte infiltrate score. High-hypoxia was independently associated with high tumour stage in nine cancer types ($FDR < 0.1$, Figure 3B). The results, combined with various pre-clinical evidences of hypoxia in increasing tumour aggressiveness⁶⁹ and the success of hypoxia-targeting treatments in limiting tumour progression⁷⁰, indicate a role of hypoxia in driving tumour progression across multiple cancer types.

Various studies reported increased DNA mutation rate among *in vitro* cell lines exposed to hypoxia⁴. We observed significantly higher mutation rates in high-hypoxia tumours across nine cancer types (*t* test, $FDR < 0.1$, Figure 3C). In multiple regression models, hypoxia was adjusted with the above mentioned variables and also tumour stage. Tumour stage was modelled as an ordered categorical variable using stage 1 as the reference level, while the effects of stage 2, 3, and 4 against stage 1 were estimated. The approach was chosen to better model the non-linear associations between tumour stage and molecular biomarkers. Hypoxia was significantly associated with a higher mutation rate independent of other variables in three cancer types, including thymoma, pancreatic adenocarcinoma, uterine corpus endometrial carcinoma ($FDR < 0.1$, Figure 3D).

High-hypoxia phenotype was also associated with increased homologous recombination deficiency in seven cancer types (t test, $FDR < 0.1$, Figure 3E). In multiple regression models, the association with hypoxia was retained in five cancer types ($FDR < 0.1$, Figure 3F).

Immune infiltrate and hypoxia are both important parts of the tumour micro-environment. Our above transcriptome-based hallmark analysis showed associations between hypoxia and various immunology response pathways. To further explore this finding, we evaluated the association between a high-hypoxia gene expression phenotype and leukocyte level determined from methylation data. Overall leukocyte content in the TCGA tumours was estimated using methylation probes which had the greatest difference between leukocyte cells and normal tissues in a mixture model⁶⁶. High-hypoxia phenotype was associated with significantly more leukocytes in 14 cancer types and significantly less leukocyte levels in four cancer types (t test, $FDR < 0.1$, Figure 3G). After correcting for confounding factors, hypoxia was associated with higher leukocytes in nine ($FDR < 0.1$, Figure 3H) and lower leukocyte levels in one cancer type. While the associations between tumour mutation rate, copy number alteration burden and response to immunotherapy were recently characterised^{34,71}, our data suggested an important yet previously less explored association between hypoxia and the tumour immunological landscape.

All the above analyses were repeated with correction for tumour purity as a confounding factor, which only marginally affected the results (data not shown). Overall, our analyses suggest a very profound association between hypoxia and tumour genetics at multiple molecular levels.

Hypoxia is A Poor Prognostic factor in Multiple Cancer Types

To investigate the impact of hypoxia on patient survival, we next evaluated the prognostic value of hypoxia gene expression phenotypes using three clinical endpoints of overall survival, disease specific survival and progression free survival⁶⁷. Patient were treated with surgery, where follow up data were censored at 10-year for each endpoint. In 16 cancer types, hypoxia was a poor prognostic

marker in at least one and often multiple clinical endpoints (Cox regression, $FDR < 0.1$ for each endpoint, Figure 4A). Although hypoxia is widely recognised as a strong mediator of resistance to radiotherapy, this analysis emphasises its importance in surgery-treated patients.

We further adjusted hypoxia with mutation load, homologous recombination deficiency score, aneuploidy score, leukocyte infiltrate score and tumour stage in multivariable Cox models. For each endpoint and each biomarker, the nominal P values from Cox models were corrected using FDR. In five cancer types (kidney renal papillary cell carcinoma, mesothelioma, lung adenocarcinoma, uveal melanoma and breast invasive carcinoma), hypoxia remained a significant poor prognostic factor for at least one clinical endpoint ($FDR < 0.1$, Figure 4B). Note that in the multivariable analyses, the prognostic significances of a high-hypoxia phenotype were comparable to other omics-based biomarkers (Figure 4C), *i.e.* high homologous recombination deficiency was associated with poor prognosis in four cancer types; high aneuploidy score was associated with poor prognosis in one cancer type; high mutational load corresponded with a good prognosis in three cancer types and poor prognosis in another; high leukocyte infiltrate was associated with a poor prognosis in two cancer types. Compared with omics biomarkers, tumour stage remained a stronger prognostic factor, in line with a previous report ⁷². Nevertheless, the consistent effect of hypoxia on patient survival again highlights that it is a clinically relevant phenotype worth intervention.

DISCUSSION

Recent advances in high-throughput technologies and establishment of large consortium like TCGA have enabled integrative and pan cancer analyses of many cancer hallmarks, including somatic mutations ⁷³, somatic copy number alterations ³⁴, DNA damage repair ⁶⁸ and immunology ⁶⁶. These analyses are generating an increasingly clear picture of the whole spectrum of changes associated with cancer. Understanding the impact of hypoxia on tumour biology is a key object in cancer research. By performing the first multi-omics pan cancer analysis of tumour hypoxia we highlight here the profound association between hypoxia and various cancer hallmarks.

It is well accepted in the literature that hypoxia results in large-scale change in transcriptional reprogramming, and the changes differed across cancer types^{15,16,18}. By surveying a large number of experimentally validated hypoxia gene sets, we identified a group of core genes consistently inducible by hypoxia. Note here that we chose to exclude hypoxia gene signatures, as their bioinformatics derivation methods differ substantially and could well reflect other biological parameters, as discussed in a previous review¹⁴. We showed that the primary tumours stratified as high-hypoxia by the highly conserved core gene set not only highly expressed the founder cancer-type-specific hypoxia gene sets in most of the cancer types (albeit perhaps not prostate carcinoma), but also were highly enriched with well-known protein and miRNA hypoxia markers. The results therefore showed that we have identified primary tumours with differential levels of hypoxia. Our pan cancer analysis also suggested that some of the hypoxia-mediated mechanisms reported in individual cancer types or *in vitro* could be generalised to multiple cancer types and *in vivo*.

We observed remarkable associations between hypoxia and many other cancer hallmarks, including high mutation rate and more DNA damage repair deficiency. Our data also showed that high-hypoxia tumours had high epithelial mesenchymal transition potential and cell proliferation signatures, consistent with the literature evidence from single tumour sites^{46,74–76}. Previous studies characterised the relationship between somatic mutations and copy number alterations with an immune landscape^{34,71}. Our analysis showed hypoxia was mostly associated with increased leukocyte content, independent from other confounding factors. It is of interest for future research to investigate whether hypoxia biomarkers could improve the prediction of response to immunology therapy.

Hypoxia has been well recognised as mediating resistance to radiotherapy^{36,50,77} and is arguably the most established target not yet translated into the clinic⁷. Although studies in some cancer types showed hypoxia was an adverse prognostic feature following surgery^{11,58,78,79}, we showed here its widespread importance for surgery-treated patients across tumour types. Importantly, the effect of

hypoxia on surgical outcomes was comparable with other omics marker. Targeting hypoxia and hypoxia signalling, either alone or in combination with other therapies, could well generate a sizable survival benefit for cancer patients.

In summary, our study highlights a high level of conservation of hypoxia response genes across tumour types. Hypoxia has a large effect on multiple cancer hallmarks with EMT, proliferation and immunology response being the most consistently enriched phenotypes. High-hypoxia tumours also have high tumour stage, mutational burden, homologous recombination deficiency and immune cell infiltrate. Our work supports the importance of translating hypoxia modification therapies into the clinic to improve patient survival rates.

METHODS

TCGA Omics Data

TCGA is a large-scale cohort for multi-omics analysis of tumours. Recently, TCGA has published the pan-cancer atlas, where all omics data were harmonised for consistent quality control, batch effect correction, normalisation, mutation calling, and curation of survival data. For this work, we extracted processed RNA sequencing V2, microRNA sequencing, reverse phase protein array and patient clinical and survival data from the Pan-cancer Atlas consortium, which were described elsewhere^{34,66–68}. Only the primary tumours were kept for analysis. We also downloaded paired mutational load, homologous recombination deficiency, aneuploidy score, leukocyte infiltrate, tumour stage and patient survival data from the consortium for evaluating the association with hypoxia.

Stratification and Analyses of Hypoxia Gene Expression Phenotypes

For each cancer type, primary tumours were stratified into two groups (high- and low-hypoxia) by applying unsupervised *K*-means clustering (*K*=2) on the mRNA expressions of the core hypoxia genes. We used *kmeans* function (1000 random starts) in R (v3.4.1). Differential expression analysis was

performed using LIMMA (v3.34.9). To identify the genes consistently associated with high- and low-hypoxia tumours in pan cancer setting, a mixed effect linear model was applied for each gene: $Expression = \beta * Hypoxia + (1|CancerType)$, where hypoxia (1 indicates high-hypoxia and 0 low-hypoxia status) is a fixed effect variable and cancer type a random effect variable. nlme library (v3.1-131) in R was used for estimating β coefficients, which reflect the sign and strength of associations between gene expression and hypoxia. Expressions of each gene were Z standardised to ensure β coefficients were comparable between different genes. When correcting tumour purity as confounding factor, purity value was added as a fixed effect variable: $Expression = \beta * Hypoxia + \gamma * Purity + (1|Cancer_type)$. Genes were ranked from high to low based on β coefficients. GSEA analysis was performed using gene pattern public server and with 50 hallmark gene sets.

Statistical Analysis

Unequal variance *t* test was used to compare if the mean of two distributions were equal. Fisher's exact test was used to test independence of distribution of high- and low-hypoxia tumours across different tumour stages. Survival estimates were performed using the Kaplan-Meier method, and hazard ratios (HR) and 95% confidence interval (CI) were obtained using the Cox proportional hazard model. All *P*-values were two-sided and multiple testing corrections were done using false discovery rate. All statistical analyses were done in R (v3.4.1).

Reference

1. Harris, A. L. Hypoxia — a key regulatory factor in tumour growth. *Nat. Rev. Cancer* **2**, 38 (2002).
2. Wouters, B. G. & Koritzinsky, M. Hypoxia signalling through mTOR and the unfolded protein response in cancer. *Nat. Rev. Cancer* **8**, 851 (2008).
3. Thienpont, B. *et al.* Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* **537**, 63 (2016).
4. Luoto, K. R., Kumareswaran, R. & Bristow, R. G. Tumor hypoxia as a driving force in genetic instability. *Genome Integr.* **4**, 5 (2013).
5. Gilkes, D. M., Semenza, G. L. & Wirtz, D. Hypoxia and the extracellular matrix: drivers of tumour metastasis. *Nat. Rev. Cancer* **14**, 430–439 (2014).
6. Salem, A. *et al.* Targeting Hypoxia to Improve Non–Small Cell Lung Cancer Outcome. *JNCI J. Natl. Cancer Inst.* **110**, djsx160–djsx160 (2018).
7. Wilson, W. R. & Hay, M. P. Targeting hypoxia in cancer therapy. *Nat Rev Cancer* **11**, 393–410 (2011).
8. Stratford, I. J., Williams, K. J., Cowen, R. L. & Jaffar, M. Combining bioreductive drugs and radiation for the treatment of solid tumors. *Semin. Radiat. Oncol.* **13**, 42–52 (2017).
9. Tharmalingham, H. & Hoskin, P. Clinical trials targeting hypoxia. *Br. J. Radiol.* 20170966 (2018).
10. Overgaard, J. Hypoxic Radiosensitization: Adored and Ignored. *J. Clin. Oncol.* **25**, 4066–4074 (2007).
11. E., G. H. *et al.* hsa-miR-210 is a marker of tumor hypoxia and a prognostic factor in head and neck cancer. *Cancer* **116**, 2148–2158 (2010).
12. Blick, C. *et al.* Identification of a hypoxia-regulated miRNA signature in bladder cancer and a role for miR-145 in hypoxia-dependent apoptosis. *Br. J. Cancer* **113**, 634 (2015).
13. Chang, Y.-N. *et al.* Hypoxia-regulated lncRNAs in cancer. *Gene* **575**, 1–8 (2016).
14. Yang, L. & West, C. M. L. Hypoxia gene expression signatures as predictive biomarkers for personalising radiotherapy. *Br. J. Radiol.* 20180036 (2018).
15. Harris, B. H. L., Barberis, A., West, C. M. L. & Buffa, F. M. Gene Expression Signatures as Biomarkers of Tumour Hypoxia. *Clin. Oncol.* **27**, 547–560 (2017).
16. Chi, J.-T. *et al.* Gene Expression Programs in Response to Hypoxia: Cell Type Specificity and Prognostic Significance in Human Cancers. *PLOS Med.* **3**, e47 (2006).
17. 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).
18. Sørensen, 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).

19. Wang, V., Davis, D. A., Haque, M., Huang, L. E. & Yarchoan, R. Differential Gene Up-Regulation by Hypoxia-Inducible Factor-1 α and Hypoxia-Inducible Factor-2 α in HEK293T Cells. *Cancer Res.* **65**, 3299 LP-3306 (2005).
20. Yang, L. *et al.* Validation of a hypoxia related gene signature in multiple soft tissue sarcoma cohorts. *Oncotarget* **9**, (2018).
21. Fjeldbo, C. S. *et al.* Integrative Analysis of DCE-MRI and Gene Expression Profiles in Construction of a Gene Classifier for Assessment of Hypoxia-Related Risk of Chemoradiotherapy Failure in Cervical Cancer. *Clin. Cancer Res.* **22**, 4067 LP-4076 (2016).
22. Halle, C. *et al.* Hypoxia-Induced Gene Expression in Chemoradioresistant Cervical Cancer Revealed by Dynamic Contrast-Enhanced MRI. *Cancer Res.* **72**, 5285 LP-5295 (2012).
23. 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).
24. Fardin, P. *et al.* A biology-driven approach identifies the hypoxia gene signature as a predictor of the outcome of neuroblastoma patients. *Mol. Cancer* **9**, 185 (2010).
25. 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).
26. Sung, F. L. *et al.* Genome-wide expression analysis using microarray identified complex signaling pathways modulated by hypoxia in nasopharyngeal carcinoma. *Cancer Lett.* **253**, 74–88 (2007).
27. Elvidge, G. P. *et al.* Concordant Regulation of Gene Expression by Hypoxia and 2-Oxoglutarate-dependent Dioxygenase Inhibition: THE ROLE OF HIF-1 α , HIF-2 α , AND OTHER PATHWAYS. *J. Biol. Chem.* **281**, 15215–15226 (2006).
28. Aprelikova, O., Wood, M., Tackett, S., Chandramouli, G. V. R. & Barrett, J. C. Role of ETS Transcription Factors in the Hypoxia-Inducible Factor-2 Target Gene Selection. *Cancer Res.* **66**, 5641 LP-5647 (2006).
29. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
30. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).
31. Nordsmark, M. *et al.* Measurements of hypoxia using pimonidazole and polarographic oxygen-sensitive electrodes in human cervix carcinomas. *Radiother. Oncol.* **67**, 35–44 (2003).
32. E., A. R. *et al.* GLUT-1 and CAIX as intrinsic markers of hypoxia in carcinoma of the cervix: Relationship to pimonidazole binding. *Int. J. Cancer* **104**, 85–91 (2003).
33. 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 LP-3449 (2007).
34. Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33**, 676–689.e3 (2018).
35. Buccitelli, C. *et al.* Pan-cancer analysis distinguishes transcriptional changes of aneuploidy

- from proliferation. *Genome Res.* **27**, 501–511 (2017).
36. Leszczynska, K. B. *et al.* Hypoxia-induced p53 modulates both apoptosis and radiosensitivity via AKT. *J. Clin. Invest.* **125**, 2385–2398 (2015).
 37. Wu, Y. & Zhou, B. P. TNF- α /NF- κ B/Snail pathway in cancer cell migration and invasion. *Br. J. Cancer* **102**, 639–644 (2010).
 38. Liu, T., Zhang, L., Joo, D. & Sun, S.-C. NF- κ B signaling in inflammation. *Signal Transduct. Target. Ther.* **2**, 17023 (2017).
 39. Biniecka, M. *et al.* Successful tumour necrosis factor (TNF) blocking therapy suppresses oxidative stress and hypoxia-induced mitochondrial mutagenesis in inflammatory arthritis. *Arthritis Res. Ther.* **13**, R121 (2011).
 40. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).
 41. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
 42. Li, J. *et al.* TCPA: a resource for cancer functional proteomics data. *Nat Meth* **10**, 1046–1047 (2013).
 43. Ho, J. J. D. *et al.* Systemic Reprogramming of Translation Efficiencies on Oxygen Stimulus. *Cell Rep.* **14**, 1293–1300 (2016).
 44. Liao, H., Hyman, M. C., Lawrence, D. A. & Pinsky, D. J. Molecular regulation of the PAI-1 gene by hypoxia: contributions of Egr-1, HIF-1 α , and C/EBP α . *FASEB J.* **21**, 935–949 (2006).
 45. Franovic, A. *et al.* Translational up-regulation of the EGFR by tumor hypoxia provides a nonmutational explanation for its overexpression in human cancer. *Proc. Natl. Acad. Sci.* **104**, 13092 LP-13097 (2007).
 46. Misra, A., Pandey, C., Sze, S. K. & Thanabalu, T. Hypoxia Activated EGFR Signaling Induces Epithelial to Mesenchymal Transition (EMT). *PLoS One* **7**, e49766 (2012).
 47. Hun, L. S., Jin, L. Y. & Jae, H. H. Role of hypoxia-induced fibronectin-integrin β 1 expression in embryonic stem cell proliferation and migration: Involvement of PI3K/Akt and FAK. *J. Cell. Physiol.* **226**, 484–493 (2010).
 48. Okano, M. *et al.* Upregulated Annexin A1 promotes cellular invasion in triple-negative breast cancer. *Oncol. Rep.* **33**, 1064–1070 (2015).
 49. Bellot, G. *et al.* Hypoxia-Induced Autophagy Is Mediated through Hypoxia-Inducible Factor Induction of BNIP3 and BNIP3L via Their BH3 Domains. *Mol. Cell. Biol.* **29**, 2570–2581 (2009).
 50. Sun, Y. *et al.* Hypoxia-induced autophagy reduces radiosensitivity by the HIF-1 α /miR-210/Bcl-2 pathway in colon cancer cells. *Int. J. Oncol.* **46**, 750–756 (2015).
 51. Stoner, M. *et al.* Hypoxia Induces Proteasome-Dependent Degradation of Estrogen Receptor α in ZR-75 Breast Cancer Cells. *Mol. Endocrinol.* **16**, 2231–2242 (2002).
 52. Wolff, M., Kosyna, F. K., Dunst, J., Jelkmann, W. & Depping, R. Impact of hypoxia inducible factors on estrogen receptor expression in breast cancer cells. *Arch. Biochem. Biophys.* **613**, 23–30 (2017).

53. Kurebayashi, J., Otsuki, T., Moriya, T. & Sonoo, H. Hypoxia Reduces Hormone Responsiveness of Human Breast Cancer Cells. *Jpn. J. Cancer Res.* **92**, 1093–1101 (2001).
54. Bos, R. *et al.* Expression of hypoxia-inducible factor-1 α and cell cycle proteins in invasive breast cancer are estrogen receptor related. *Breast Cancer Res.* **6**, R450 (2004).
55. Generali, D. *et al.* Phosphorylated ER α , HIF-1 α , and MAPK Signaling As Predictors of Primary Endocrine Treatment Response and Resistance in Patients With Breast Cancer. *J. Clin. Oncol.* **27**, 227–234 (2009).
56. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509 (2014).
57. Chu, A. *et al.* Large-scale profiling of microRNAs for The Cancer Genome Atlas. *Nucleic Acids Res.* **44**, e3–e3 (2016).
58. Camps, C. *et al.* hsa-miR-210 Is Induced by Hypoxia and Is an Independent Prognostic Factor in Breast Cancer. *Clin. Cancer Res.* **14**, 1340 LP-1348 (2008).
59. Mace, T. A. *et al.* Hypoxia induces the overexpression of MicroRNA-21 in pancreatic cancer cells. *J. Surg. Res.* **184**, 855–860 (2013).
60. Wu, C. *et al.* Hypoxia Potentiates MicroRNA-Mediated Gene Silencing through Posttranslational Modification of Argonaute2. *Mol. Cell. Biol.* **31**, 4760–4774 (2011).
61. Krishnan, K. *et al.* miR-139-5p is a regulator of metastatic pathways in breast cancer. *RNA* **19**, 1767–1780 (2013).
62. Yang, X., Bai, F., Xu, Y., Chen, Y. & Chen, L. Intensified Beclin-1 Mediated by Low Expression of Mir-30a-5p Promotes Chemoresistance in Human Small Cell Lung Cancer. *Cell. Physiol. Biochem.* **43**, 1126–1139 (2017).
63. Saad, M. A. *et al.* Alcohol-dysregulated miR-30a and miR-934 in head and neck squamous cell carcinoma. *Mol. Cancer* **14**, 181 (2015).
64. Wang, L. *et al.* IL-1 β -Mediated Repression of microRNA-101 Is Crucial for Inflammation-Promoted Lung Tumorigenesis. *Cancer Res.* **74**, 4720 LP-4730 (2014).
65. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291–304.e6 (2018).
66. Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* **48**, 812–830.e14 (2018).
67. Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400–416.e11 (2018).
68. Knijnenburg, T. A. *et al.* Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep.* **23**, 239–254.e6 (2018).
69. Semenza, G. L. The hypoxic tumor microenvironment: A driving force for breast cancer progression. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**, 382–391 (2016).
70. Shay, J. E. S. *et al.* Inhibition of hypoxia-inducible factors limits tumor progression in a mouse model of colorectal cancer. *Carcinogenesis* **35**, 1067–1077 (2014).
71. Goodman, A. M. *et al.* Tumor Mutational Burden as an Independent Predictor of Response to

- Immunotherapy in Diverse Cancers. *Mol. Cancer Ther.* **16**, 2598 LP-2608 (2017).
72. Yuan, Y. *et al.* Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat. Biotechnol.* **32**, 644–652 (2014).
 73. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* **45**, 1113–1120 (2013).
 74. HUMAR, R. O. K., KIEFER, F. N., BERNIS, H., RESINK, T. J. & BATTEGAY, E. J. Hypoxia enhances vascular cell proliferation and angiogenesis in vitro via rapamycin (mTOR) -dependent signaling. *FASEB J.* **16**, 771–780 (2002).
 75. Li, L., Li, L., Zhang, Z. & Jiang, Z. Hypoxia promotes bone marrow-derived mesenchymal stem cell proliferation through apelin/APJ/autophagy pathway. *Acta Biochim. Biophys. Sin. (Shanghai)*. **47**, 362–367 (2015).
 76. Sato, M. *et al.* Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations. *Mol. Cancer Res.* **11**, 638–650 (2013).
 77. Toustrup, K. *et al.* Development of a Hypoxia Gene Expression Classifier with Predictive Impact for Hypoxic Modification of Radiotherapy in Head and Neck Cancer. *Cancer Res.* **71**, 5923 LP-5931 (2011).
 78. Forker, L. *et al.* The hypoxia marker CAIX is prognostic in the UK phase III Vortex-Biobank cohort: an important resource for translational research in soft tissue sarcoma. *Br. J. Cancer* (2017).
 79. Yang, L. *et al.* Development and Validation of a 28-gene Hypoxia-related Prognostic Signature for Localized Prostate Cancer. *EBioMedicine* **31**, 182–189 (2018).

Figures and Figure legends

Figure 1. Core hypoxia genes define *in vivo* hypoxia transcriptional programme. **A)** 12 founder gene sets of reliable hypoxia inducible genes were curated from literature. The genes included in more than 4 founder gene sets made up the final core hypoxia gene set. In the Chord diagram, relative circular lengths indicate size of founder gene sets, and thickness of the links between founder gene sets and the core gene set are proportional to the number of member genes from founder set included in the final core gene set. **B)** The distributions of genome-wide t statistics from LIMMA analysis comparing high- vs. low-hypoxia tumours were used as null, where the distributions of t statistics for 8 founder gene sets were significantly higher than the null distributions. **C)** Similarity of hypoxia transcriptional programme between two cancer types was quantified with Spearman correlation of genome-wide t statistics. With an arbitrary cut-off of 0.3, 62 similar pairs could be identified with the strength of association being proportional to the thickness of links.

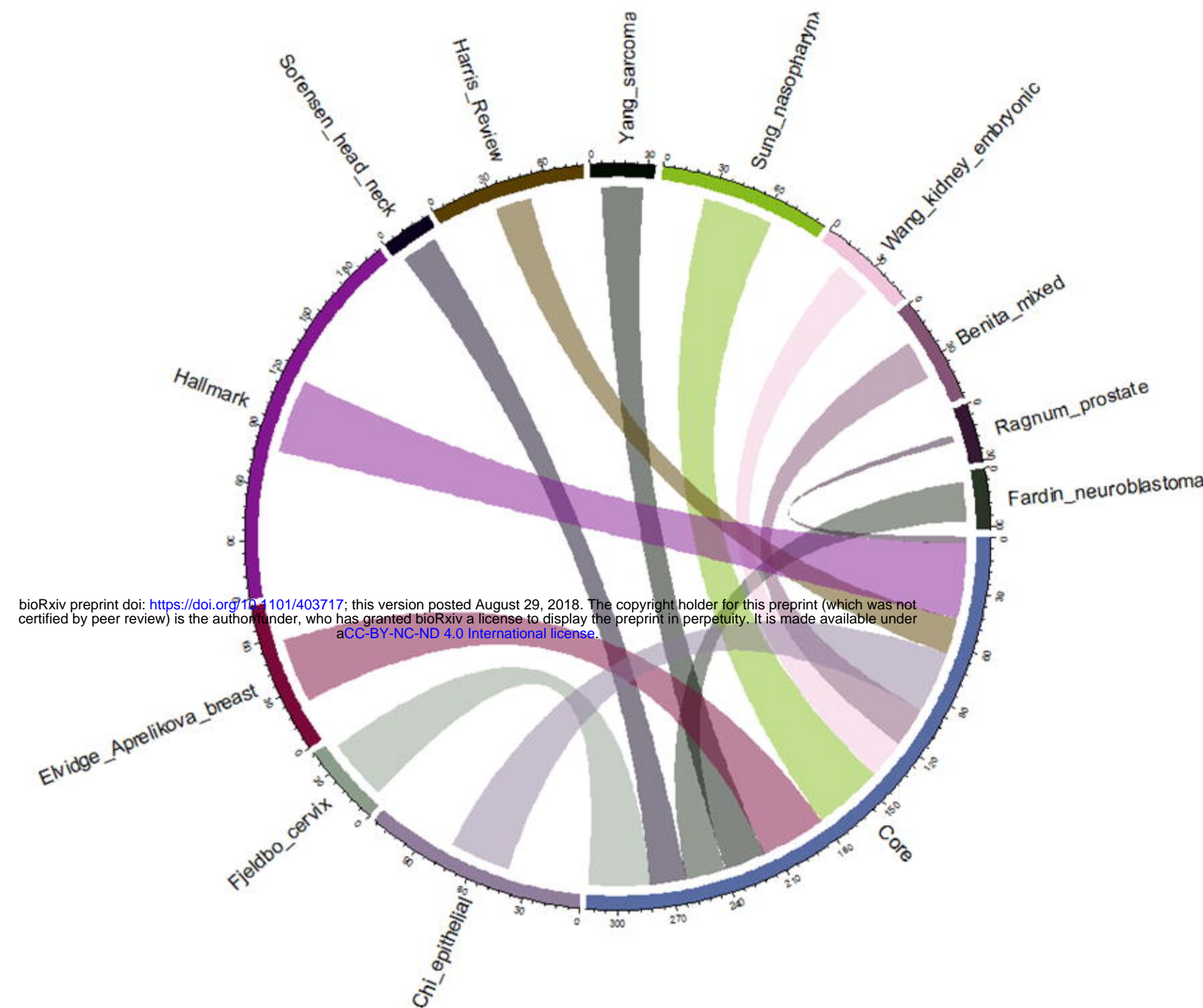
Figure 2. Pan cancer gene set enrichment analysis of the high-hypoxia phenotype. Genes were ranked according to their strength of association with the high-hypoxia phenotype in a pan cancer analysis of 32 cancer types available in TCGA. The high-hypoxia phenotype was defined from expression of 52 core hypoxia inducible genes. GSEA analysis was performed testing the enrichment of 50 hallmark pathways in high-hypoxia tumours. Some of the most enriched pathways, besides hypoxia, glycolysis and angiogenesis, are shown.

Figure 3. Associations between hypoxia phenotype, tumour stage and other cancer hallmarks in 32 cancer types. **A)** Associations between hypoxia and tumour stages were evaluated using χ^2 test. **B)** Associations between hypoxia and tumour stages were analysed in multiple regression models adjusting the confounding factors. **C)** and **D)** Associations between hypoxia and mutation rate were analysed using both t test and multiple regression models adjusting the confounding factors. **E)** and

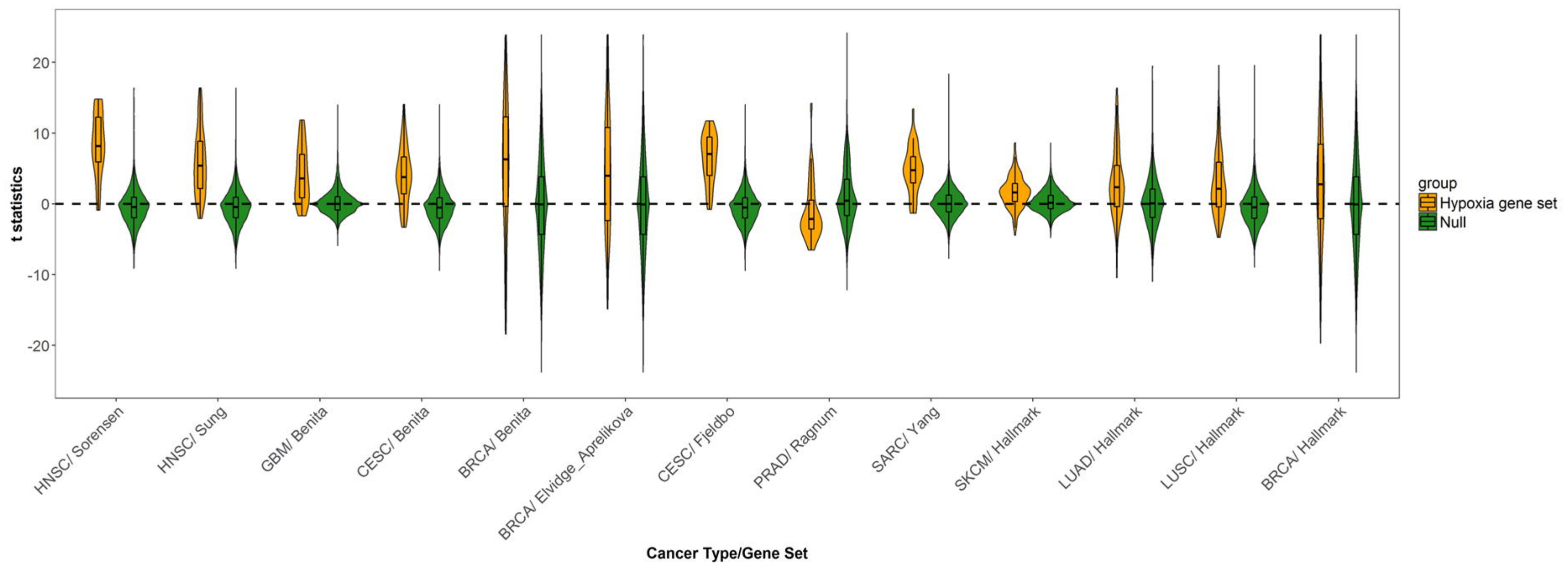
F) Associations between hypoxia and homologous recombination deficiency scores were analysed with both *t* test and multiple regression analyses. **G)** and **H)** Associations between hypoxia and leukocyte were analysed in both *t* test and multiple regression analyses. In each plot cancer types were ranked according to the effect size. Numerical values in each plot indicate FDR values.

Figure 4. Prognostic value of hypoxia phenotype in 32 cancer types. **A)** Prognostic significances of the 52 core hypoxia gene expression phenotype (high vs. low) alone were evaluated in each cancer type using Cox proportional hazard model. The analysis was performed for overall survival, progression free survival and disease specific survival separately. Dots indicate coefficients from Cox model and lines indicate 95% confidence intervals. In each endpoint *P* values were corrected into FDR (values on the side). **B)** The hypoxia gene expression phenotype was adjusted with mutation load, homologous recombination deficiency score, aneuploidy score, leukocyte infiltrate and tumour stage (using stage 1 as reference). **C)** The same multivariable analyses done in **B)**. For each endpoint and each biomarker, *P* values from 32 Cox models were corrected into FDR. Y axis is $-\log(\text{FDR}, 10)$, with higher values indicating more significance. The dashed horizontal line indicates a FDR of 0.1.

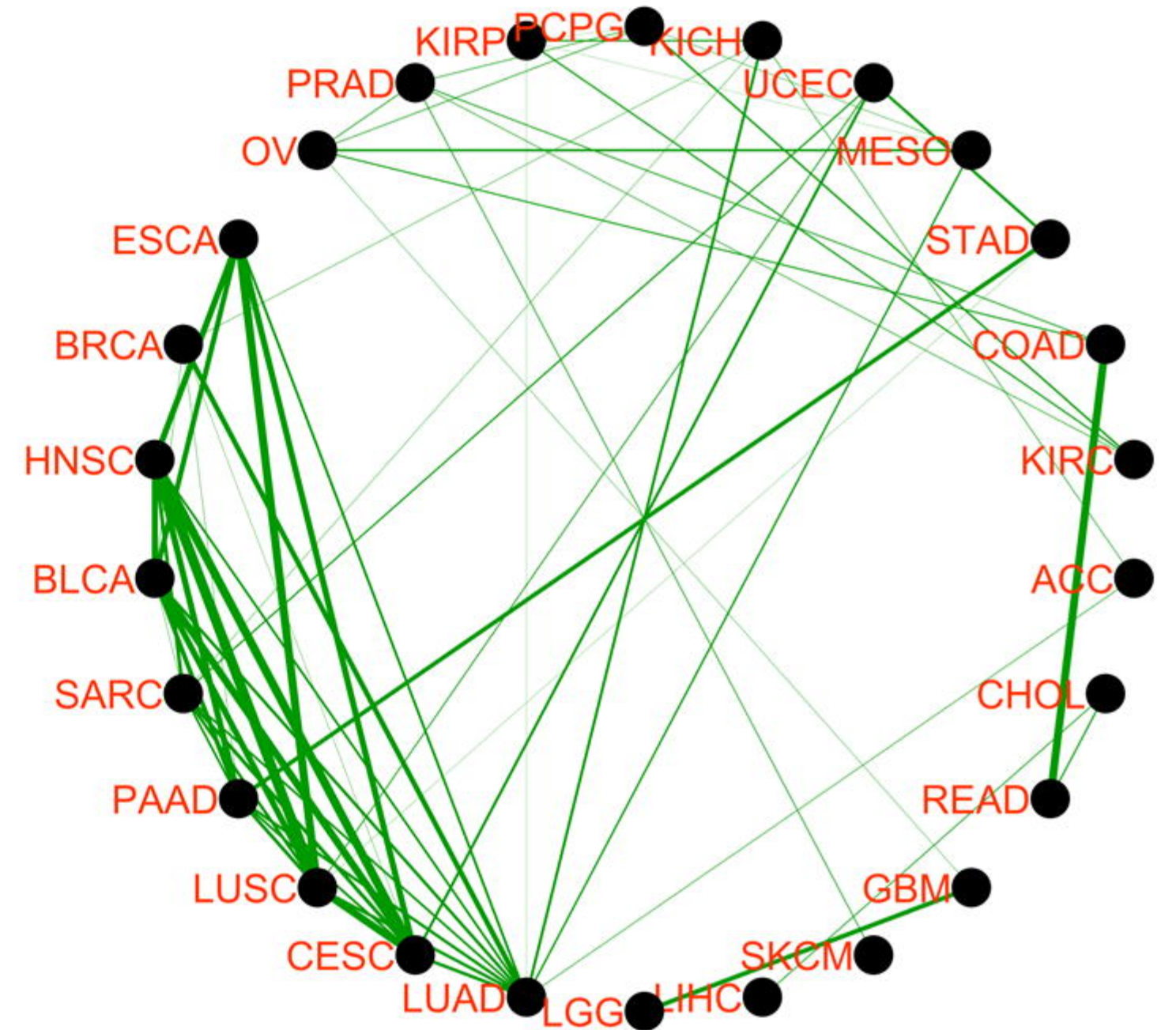
A

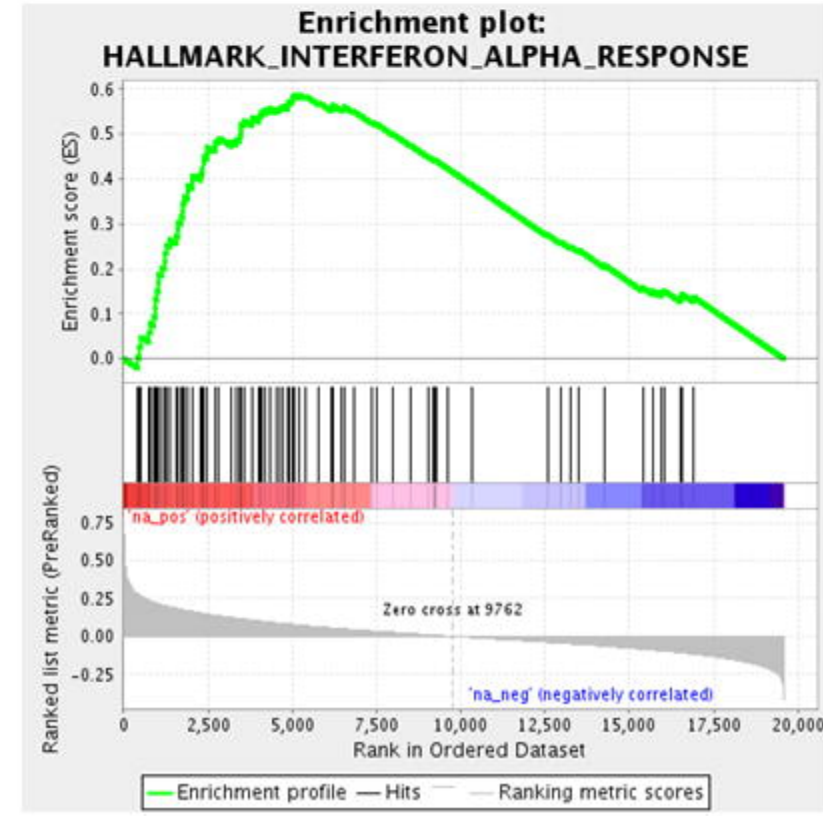
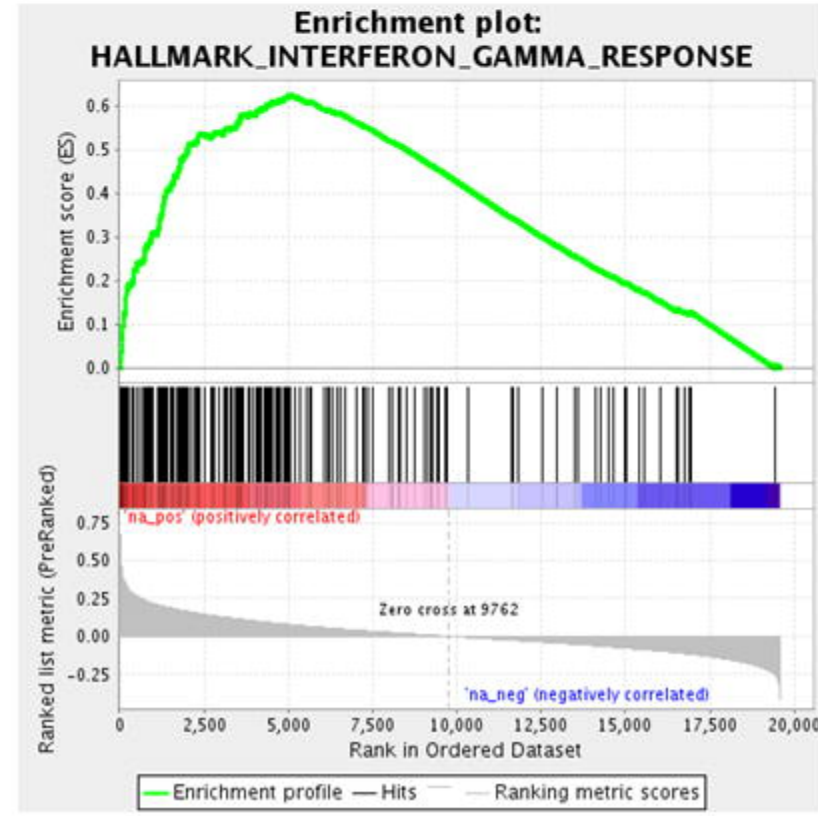
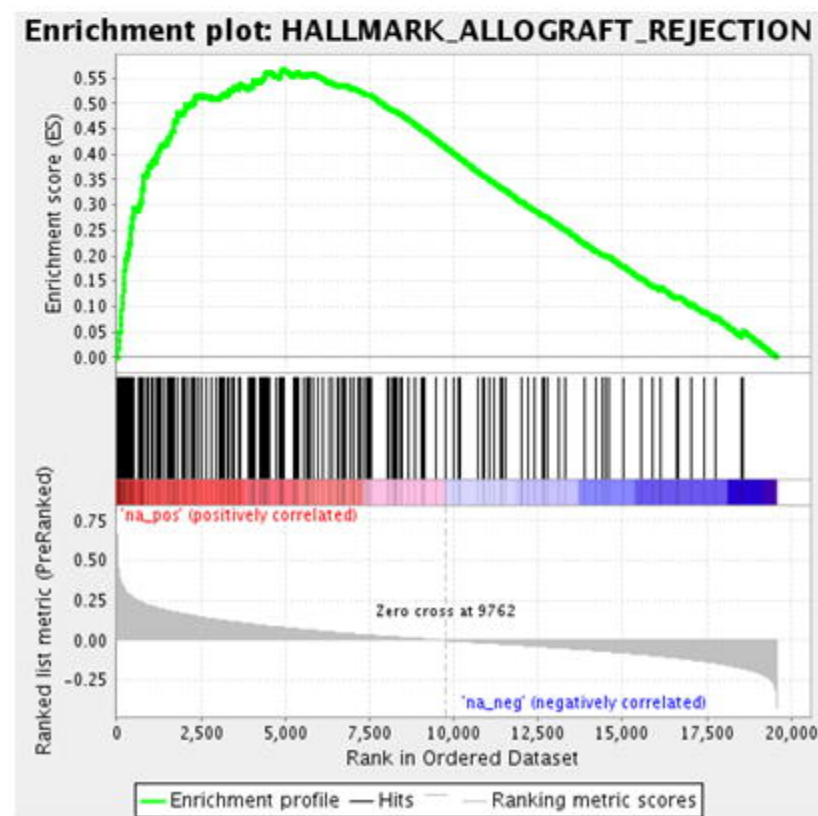
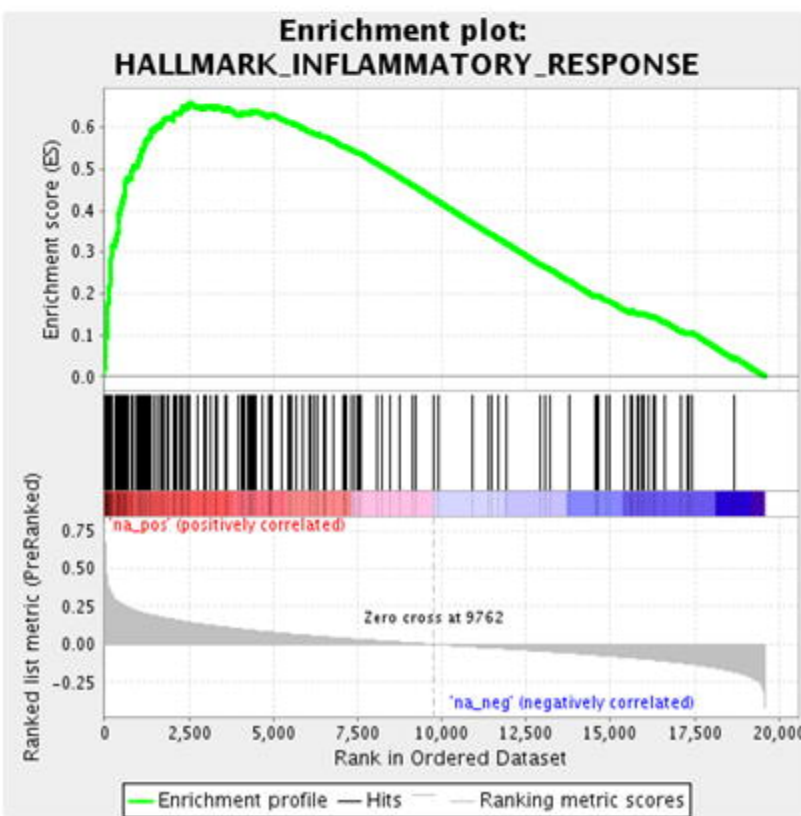
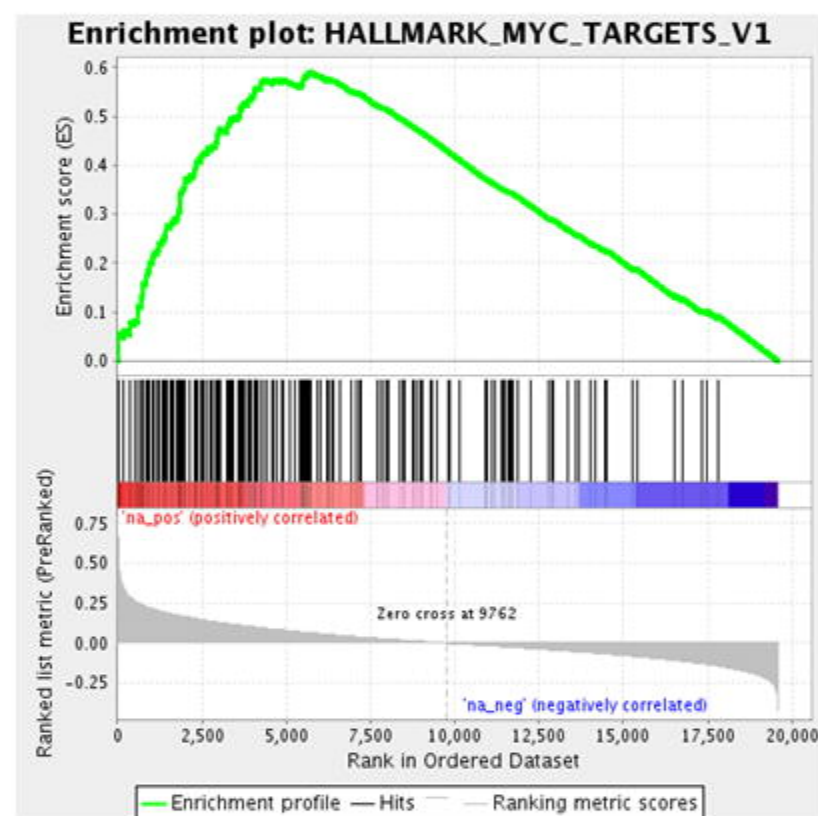
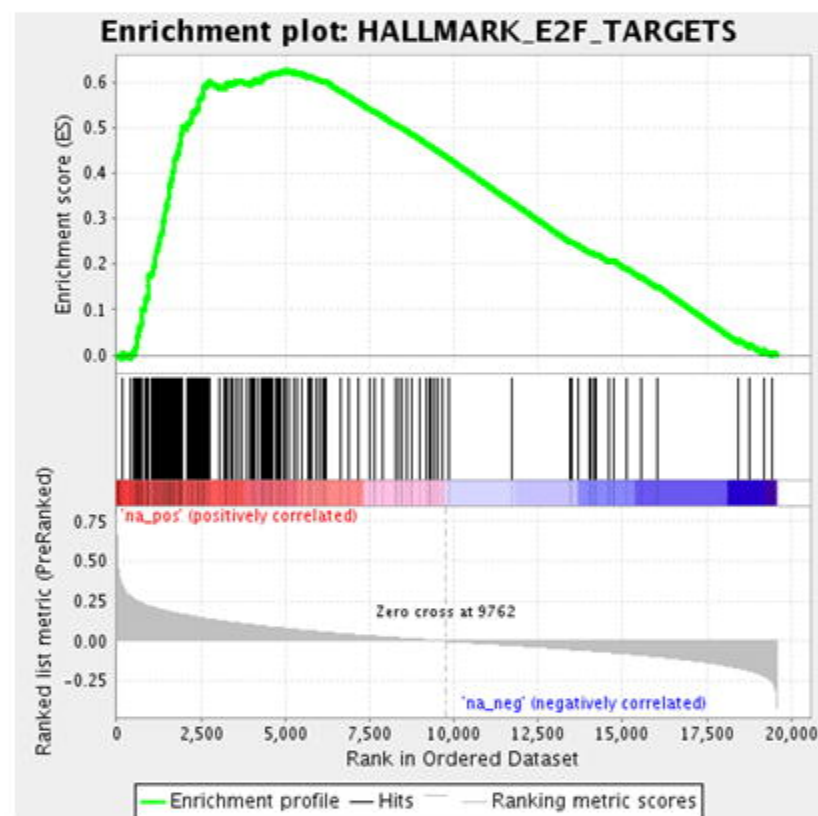
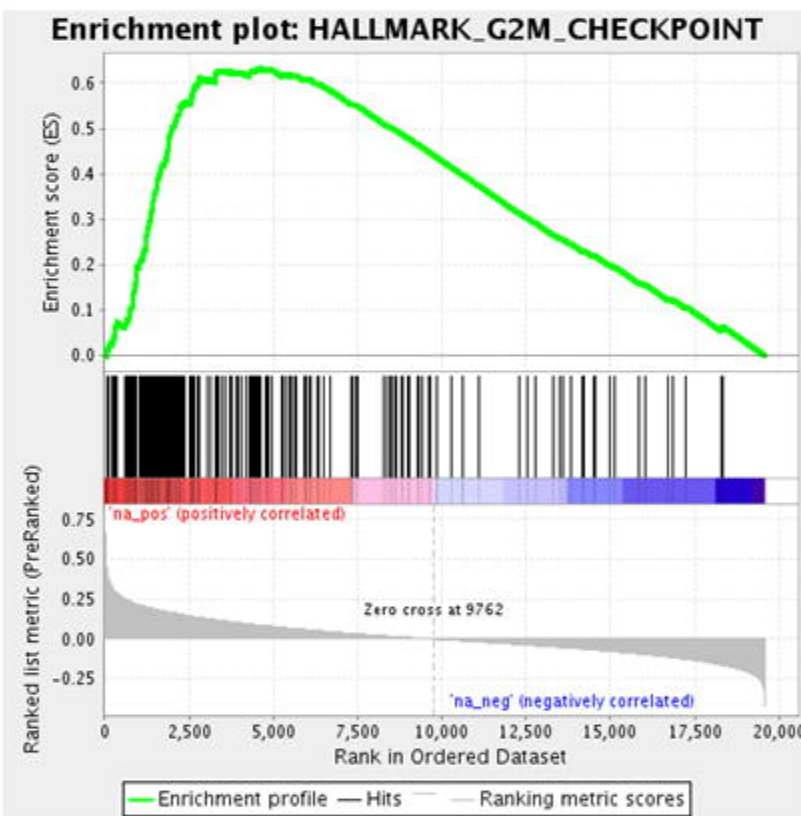
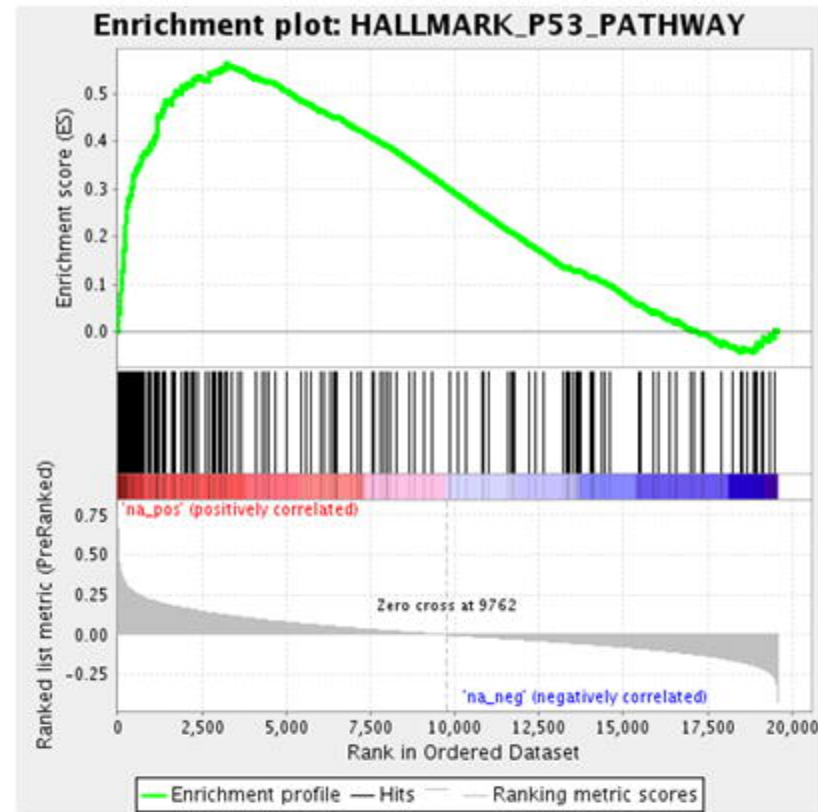
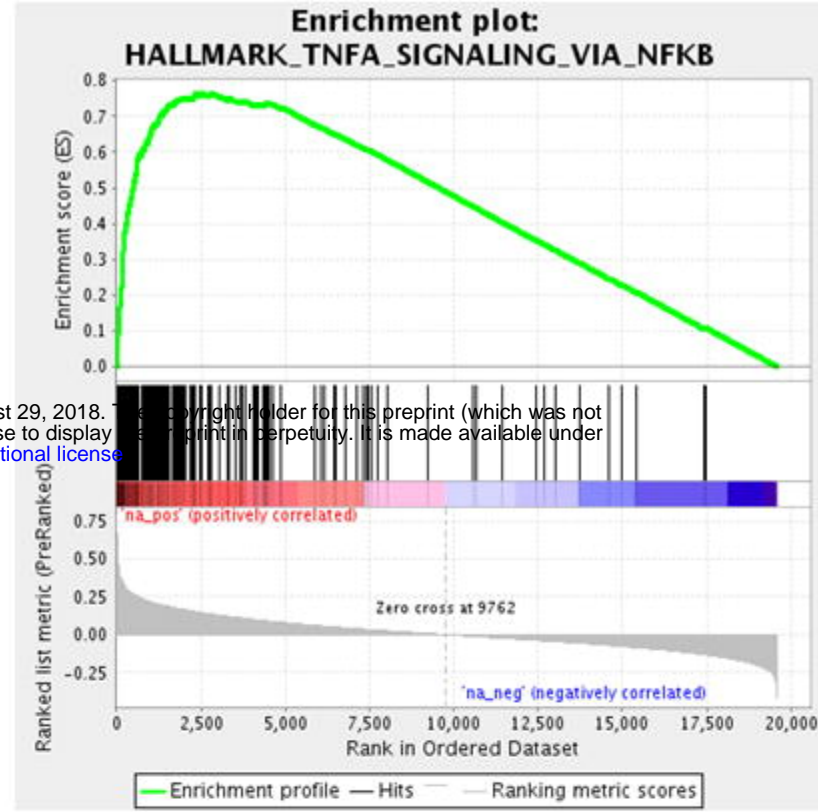
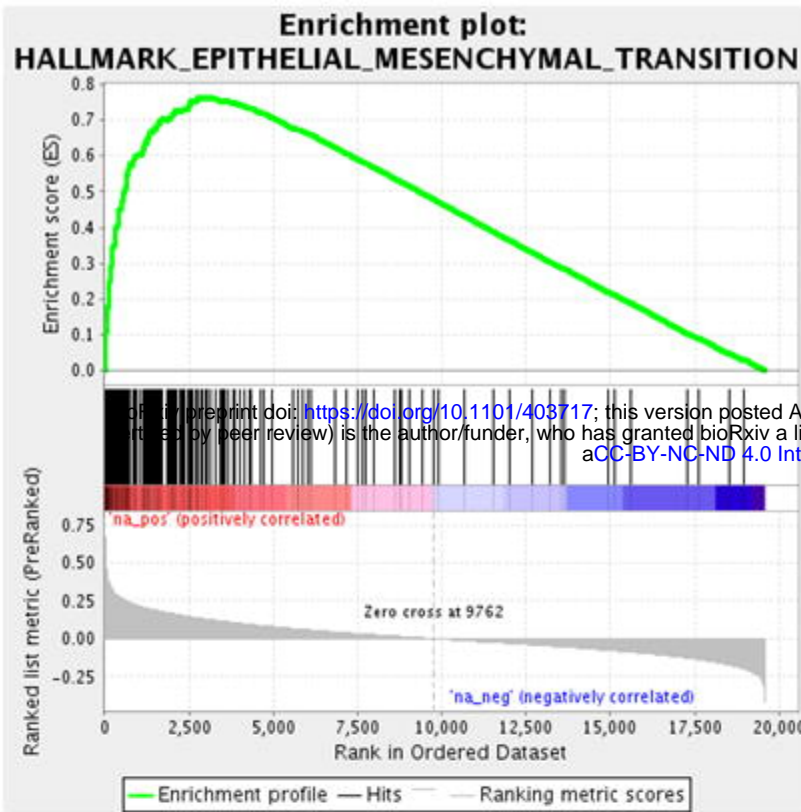


B

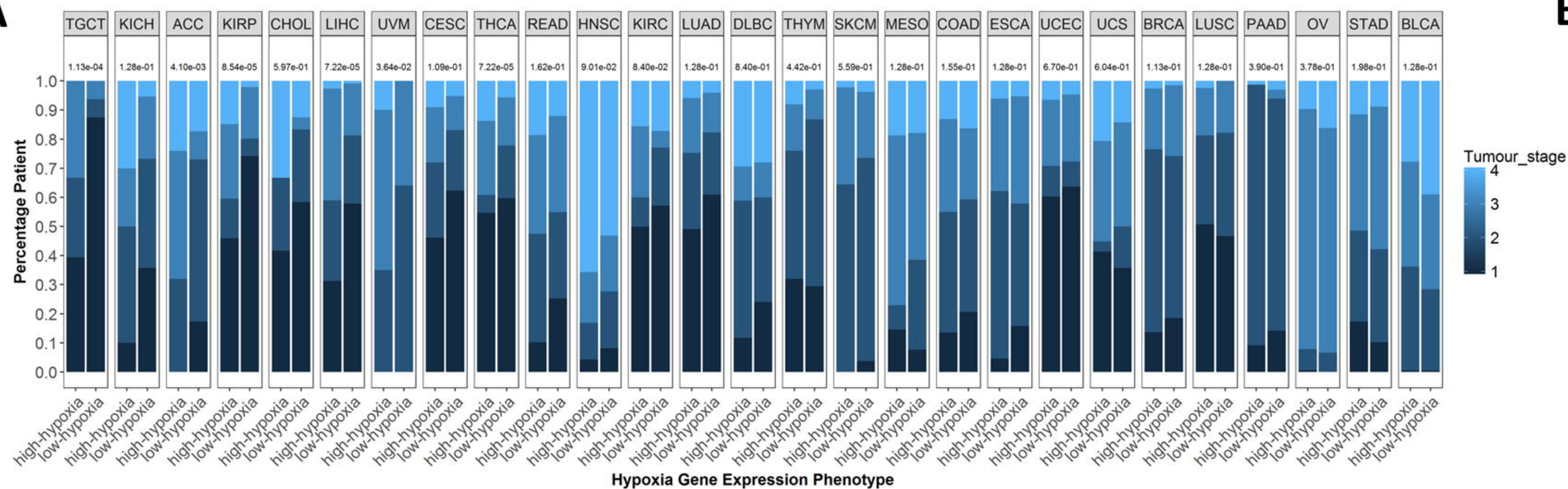


C

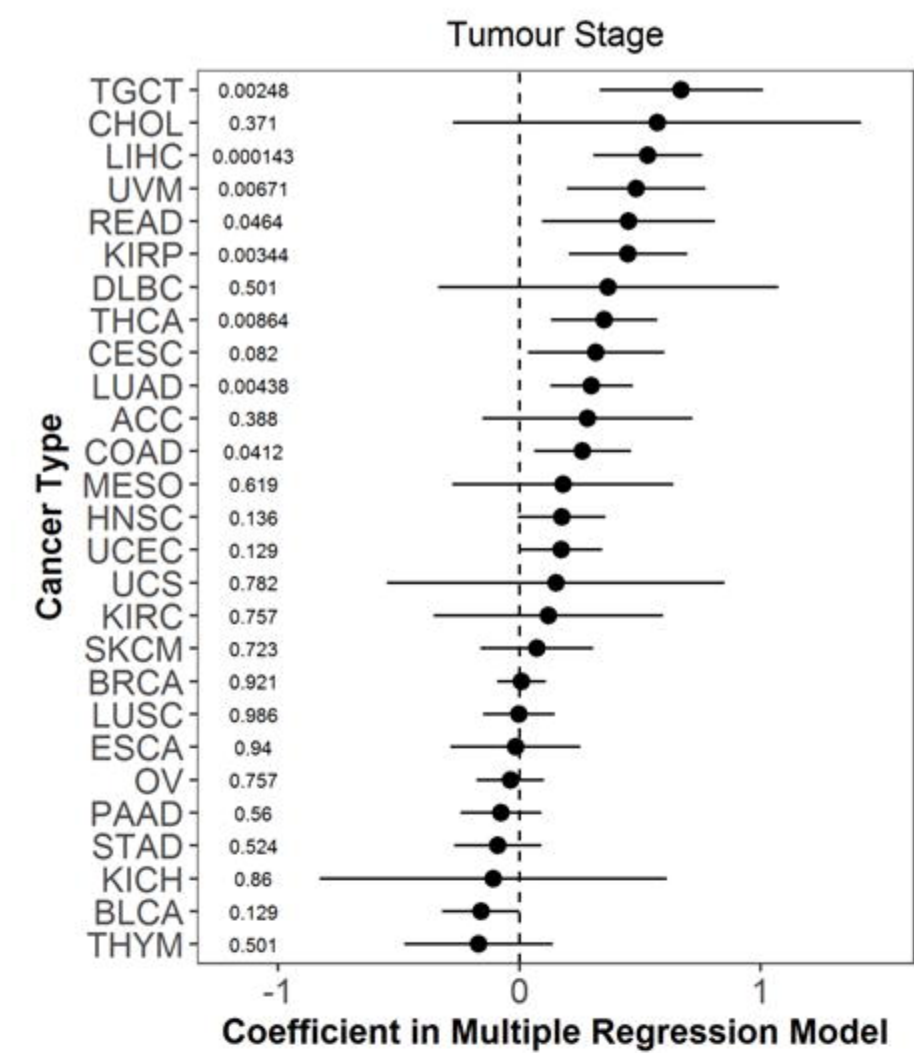




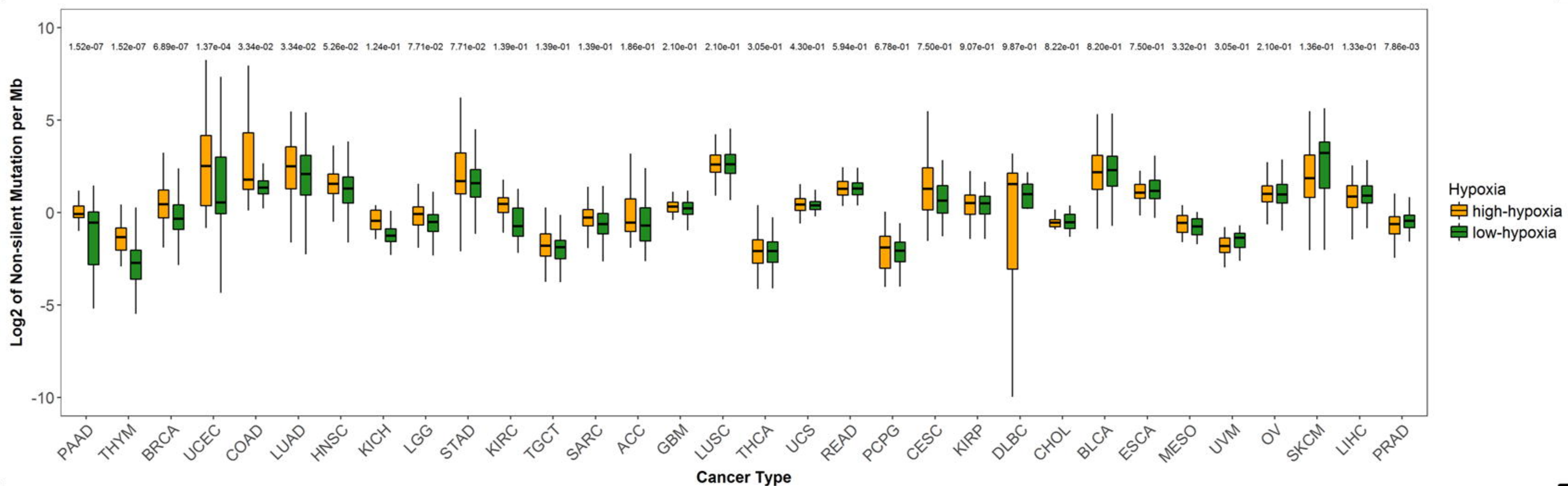
A



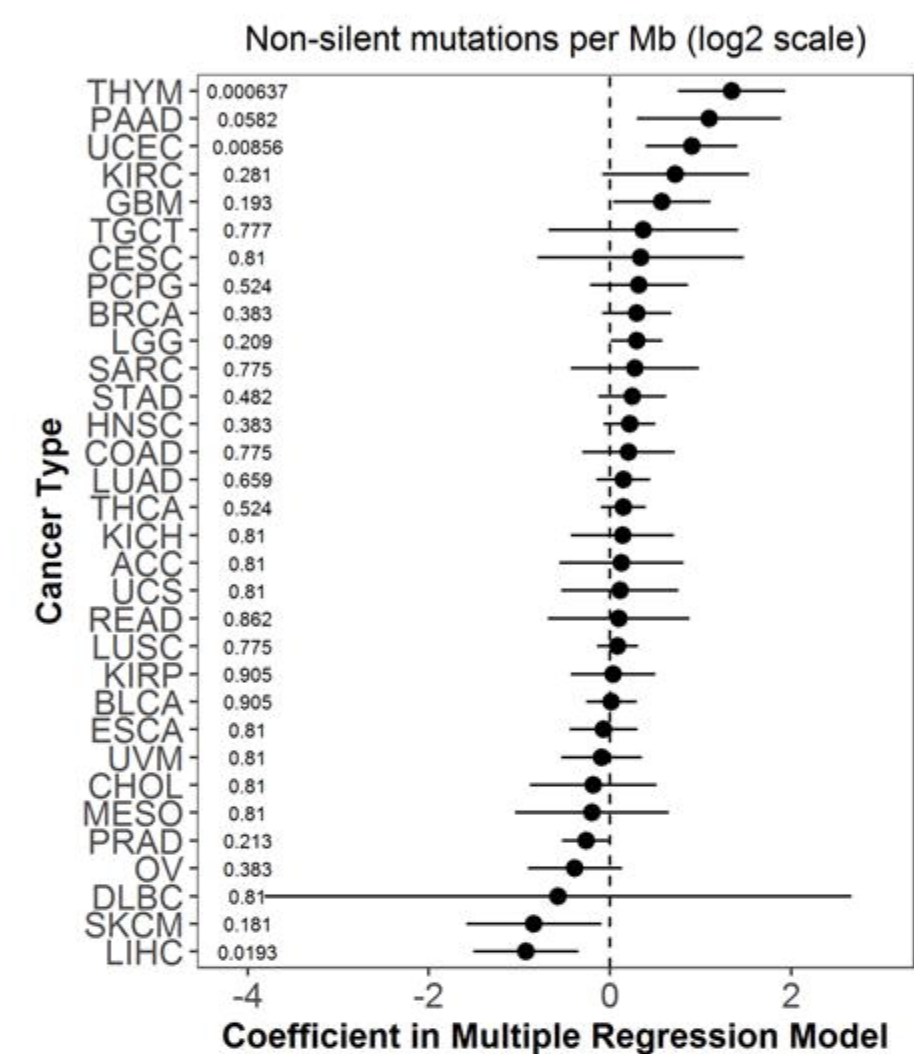
B



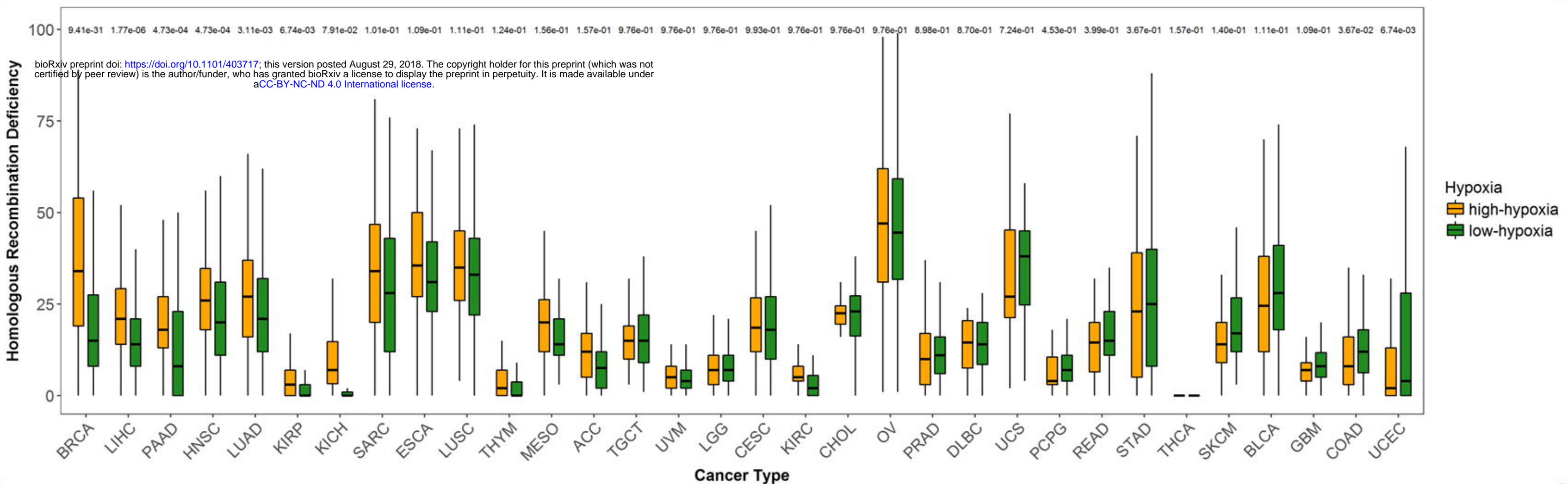
C



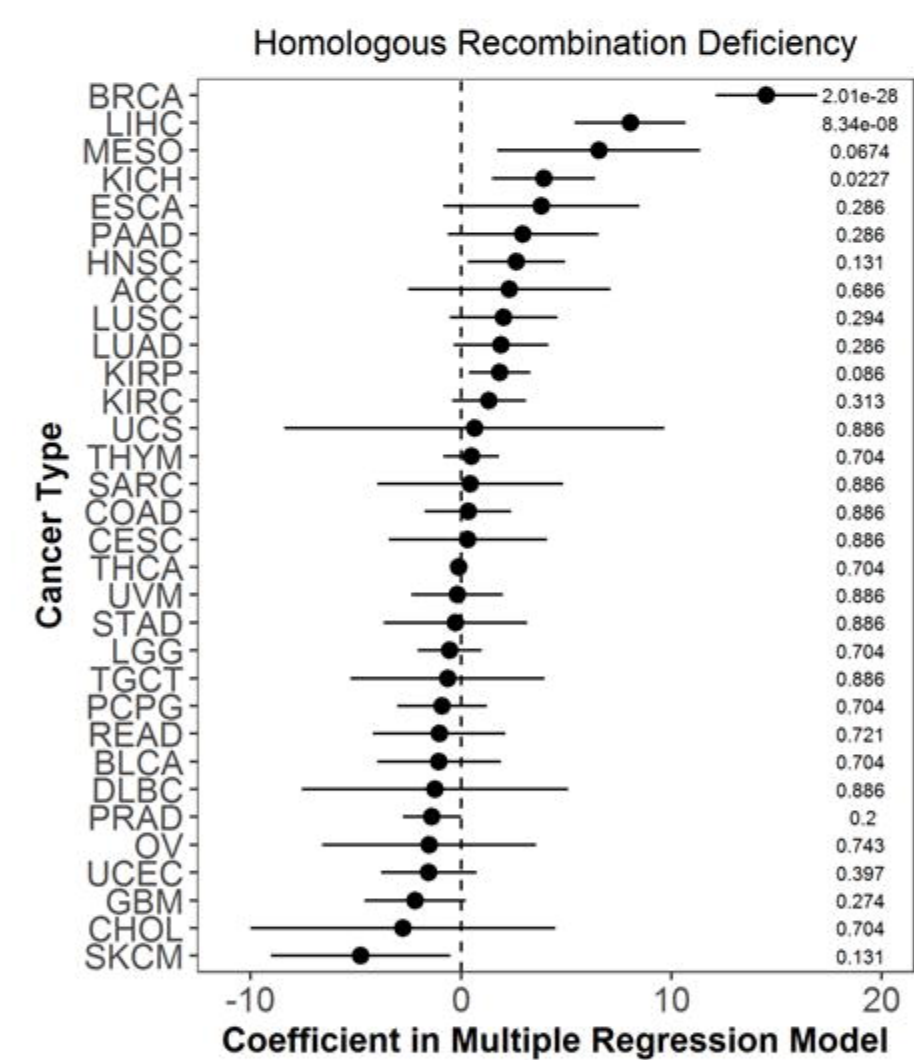
D



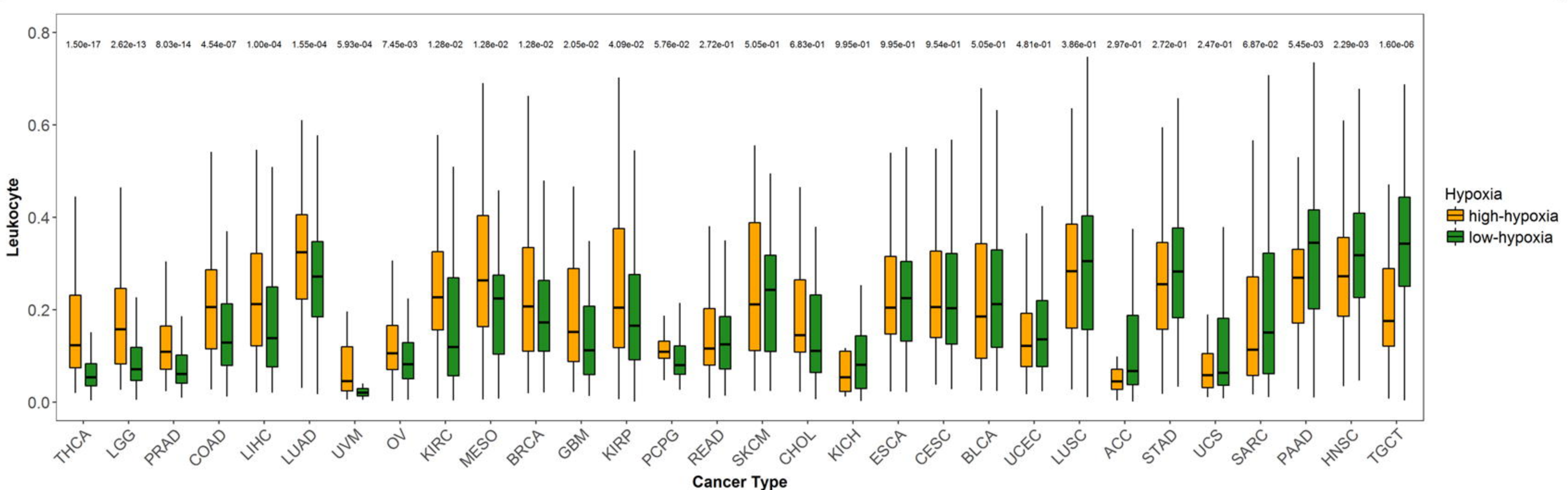
E



F



G



H

