Assessing hypoxia in cancer: a data-driven approach

Matteo Di Giovannantonio, Linacre College

Primary Supervisor: Professor Francesca Buffa

Other Supervisors: Professors Eric O'Neill & Benjamin Harris

Department of Oncology, University of Oxford

SARS-CoV-2 Pandemic Impact Statement

During my DPhil, the SARS-CoV-2 pandemic disrupted all aspects of our lives. My research and therefore this thesis were not exempt. This influenced the generation of benchside data that would have fed into my thesis, therefore I focussed mainly on data generated outside of the pandemic from publicly available sources.

Statement of originality

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own. This is also reflected in the authorship of the related submitted papers.

In all research, collaboration is a key part. Therefore I must sincerely thank those who have helped me throughout my DPhil and in my development as a scientist (see acknowledgments).

Submitted and published works related to my DPhil

- **Di Giovannantonio M.**, Harris B.H.L., Elshenawy B., Harris D.A., Barberis A., Harris A.L, Buffa F.M. Defining hypoxia through big data: a landmark study evaluating hypoxia gene expression signatures *Nature Genetics* [submitted]

  Role: study design, data cleaning, bioinformatics analysis, paper drafting, iteration and submission. This paper is the major output of my DPhil thesis and encompasses elements of all chapters.

- **Di Giovannantonio M.***, Harris B.H.L.* , Zhang P., Kitchen-Smith I., Xiong L., Sahgal N., Stracquadanio G., Wallace M., Blagden S., Lord S., Harris D.A., Harris A.L, Buffa F.M., Bond G. Heritable genetic variants in key cancer genes link cancer risk with anthropometric traits *Journal of Medical Genetics* 2021 58(6): 392-399.

  Role: study design, data cleaning, bioinformatics analysis and paper drafting.

- Harris B.H.L.*, Zuhair M.*, **Di Giovannantonio M.**, Rosadas C., Khan M., Short C., Thaventhiran T., Quinlan R., Taylor A, Calvez R., Taylor G.P, Tedder R.S, McClure M.O, Fertleman M. Asymptomatic Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection in a Rehabilitation Facility: Evolution of the Presence of Nasopharyngeal SARS-CoV-2 and Serological Antibody Responses
  The Journal of Infectious Diseases 2021 223(2): 192-196.

  Role: statistical analysis and paper drafting.

- Harris B.H.L.*, Pereira C.*, **Di Giovannantonio M.**, Rosadas C., Short C., Quinlan R., Sureda-Vives M., Fernandez N., Day-Weber I., Khan M., Marchesin F., Katsanovskaja K., Parker E., Taylor G.P, Tedder R.S, McClure M.O, Dani M., Fertleman M. The Association Between Antibody Response to Severe Acute

Respiratory Syndrome Coronavirus 2 Infection and Post-COVID-19 Syndrome in Healthcare Workers. *The Journal of Infectious Diseases* 2021 223(10): 1671-1676.

Role: statistical analysis and paper drafting.

- Pereira C, Perera AH, Rudarakanchana N, Rudarakanchana N, Harris BHL, **Di Giovannantonio M**, Taylor-Robinson SD, Dani M, Fertleman M. Cytokine Changes in Cerebrospinal Fluid Following Vascular Surgery on the Thoracic Aorta. *Scientific reports*, 2022.

  Role: statistical analysis and paper drafting.

- Fertleman M., Pereira C., Dani M., Harris B.H.L., **Di Giovannantonio M.**, Taylor-Robinson S.D. Cytokine changes in cerebrospinal fluid and plasma after emergency orthopaedic surgery. *Scientific Reports*, 2022.

  Role: statistical analysis and paper drafting.

- **Di Giovannantonio M.*** ,Harris B.H.L.* , Zhang P., Harris D.A, Karpe F., Lord S., Allen N.E., Maughan T.S., Bryant R. J., Harris A.L., Bond G., Buffa F.M. Integrated phenotypic and genetic analyses highlight the important role of fat-free mass in cancer risk. *Journal of Medical genetics* [in review]

  Role: study design, data cleaning, bioinformatics analysis and paper drafting.

List of Abbreviations/Acronyms

CNS	Central Nervous System
CPH	Cox Proportional Hazard
DMOG	Dimethyloxalylglycine, N-(Methoxyoxoacetyl)-glycine methyl ester
GEO	Gene Expression Omnibus
GLM	Generalized Linear Models
GSVA	Gene Set Variation Analysis
Hyp	Hypoxia
KM	Kaplan-Meier
NAT	Normal Adjacent Tissue
Norm	Normoxia
NSCLC	Non-Small Cell Lung Cancer
PC1	Principal Component 1

PLAGE	Pathway-level activity of gene expression
RGS	Random Gene Set
ssGSEA	single-sample Gene Set Enrichment Analysis
SPI	Signature Performance Index
TAMs	Tumour Infiltrating Macrophages
TCGA	The Cancer Genome Atlas

Index

Abstract

Low levels of cellular oxygen, also known as hypoxia, is a major characteristic of solid tumours. Cancer hypoxia is associated with poor prognosis, resulting in prometastatic and angiogenic effects as well as inducing resistance to both chemo- and radiotherapeutic treatments. Despite the critical role that low oxygen tensions play in cancer progression, the clinical use of hypoxia-targeting treatments is still extremely low due to the absence of both accurate and cost-effective methods available to measure hypoxia.

One promising technique to measure hypoxia are gene expression signatures. As opposed to other methods, usually more expensive or invasive, gene expression signatures can be used to measure hypoxia from cancer biopsies, and even in retrospective cohorts. During the last twenty years, more than fifty hypoxia gene expression signatures have been developed. These expression signatures were derived using a variety of computational methods and different experimental conditions. For example, some signatures have been derived *in vitro*, others *in vivo* only and others using both approaches including one or more cancer types. As a result, the scientific community has to decide which of the plethora of signatures to use in their experiments and clinical trials, but which signature and which way of summarising a signature into a “score” is most appropriate for which scenario is currently unknown.

This work is the largest and most comprehensive analysis and validation of the 53 published hypoxia gene expression signatures to date. Hypoxia gene expression scores were calculated on publicly available gene expression data from ~1,000 cell line samples (the Gene Expression Omnibus) and ~6,000 clinical samples (The Cancer Genome Atlas) and compared to the same score on normoxic samples. These hypoxia signature scores were compared to the scores derived from more than 7.5 million Random Gene Signatures (RGS) to determine whether they were truly able to measure hypoxia.

The overall most effective signature and score combination on cell line samples was the Sorensen 2010 signature using the median score. This achieved an impressive 92.84% accuracy in identifying hypoxic samples in 98 cell lines. In clinical samples, the Buffa 2010 signature using the mean score appears most appropriate as it fulfils the three key criteria of a) being on average higher in tumour samples than in normal tissue samples, b) differing in performance compared to random gene signatures c) serving as a strong prognostic marker (using a number of different thresholds) across 10 important tumour types. However, large prospective clinical studies with multiple measures of hypoxia are urgently needed to affirm our recommendation. This work lays down a new benchmark in how to measure hypoxia and the novel method used may transform how gene expression signatures in multiple fields might be evaluated in years to come.

Chapter 1: Gene expression signatures as markers of hypoxia and the concept of synthetic lethality

1.1 Introduction

The lifetime incidence of cancer is currently one in two[1]. This increasing burden of cancer on society necessitates new therapeutic interventions. The three pillars of cancer treatment: surgery, radiotherapy and targeted therapies, need optimisation to improve patient survival. By far the least optimised are targeted therapies, perhaps due to the biological heterogeneity and evolutionary nature of malignancies. However, over the past 20 years, there have been some notable successes, for instance BRAF inhibitors and then checkpoint inhibitors transforming patient survival in melanoma[2,3,4]. A major driver of treatment resistance, and hence a promising area for targeted therapy is low oxygen tension (hypoxia) which is an extremely common characteristic of the tumour microenvironment.

1.2 Hypoxia in cancer

Hypoxia is associated with poor prognosis in many different tumour types, including breast[5], bladder[6], gastric[7], head and neck[8], liver[9], lung[10], oesophageal[11] and prostate[12]. Indeed, several factors combine to induce hypoxia in tumour cells, particularly a mismatch between oxygen supply and demand. Proliferating cancer cells provide an increasing demand for oxygen, where aberrant angiogenesis cannot keep pace. Further, the neovasculature associated with tumours tends to be poorly organised and not placed optimally to provide oxygen to the tumour cells: oxygen can only diffuse about 160 mm in tissue[13]. This low oxygen environment places a selective pressure upon cells as they are forced to either adapt or die.

Humans have an inbuilt oxygen sensing mechanism, the hypoxia-inducible factor (HIF) system[14], which is a major effector of metabolic changes in hypoxia. HIF is an αβ-heterodimer that was originally identified as a DNA-binding factor that mediates hypoxia-inducible activity of the erythropoietin 3′ enhancer[15]. Later it was identified in a wide variety of cells suggesting the HIF system had a broader role than simply in erythropoietin synthesis[16]. It is important to note that HIF-α and HIF-β subunits have different isoforms which can have different functions and abundances, including HIF-1α, HIF-2α, HIF-3α and HIF-1β. Importantly, alpha subunits (HIF-α) rise in low oxygen tensions. HIF-1α and HIF-2α appear closely related and able to interact with hypoxia response genetic elements (HREs) to induce transcription. HIF-3α is less well understood, but appears to be involved in negative regulation of the hypoxic response through an alternatively spliced transcript[17]. In contrast, beta subunits (HIF-β) interact with DNA and are constitutively-expressed proteins classified as aryl hydrocarbon receptor nuclear translocators (ARNT). Only the dimeric form, containing both α and β subunits, is active as a transcription factor.

The HIF system is responsible for oxygen sensing in mammalian cells. HIF-1α is its cytoplasmic component and is normally maintained at a low level, its production being balanced by destruction under the influence of the Von Hippel-Lindau (VHL) protein, a ubiquitin ligase. When destruction is prevented, HIF-1α builds up in the cytoplasm, moves to the nucleus and, in combination with HIF-1β, triggers the expression of a set of genes associated with hypoxia (see below).

It is the stability of cytoplasmic HIF-1α that is dependent on oxygen. When normal levels of oxygen are present, a cytoplasmic prolyl hydroxylase will hydroxylate HIF-1α and the hydroxylated form is subject to ubiquitinylation/destruction. Levels of HIF-1α are thus kept low. In hypoxia, proline hydroxylation, which requires oxygen as a substrate, is prevented and HIF-1α destruction does not occur. At low oxygen tensions, therefore, HIF-1α can move to the nucleus to generate a transcription factor HIF-1α/HIF-1β (Fig 1.1), whose target is the 'hypoxia-responsive element', found upstream of a number of genes, with the consensus sequence 5′-RCGTG-3′ (R = purine).

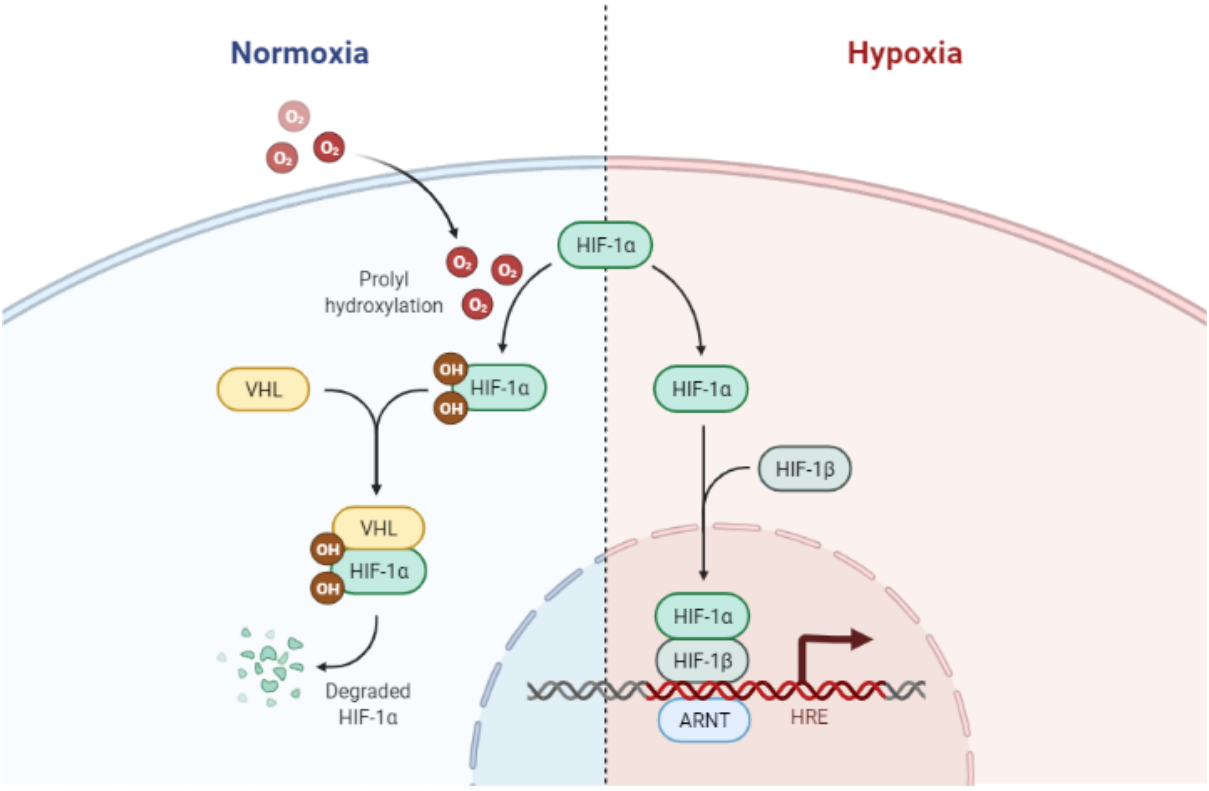


Figure 1.1: *The hypoxia inducible factor (HIF) system*

In normoxia, HIF-1α in the cytoplasm is hydroxylated on pro402 and pro564, leading to its ubiquitinylation by the VHL protein and subsequent destruction in the proteasome. In low oxygen tensions hydroxylation cannot occur, and HIF1α levels rise in the cytoplasm. HIF-1α moves into the nucleus where it combines with HIF-1β, forming a transcription factor, and stimulates transcription of hypoxia responsive genes via a hypoxia response genetic element (HRE). Figure from BioRender.

The stabilisation of HIF-1α has a number of downstream effects directly and indirectly on a variety of pathways, including glycolysis (e.g. LDHA), iron metabolism (e.g. TF) and angiogenesis (e.g. VEGF-A). Activation of the HIF system promotes tumour growth particularly by stimulating angiogenesis (VEGF genes are activated), enabling a good supply of nutrients to the growing tumour. HIF-1α stabilisation also changes the cellular milieu of non-coding RNAs that can influence gene expression, particularly microRNAs such as the oncomiRs miR-182[18] and miR-210[19]. Thus perhaps it is unsurprising that hypoxia can influence several of the classical and emerging Hallmarks of Cancer[20], including sustaining angiogenesis[21] and reprogramming energy metabolism[22]. In support of this model, defects in VHL promote tumour growth and VHL itself can be thought of as a tumour suppressor protein.

As well as changing the cells themselves, hypoxia influences the efficacy of treatment. One key area in which it plays a role is in external beam radiotherapy with photons. The majority of damage done to cancer cells using this modality is through indirect effects, with hydroxyl radicals from water needing oxygen in order for them to cause damage to DNA. Indeed, hypoxic cells are about three-fold more resistant to photon based ionising radiation. In systemic anti-cancer therapy, hypoxia too provides a challenge as if a tissue is hypoxic, it is suggestive that conventional drug delivery through blood supply will also be reduced. Therefore there have been efforts to combat hypoxia in a number of different ways ranging from reducing hypoxia in tumours to targeting hypoxic cells specifically. To address the low oxygen levels, red cells transfusions and erythroid stimulating agents to increase haemoglobin levels have been trialled[23,24] as well as oxygen mimetics (e.g. nimorazole[25]), carbogen (98% oxygen with 2% carbon dioxide)[26] and even hyperbaric oxygen[27]. As regards exploiting the hypoxic environment to target treatments, viral vectors[28], hypoxia-activated prodrugs[29] and drugs targeted at specific hypoxia-induced cellular pathways[30] have all been investigated. Recent work has highlighted promise in repurposing drugs from other fields to

actually decrease tumour oxygen consumption, thus minimising hypoxia. In particular, the anti-protozoal drug atovaquone, used to treat *Pneumocystis jiroveci* pneumonia and commonly in combination with proguanil hydrochloride in prophylaxis of malaria has perhaps unexpectedly shown promise. Atovaquone has been shown to significantly reduce oxygen consumption across a variety of tumour cell lines (FaDu, HCT116, DLD-1, H1299 lung carcinoma, A549, H460, MCF7, T24 and PSN-1) both in vitro and in mouse models[31]. The mechanism responsible for the decrease in oxygen consumption secondary to atovaquone is the inhibition of oxidative phosphorylation specifically at mitochondrial complex III of the electron transport chain. Subsequent work in humans (n=30) has confirmed that hypoxic areas in non-small cell lung cancer appear to reduce following administration of the well-tolerated atovaquone[32], paving the way for a phase III trial to assess its effect on patient outcomes. However, as yet, no hypoxia modifier or targeted therapy is routinely used in the UK and a multitude of trials have not shown clinical benefit in humans. A commonly cited reason for this has been the complexity surrounding stratification of patients, e.g. if patients are stratified to receive hypoxia modifying therapy and they have non-hypoxic tumours, they may not receive any clinical benefit. Therefore it is important to ascertain which tumours are hypoxic to change these promising benchside therapies into clinical staples.

Hypoxia can be measured in a number of ways, some more invasive than others but as yet none is used routinely in the clinic. The most direct and considered the “gold standard” measure of oxygen tension is an oxygen electrode. However its utility is limited by accessibility and also limited areas of the tumour can be sampled. Further, this approach does not differentiate between regions of hypoxia and areas of necrosis which may be less biologically relevant. Oxygen electrodes have been used in clinical studies but other approaches that could be easily integrated into the standard clinical pathway are looking more promising.

Other hypoxia measuring solutions can be split into cross-sectional imaging techniques, the more experimental area of circulating markers of hypoxia and a variety of methods that can be applied to tumour biopsy. A number of cross-sectional imaging approaches exist tending to use magnetic resonance imaging (MRI) or positron emission tomography (PET). Two notable examples using MRI include blood oxygenation level dependent (BOLD) MRI and oxygen-enhanced MRI (OE-MRI or tissue oxygenation level dependent (TOLD) MRI)[33]. BOLD MRI is based upon the intrinsic effective transverse relaxation rate whereas, TOLD MRI is based on T1-weighted contrast. Both have been widely debated, and perhaps are best employed in combination to overcome the drawbacks of both, e.g. TOLD MRI seems to relate more to radiation response whereas BOLD MRI seems to show larger size changes under hypoxia[34]. PET scans to measure hypoxia have tended to use Fluorine-18 (^{18}F) fluoromisonidazole [18F-MISO] or Copper-64 (^{64}Cu) diacetyl-bis(N4-methylthiosemicarbazone) [64Cu-ATSM] as tracers, with increased retained Fluorine-18 (^{18}F) or Copper-64 (^{64}Cu) implying low oxygen levels within the tumour. Good correlation between these two tracers has been described in rodent models of brain tumours ($r \sim 0.8$)[35]. Notably, these techniques rely on clinical facilities having suitable strength MRI machines and also tracer availability, but these techniques theoretically can be repeated over time.

Serological markers of hypoxia too have the advantage that they can be measured serially. The current frontrunner in this field is the hypoxia-induced tumour-associated glycoprotein osteopontin[36]. Elevated plasma levels have been shown to be associated with poor survival following curative intent radiotherapy in head and neck cancer[37] as well as in sarcoma patients[38]. Other proposed biomarkers of hypoxia are hepatocyte growth factor (HGF) and interleukin-8 (IL-8)[39] but in all cases, further evaluation is needed, particularly in intercorrelation with other validated hypoxic markers. One potentially salient limitation is that

the serological biomarker cannot differentiate the origin of hypoxic tissue in patients with multiple tumour sites (i.e. in metastasis) and this might be relevant in radiotherapy planning.

A more established benchside approach for measuring hypoxia tumours is using immunohistochemistry for targets downstream of HIF-1α. Commonly used protein targets include carbonic anhydrase-9 (CA9) and glucose transporter 1 (GLUT1)[40]. An advantage of this approach is that immunohistochemistry can be used to identify the intracellular location of protein (e.g. membrane, nuclear, cytoplasmic) and help in protein quantification. However this approach is somewhat subjective even in the most automated protocols, and variation in antibody performance may influence ultimate interpretation.

A very promising approach that can be applied to routine tumour biopsies relates to measuring RNA gene expression related to hypoxia. Although biopsies can be challenging or even impossible in some tumours, they are the gold standard before treatment is commenced. Further with the reduction in sequencing costs, measuring “gene expression signatures” as surrogates of hypoxia has become an attractive option. However, one must be mindful that assessing hypoxia by measuring cellular adaptation i.e. by genes or proteins can potentially be subject to error if attention is not paid to mutational status. For instance, mutations in the key HIF pathway player VHL should cause reduced breakdown of HIF-1α (see *Fig. 1.1*), and thus should make a cell appear “hypoxic” judging on gene/protein expression even in normoxic conditions. However, this assumption has not as yet been proven using a variety of gene expression signatures. And this issue is potentially easily circumvented by undertaking a parallel salient mutational screen as part of the RNASeq analysis needed to work out the hypoxia gene expression signature score.

Unlike the aforementioned strategies, these hypoxia gene expression signatures have the potential to stratify patients in retrospective data series. This opens up other data sources not specifically aimed at studying hypoxia as useful additions to hypoxia analyses;

one such resource could be the The Cancer Genome Atlas (TCGA) dataset. This allows advanced big data and deep learning approaches to be employed which otherwise might be less fruitful, e.g. artificial neural networks.

A number of hypoxia gene expression signatures have been developed using a variety of approaches[13]. A common approach has been to derive signatures *in vitro* using microarrays by comparing cells in normoxia and hypoxia. Focus was mainly placed on upregulated genes under hypoxia, and genes significantly upregulated or reaching a research determined threshold were termed as “hypoxia gene expression signature”. Such signatures have attracted some scrutiny when clinical translation was proposed due to the difference between cell lines and the human tumour microenvironment. Therefore clinical datasets started to become the focus when deriving new hypoxia signatures. Previous knowledge of hypoxia-regulated genes was used to identify coregulated genes using a variety of bioinformatic methods. A signature from our laboratory was pioneering in this field. Winter et al. 2007 developed a “hypoxia metagene” through clustering genes from *in vivo* human head and neck cancer samples (n = 59) with 10 well-known hypoxia-regulated genes (VEGF, ADM, GLUT1, PDK-1, ENO1, HK2, PFKB3, CA9, AK3, and CCNG2)[41]. The median RNA expression for all the genes in the metagene (signature of 99 genes) was found to be a prognostic factor for recurrence-free survival in an independent cohort of head and neck cancer patients (n = 60). Further, the median score for this signature was also found to be a marker of metastasis-free and overall survival in a human breast cancer dataset (n = 295), suggesting such signatures of hypoxia could be used in different cancer types aside from the ones in which they are derived.

This approach was then refined and expanded to create a common, compact and highly prognostic hypoxia metagene. Buffa and colleagues in 2010 used previously validated “seed genes” which were known to be hypoxia responsive. A larger number of *in vivo* cancer

samples from three head and neck and five breast cancer studies (combined n = 1,136) were used to identify co-expression patterns which were then used to define a hypoxia signature or “metagene”. The hypoxia metagene (51 genes) was then seen to be prognostic in independent breast and lung cancer datasets.

Subsequently, comprehensive work from our group then went on to mine the literature for hypoxia signatures to identify commonalities and potential differences: Harris et al. identified 32 published hypoxia signatures in 2015 (*Table 1.1*), 27 derived from cell lines and five originating from human samples. Further, Harris et al. highlighted that the size of the signatures varied markedly from four to 759 genes and clinically derived signatures were found to have less genes on average than their in vitro derived counterparts (61 vs 85 genes). This data was recovered by a systematic interrogation of Pubmed, Scopus, Web of Science, hypoxiadb and genesigdb using the search terms ‘hypoxia signature’ or ‘hypoxia’ AND/+ ‘signature’. They also highlighted the important issue of gene reannotation of the generated signatures, potentially influencing their utility. For instance, the understanding of genes and proteins change as our biological understanding and resolution evolves, for instance VEGF was discovered initially, but subsequently VEGF-B, VEGF-C and VEGF-D were discovered and the “original” VEGF was renamed VEGF-A. Further, it is now acknowledged that there are different splicing isoforms of various transcripts, leading to another level of complexity. Since this publication RNASeq has become a more common method of RNA quantification but similar issues still remain as authors still tend to report gene symbols rather than transcript ids. It was also highlighted that both *in vitro* and clinically derived signatures have both been shown to be prognostic and predictive signatures in certain settings[42,43], however more essential work is needed on evaluating the signatures in other datasets. In addition, there is minimal discussion on the most appropriate way to apply these signatures in terms of summary statistics (e.g. median vs mean etc) or their utility across multiple datasets and cancers. Discussing with the authors [personal

communication], this was felt to be an area of interest but out of the scope of this particular manuscript as this would have formed more of a research focus rather than a review. Therefore, although slightly dated, this review lays a solid foundation for this work, leaving the important open research questions of where these hypoxia signatures can be used and how best to apply them. Indeed, the true versatility of signatures in denoting hypoxia across cell lines and different experimental conditions and thus clinical samples is not well described and needs revisiting as few signatures have been evaluated over large cohorts or over a number of different cancer types. Also, perhaps surprisingly, there is no consensus on how best to employ these signatures in terms of their summary statistics, to define hypoxic tumours within the wealths of publicly available genomic data (e.g. does Principal Component Analysis 1[44] (PCA1) perform better than Gene Set Variation Analysis[45] (GSVA) or vice versa). Even the recent extensive study published by et Bhandari et al. did not address this[46] despite looking for the molecular hallmarks of tumour hypoxia across the TCGA dataset. Here, the authors used several signatures and showed the utility of gene expression signatures in classifying tumour hypoxia based on their molecular profile, such as the mutational landscape of hypoxia in prostate cancer or hypoxia and tumour subclonal evolution. This study showed the clinical relevance of hypoxia signatures and their ability in quantifying hypoxia across different cancer types. However, a systematic comparison of different signatures and scores was still missing and as Yang et al point out there is *no consensus on the ideal method to define tumors as hypoxic and no methodological study to date assessing and comparing the performance of the different methods*[47] and I hope to address this.

Pubmed ID	*Author & publication date*	*Number of gene symbols*	*Reference*
10706099	Koong et al, 2000	10	[48]
12947397	Denko et al, 2003	80	[49]
15093745	Jogi, et al, 2004	107	[50]
15100389	Ning et al, 2004	104	[51]
15374877	Manalo et al, 2005	107	[52]
15833863	Wang et al, 2005	56	[53]
15994966	Detwiller et al, 2005	27	[54]
16849508	Bosco et al, 2006	177	[55]
16507782	Mense et al, 2006	111	[56]
16740701	Aprelikova et al, 2006	236	[57]
16417408	Chi et al, 2006	111	[58]
16565084	Elvidge et al, 2006	181	[59]
16595741	Peters et al, 2006	159	[60]
17532074	Seigneuric et al, 2007 (early 0% O_2)	71	[61]
17532074	Seigneuric et al 2007, (early 2% O_2)	34	[61]

17532074	Seigneuric et al, 2007 (common genes between early 0% and 2% O_2 signatures)	14	[61]
17409455	**Winter et al, 2007**	**99**	[41]
17187782	Shi et al, 2007	32	[62]
17320280	Sung et al, 2007	90	[63]
18984585	Beyer et al, 2008	159	[64]
20592013	Van Malenstein et al, 2009*	7 (3*)	[65]
19491311	Benita et al, 2009*	81 (57*)	[66]
19832978	Fardin et al, 2009	8	[67]
19291283	**Hu et al, 2009**	**13**	[68]
20652058	Fardin et al, 2010	35	[69]
20429727	Sorensen et al, 2010	27	[70]
20087356	**Buffa et al, 2010**	**51**	[71]
20416888	**Ghorbel et al, 2010**	**166**	[72]
21846821	Toustrup et al, 2011	15	[42]
21325071	**Ghazoui et al, 2011**	**70**	[73]
22356756	Starmans et al, 2012	759	[74]
23820108	**Eustace et al, 2013**	**26**	[43]

Table 1.1: *Hypoxia gene expression signatures found by Harris et al. 2015*

Emboldened signatures are those which used approaches involving analysis of clinical samples whereas starred (*) signatures contain both up- and downregulated genes.

1.3 Aims

Bearing the aforementioned in mind, the aims of my DPhil are as follows:

- *Overarching:* complete the most comprehensive analysis of hypoxia gene expression signatures to date
    - identify all published hypoxia gene expression signatures in the literature
    - determine the best summary scores to be used with hypoxia signatures
    - determine which hypoxia signatures and scores best identify cells under hypoxia in retrospective *in vitro* datasets
    - explore the performance of the hypoxia signatures and their scores in a large clinical dataset spanning common cancer types (TCGA)

Chapter 2: Materials and methods

Assessment of hypoxia signatures requires several approaches. This section presents the different datasets and methods that have been used for this project, including some statistical concepts to better understand the power of the analyses as well as the critical points and future developments. Code related to this work can be found in *Appendix 6*.

2.1 Hypoxia gene expression signatures

A gene expression signature consists of a collection of one or more genes whose expression levels reflect a specific biological status of the sample under examination. These can be derived in a number of different ways, including training mathematical classification models such as Generalized Linear Models (GLM) to identify the genes that are able to better discriminate between two different classes (e.g. Hypoxia vs Normoxia). Harris et al.[13] defined two main methods for the identification of hypoxia gene expression signatures: (i) the "traditional" *in vitro* approach, where genes have been selected according to their expression in cell lines, and (ii) an alternative approach in which a combination of cell lines data and clinical samples have been used for generating the signatures.

Published hypoxia gene expression signatures were identified from two main sources (summarised in *Fig. 2.1*):

1. A previously published extensive literature review from our group[13] containing 32 signatures
2. Interrogating Web of Science, Scopus and Pubmed databases using the query: *("hypoxia signature"*) OR (*"hypoxia"* AND *"signature"*)

This approach undercovered signatures that had been developed on specific cancer types and in different oxygen tensions. For example, the Ragnum signature was developed

studying the gene expression patterns in prostate cancer cell lines whereas the Elvidge signature only used MCF7 breast cancer cell lines. Another interesting signature is the Seigneuric score which was obtained under different conditions of hypoxia (anoxia and 2% oxygen).

Pubmed ID	Author and publication date	Number of gene symbols
10706099	Koong *et al.*, 2000	10
12947397	Denko *et al.*, 2003	80
15093745	Jogi *et al.*, 2004	107
15100389	Ning *et al.*, 2004	104
15374877	Manalo *et al.*, 2005	107
15833863	Wang *et al.*, 2005	56
15994966	Detwiller *et al.*, 2005	27
16849508	Bosco *et al.*, 2006	177
16507782	Mense *et al.*, 2006	111
16740701	Aprelikova *et al.*, 2006	236
16417408	Chi *et al.*, 2006	111
16565084	Elvidge *et al.*, 2006	181
16595741	Peters *et al.*, 2006	159
17532074	Seigneuric *et al.*, 2007 (early 0% O_2)	71
17532074	Seigneuric *et al.* 2007, (early 2% O_2)	34
17532074	Seigneuric *et al.*, 2007 (common genes between early 0% and 2% O_2 signatures)	14
17409455	**Winter *et al.*, 2007**	**99**
17187782	Shi *et al.*, 2007	32
17320280	Sung *et al.*, 2007	90
18984585	Beyer *et al.*, 2008	159
20592013	Van Malenstein *et al.*, 2010*	7 (3*)
19491311	Benita *et al.*, 2009*	81 (57*)
19832978	Fardin *et al.*, 2009	8
19291283	**Hu *et al.*, 2009**	**13**
20652058	Fardin *et al.*, 2010	35
20429727	Sorensen *et al.*, 2010	27
20087356	**Buffa *et al.*, 2010**	**51**
20416888	Ghorbel *et al.*, 2010	166
21846821	Toustrup *et al.*, 2011	15
21325071	**Ghazoui *et al.*, 2011**	**70**
22356756	Starmans *et al.*, 2012	759
23820108	**Eustace *et al.*, 2013**	**26**

+

Figure 2.1: *Identification of gene expression hypoxia signatures*

This approach identified 53 signatures by combining a) the list from Harris et al. published in 2015 (the table on the left is adapted from the publication) with b) the results of a search on Web of Science, Scopus and Pubmed to identify the most recent signatures. The review collected 32 signatures (those highlighted in bold have been derived using clinical samples). The other 21 signatures were identified using the query on the three comprehensive online resources.

2.2 Reannotation of hypoxia signatures

Gene expression signatures have been developed using a variety of sequencing technologies. Over the last decade, before the advent of Next Generation Sequencing (NGS) at reasonable cost, microarrays were widely used to derive gene expression signatures. However, the most recent signatures tend to use RNA sequencing data or other gene expression technologies (e.g. Illumina Beadchip) for deriving hypoxia signatures. The authors of the signature related papers mainly report the genes in their publications as Gene Name (or HGNC symbol) and rarely include more specific elements such as microarray Probe ID or the Ensembl/Entrez Gene ID. This provides a challenge as biological understanding evolves over time as does gene nomenclature. A commonly used way to uniform gene annotations across signatures involves converting all the gene names from the articles into more stable gene identifiers, such as Ensembl gene IDs, using Ensembl biomaRt[75]. However, this conversion might be quite problematic for two reasons:

1. **Gene names change over time**: for example, the Vascular Endothelial Growth Factor A (VEGFA) gene, which plays a key role in hypoxia, was previously reported as VEGF. Thus any signatures that contain VEGF, will lose this gene during the conversion process using Ensembl BiomaRt.
2. **Ensembl annotation database change over time**: a new release of the Ensembl dataset could result in loss of information about Probe IDs or gene names if these are outdated, since Ensembl does not include outdated annotations in its relational database. For example, the above-mentioned VEGF gene name is not available on the recent Ensembl release.

Therefore, in order to convert gene names from papers into Ensembl gene IDs using biomaRt, it is critical to reannotate gene names into their most updated version. For this task, the multi-symbol checker from HGNC was used[76]. *Fig.2.2* shows an example output of the multi-symbol checker. The “Approved symbol” column reports the most recent gene

symbol from the input gene symbol. Each gene expression signature was reannotated using the multi-symbol checker to convert each gene symbol to the latest “Approved symbol”. In case more Alias symbols were suggested, the Relevance Score on GeneCards[77] was manually used to select the most relevant annotation. Gene names annotated as protein coding were prioritised.

Once updated to the most recent and relevant gene symbol, genes in the signature were converted into Ensembl Gene IDs using biomaRt (Ensembl 107: Jul 2022). Finally, a conversion dictionary was created to match each Ensembl Gene ID with any other external annotation such as Probe IDs or Ensembl transcript IDs. A list of all gene symbols and Ensembl Gene IDs is available in *Appendix 1*.

Input	Match type	Approved symbol	Approved name	HGNC ID	Location
CART	Alias symbol	CARTPT	CART prepropeptide	HGNC:24323	5q13.2
CPO	Approved symbol	CPO	carboxypeptidase O	HGNC:21011	2q33.3
CPO	Previous symbol	CPOX	coproporphyrinogen oxidase	HGNC:2321	3q11.2
HK2	Approved symbol	HK2	hexokinase 2	HGNC:4923	2p12
HK2	Alias symbol	KCNA5	potassium voltage-gated channel subfamily A member 5	HGNC:6224	12p13.32
HK2	Alias symbol	KIF2A	kinesin family member 2A	HGNC:6318	5q12.1
HK2	Alias symbol	HOOK2	hook microtubule tethering protein 2	HGNC:19885	19p13.13
SCL2A1	Unmatched				

Figure 2.2: *Multi-symbol checker from HGNC*

Output from the HGNC multi-symbol checker using the following hypoxia-associated genes as input: CART, CPO, HK2 and SCL2A1. For HK2 with multiple aliases, the HK2 symbol was kept.

2.3 Calculation of hypoxia scores

Hypoxia gene expression signatures can be reduced to a single measure, known as a hypoxia score. Hypoxia scores are single values indicating the level of hypoxia in biological samples and can be calculated using different techniques[78]. Six main scores are commonly used in the literature: Mean score, Median score, PCA1 score, ssGSEA, GSVA and PLAGE. Each of these scores are derived using different methods:

1. **Mean** score: mean value of the expression of the genes in the signature.
2. **Median** score: median value of the expression of the genes in the signature.
3. **PCA1** score: this score is obtained by selecting the first element in the Principal Component Analysis (PCA1)[44].
4. **ssGSEA** score: a single-sample extension of Gene Set Enrichment Analysis (GSEA). The ssGSEA enrichment score represents the degree to which the genes in a particular gene set are coordinately up- or down-regulated within a sample[79].
5. **GSVA** score: the Gene Set Variation Analysis (GSVA) score is a method designed to assess the collective behaviour of functionally related genes forming a set[45].
6. **PLAGE** score: the Pathway-level activity of gene expression (PLAGE) operates by quantifying the level of 'activity' of each pathway in different samples[80].

Each score gives a different summary of the data and has potential advantages and drawbacks. For example, the mean score is more likely to be swayed by outliers compared to the median. However, due to methodological restrictions only Mean, Median, GSVA and ssGSEA could be evaluated (discussed further in Chapter 2.5). In order to reduce bias, all 53 signatures were examined using these summary scoring methods. This was done to ascertain the most appropriate combination of score and signature for both cell lines and clinical samples. The scores were calculated using the sigQC package[81] using R v4.2.0.

2.4 Datasets used

Two main data repositories have been used in this study: *(i)* the Gene Expression Omnibus (GEO) archive to determine the performance of hypoxia signatures *in vitro* on both cancer and non-cancer cell lines and *(ii)* The Cancer Genome Atlas (TCGA) for evaluating the effectiveness of hypoxia signatures in clinical samples.

2.4.1 The Gene Expression Omnibus

The Gene Expression Omnibus (GEO) archive[82] is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data submitted by the research community. The GEO archive genomic data are organised in ***Series***, identified by the letters “GSE” followed by the series number (e.g. GSE153291). Series link together a group of related samples and provide a focal point and description of the whole study, including information about extracted data, processing, analyses and conclusions. Each Series contains information about:

- ***Platforms***: A platform record is composed of a summary description of the array or sequencer. A Platform ID is reported with the letters “GPL” and the platform number. For example, GPL570 is the corresponding ID for the Affymetrix Human Genome U133 Plus 2.0 Array. One platform can be assigned to multiple Series.
- ***Samples***: a Sample record describes all the conditions under which an individual Sample was handled. Each Sample is identified with a unique ID, the letters “GSM” followed by the sample number (e.g. GSM71498).

Batch effects can occur due to different technical and experimental conditions in addition to the sequencing technology used, such as the lab users and the consumables utilised. Combining all samples from different Series together in a unique dataset could

introduce batch effects that could lead to inaccurate conclusions. To overcome this potential problem, I assessed the performance of hypoxia signatures by evaluating the difference between paired hypoxic/normoxic samples according to the following rules:

1. Both samples must be part of the same Series and sequenced using the same technology (same Platform ID, e.g. both sequenced using GPL570 from Series GSE153291)
2. Both samples are from the same cell line (for example, MDA-MB-231 samples cannot be combined with MCF-7 samples)

To date, GEO stores more than *175,000* Series, *23,200* Platforms and *550,000* Samples. Unfortunately, to ease the uploading process for researchers, the GEO archive collects most of the information as *unstructured* data (e.g. free text). Thus, all the information such as oxygen concentration, time under hypoxia, cell type or additional treatments need manual curation. In this study, the GEO database was queried to identify cell line experiments testing hypoxic conditions at different time points and oxygen concentrations. Other exogenous conditions in addition to hypoxia such as exposure to a low glucose environment and experiments involving hypoxia mimicking agents were also identified. The following query was used on the Advanced Search engine on GEO:

```
("cancer"[All Fields] AND "hypoxia"[All Fields]) AND
"human"[Organism] AND ("gse"[Entry Type]) AND ("expression profiling
by array"[DataSet Type] OR "expression profiling by high throughput
sequencing"[DataSet Type] OR "expression profiling by rt
pcr"[DataSet Type])
```

To summarise, the following criteria were included in the query:

- Series must contain the word "*Hypoxia"* and "*Cancer"* in any of the fields in the Series description

- Experiments must have been performed using human-derived biological samples
- Gene expression profiles were obtained using high-throughput sequencing or arrays

The query identified 204 Series and a total number of 2,134 Samples. Three sequential filtering rules were then applied:

1. ***Rule #1***: The first filtering was performed on each Series by looking at its experimental description. Series containing at least one *in vitro* hypoxic sample and one normoxic control were selected leaving 103 Series and 1511 samples after the filtering
2. ***Rule #2***: The second filtering consisted of removing all Series that have not been published or referenced in any peer-reviewed article, leaving 97 Series and 1423 samples after this step
3. ***Rule #3***: The third filtering was performed by manual curation of all the Series and their corresponding papers, retrieving experimental information such as cell lines, oxygen tensions used, etc. Series were excluded if any samples in the Series reported different annotation between the Series information in GEO and the information in the scientific article.

A total number of 88 Series and 973 samples that survived the final filtering step are discussed more in Chapter 3 and are investigated in detail in Chapters 4 and 5. Both RNA-seq and microarray gene expression data were used in their post-processed form, as used in their published reference papers. A further normalisation step was included for RNA-seq data, converting each gene expression value into Transcript Per Million (TPM).

2.4.2 The Cancer Genome Atlas (TCGA)

The Cancer Genome Atlas (TCGA) is a “landmark cancer genomics program”, coordinated by the National Cancer Institute and National Human Genome Research (USA)[83]. The project includes profiling of 20,000 primary cancer samples from 33 cancer types with in some cases matched normal samples. Alongside this there is a wealth of detailed clinical information including, e.g. stage, histological type etc. the TCGA data has been made publicly available in an anonymised fashion for the scientific community. The analyses presented in this work focus on mRNA expression in the TCGA samples in 10 key solid tumours where at least 30 normal tissue samples were available to evaluate against to investigate the performance of hypoxia signatures (discussed further in Chapter 2.6).

2.5 Evaluating hypoxia signatures and scores in cell lines

The performance of hypoxia signatures was evaluated against Random Gene Sets (RGS) of the same length as the original hypoxia signature to ensure they performed significantly better than random sets of genes. This is particularly important as previous work has highlighted the interesting finding that unrelated and random gene signatures have been linked to clinical outcomes likely by chance[84]. Thus, this study seeks signatures and summary scores for hypoxia that outperform RGS using a permutation-based analysis. Further, each Series was tested separately to avoid experimental biases and batch effects, i.e. samples from GSE15530 were not compared to samples from GSE3188, even if belonging to the same cell line as discussed in Chapter 2.4.1.

First and foremost, it is important to define how the hypoxia scores can be evaluated in controlled cell line experiments. These scores are assessed by looking at their value according to the oxygen status of the samples (hypoxic or normoxic). For example, a hypoxic sample would be expected to have a higher hypoxia score than a normoxic sample. Which score and signature is chosen may influence this as illustrated in *Fig. 2.3*. Here, the ssGSEA score using the Buffa signature for the hypoxic sample is lower than the ssGSEA score for the normoxic sample (*Fig. 2.3a)*. This suggests the scoring system might be suboptimal compared to the mean score where the score for the hypoxic sample is greater than that for the normoxic sample (*Fig. 2.3b)*.

One way of measuring how a signature differentiates between hypoxic/normoxic pairs consists of calculating the absolute value between their two hypoxia scores (Euclidean distance) (*Fig. 2.4*). Measuring this Euclidean distance lays the foundation of the innovative method presented in this study and the evaluation against RGS. Given a hypoxia signature and a score method, the Euclidean distance $d_{SIG}(hyp, norm)$ is calculated as follows, with h_{score} and n_{score} representing the score of the hypoxic and normoxic samples respectively:

$$d_{SIG}(hyp, norm) = \sqrt{(h_{score} - n_{score})^2}$$

If a hypoxia signature and score combination is working well, you may expect a large distance, $d_{SIG}(hyp, norm)$. The same approach also is carried out using an RGS and this too gives a Euclidean distance, defined as $d_{RGS}(hyp, norm)$. One may expect the $d_{RGS}(hyp, norm)$ to be less than the hypoxia signature $d_{SIG}(hyp, norm)$ if the hypoxia signature is working correctly. This effect is demonstrated in *Fig. 2.3* using the mean score for both the Buffa signature (*Fig. 2.3c*) and one RGS (*Fig. 2.3d*) on one hypoxic/normoxic pair. For ease of illustration, only one RGS is shown, but to further enhance statistical robustness, multiple different RGS can be used to calculate multiple $d_{RGS}(hyp, norm)$. These distances from multiple RGS can then be used to derive a null distribution and can be compared to the hypoxia signature $d_{SIG}(hyp, norm)$ on each individual pair of normoxic and hypoxic samples. This is demonstrated in *Fig. 2.5* on one pair of samples: $d_{SIG}(hyp, norm)$ was compared to the $d_{RGS}(hyp, norm)$ from 500 different RGS. The red bar in *Fig. 2.5a* indicates that only the hypoxia signature belongs to the last quantile (Frequency = 1) when using the mean score. However, in *Fig. 2.5b*, the $d_{SIG}(hyp, norm)$ using ssGSEA score falls in the same quantile as 16 other $d_{RGS}(hyp, norm)$. This suggests that mean score works better than ssGSEA in this experiment using the Buffa signature. This can be measured by calculating a significance value (p) as below (with $\varepsilon = 0.00199$ representing the correction factor) and this was repeated for all hypoxic/normoxic pairs of samples in each GEO *Series*:

$$p = 1 - (percentile\ of\ score \div 100) + \varepsilon$$

The percentile of score is the number of distances (expressed as percentage) in the distribution that are lower than the $d_{SIG}(hyp, norm)$. For example, a percentile of score of 100% means that all the 500 $d_{RGS}(hyp, norm)$ are lower than the $d_{SIG}(hyp, norm)$ as in *Fig. 2.5b.* However, if a score in the normoxic sample is greater than that in the hypoxic sample

using the hypoxia gene expression signature, the significance between the two samples is automatically made non-significant ($p = 1$). The correction factor ε is used in the permutation test to prevent a zero p-value, occurring when no RGS outperforms the original signature. This correction factor is calculated as the lowest possible p obtainable from the permutation test when only one random signature performs better than the original signature ($1/501 = 0.00199$).

GEO Series have multiple hypoxic and normoxic samples, including replicates. All combinations of hypoxic samples and normoxic samples from the same GEO series were analysed as pairwise combinations, as opposed to averages across samples, to provide more detailed results and insights (*Fig. 2.4*).

The last step of the analysis consists of calculating the classification accuracy of each signature as the ratio between the number of significant pairwise combinations (threshold *p<0.005*) over the total number of combinations for each cancer type or experimental condition. The p-value corresponds to the probability of rejecting the null hypothesis, in which RGS perform better than the original signature. Therefore, a stringent threshold of p=0.005 was chosen in order to allow a maximum of three RGS to outperform the original signature and allow the original signature to still remain significant ($3/501 = \sim 0.005$). Accuracy is reported as the percentage of correct classifications. For example, if 80 hypoxic/normoxic pairwise combinations out of 100 tested on all the breast cancer cell lines are significant compared to their respective RGS (as described above), the accuracy is stated as 80% in breast cancer. This approach was repeated for each of the four scoring metrics (Mean, Median, ssGSEA[79] and GSVA[45]). A summary of the entire data workflow is shown in *Fig. 2.6*.

N.B. PCA1[44] and PLAGE[80] cannot be assessed using the above method as there is no way to tell if the direction of the implied variance using the hypoxia signature is the same as in

any RGS. For instance, PCA1 using the Buffa hypoxia signature may have a score of -0.5 in one sample and 0.5 in one RGS in the same sample. However, as the score is based on variance, positivity and negativity can vary and thus it is impossible to tell if the hypoxia score of -0.5 and 0.5 are both actually the same score or opposite.

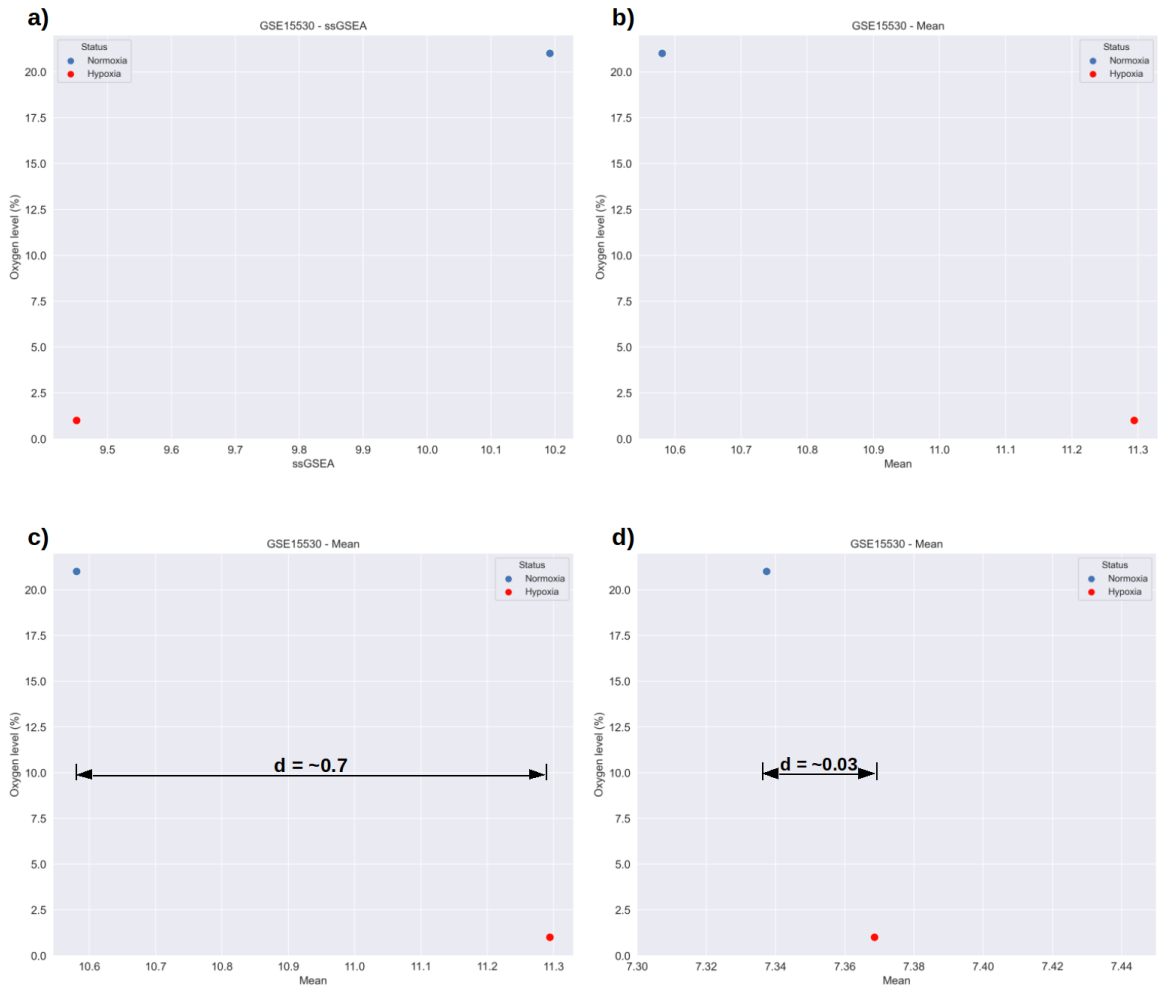
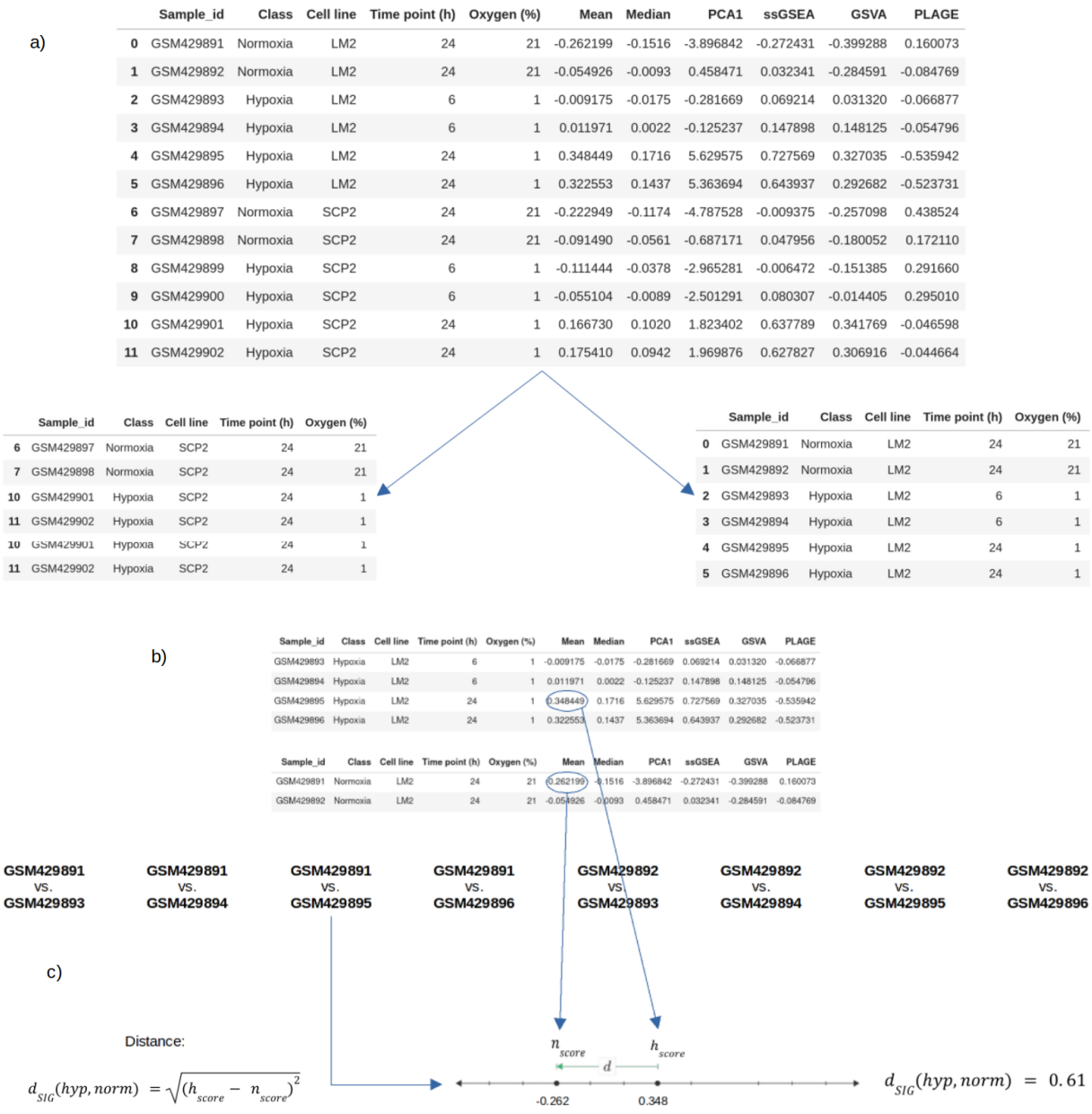


Figure 2.3: *Hypoxia scores and distance calculation against RGS*

The scatterplots were generated using Buffa signature on one normoxic (GSM390200, highlighted in blue) and one hypoxic (GSM390196, highlighted in red) sample from the Series GSE15530. In Panel a) the Mean (right) and the ssGSEA (left) scores are plotted on the x-axis against the percentage of Oxygen. The Mean score of the hypoxic sample is greater than the score of the normoxic sample. Viceversa, the ssGSEA score of the hypoxic

sample is lower than the score of the normoxic sample, resulting in a wrong classification. Panel b) shows that the distance calculated using the mean score between the two samples is ~24x higher (left image, $d = \sim 0.7$) than the distance observed between the same samples using a RGS of the same length (right image, $d = \sim 0.03$).



	Sample_id	Class	Cell line	Time point (h)	Oxygen (%)	Mean	Median	PCA1	ssGSEA	GSVA	PLAGE
0	GSM429891	Normoxia	LM2	24	21	-0.262199	-0.1516	-3.896842	-0.272431	-0.399288	0.160073
1	GSM429892	Normoxia	LM2	24	21	-0.054926	-0.0093	0.458471	0.032341	-0.284591	-0.084769
2	GSM429893	Hypoxia	LM2	6	1	-0.009175	-0.0175	-0.281669	0.069214	0.031320	-0.066877
3	GSM429894	Hypoxia	LM2	6	1	0.011971	0.0022	-0.125237	0.147898	0.148125	-0.054796
4	GSM429895	Hypoxia	LM2	24	1	0.348449	0.1716	5.629575	0.727569	0.327035	-0.535942
5	GSM429896	Hypoxia	LM2	24	1	0.322553	0.1437	5.363694	0.643937	0.292682	-0.523731
6	GSM429897	Normoxia	SCP2	24	21	-0.222949	-0.1174	-4.787528	-0.009375	-0.257098	0.438524
7	GSM429898	Normoxia	SCP2	24	21	-0.091490	-0.0561	-0.687171	0.047956	-0.180052	0.172110
8	GSM429899	Hypoxia	SCP2	6	1	-0.111444	-0.0378	-2.965281	-0.006472	-0.151385	0.291660
9	GSM429900	Hypoxia	SCP2	6	1	-0.055104	-0.0089	-2.501291	0.080307	-0.014405	0.295010
10	GSM429901	Hypoxia	SCP2	24	1	0.166730	0.1020	1.823402	0.637789	0.341769	-0.046598
11	GSM429902	Hypoxia	SCP2	24	1	0.175410	0.0942	1.969876	0.627827	0.306916	-0.044664

	Sample_id	Class	Cell line	Time point (h)	Oxygen (%)
6	GSM429897	Normoxia	SCP2	24	21
7	GSM429898	Normoxia	SCP2	24	21
10	GSM429901	Hypoxia	SCP2	24	1
11	GSM429902	Hypoxia	SCP2	24	1
10	GSM429901	Hypoxia	SCP2	24	1
11	GSM429902	Hypoxia	SCP2	24	1

	Sample_id	Class	Cell line	Time point (h)	Oxygen (%)
0	GSM429891	Normoxia	LM2	24	21
1	GSM429892	Normoxia	LM2	24	21
2	GSM429893	Hypoxia	LM2	6	1
3	GSM429894	Hypoxia	LM2	6	1
4	GSM429895	Hypoxia	LM2	24	1
5	GSM429896	Hypoxia	LM2	24	1

Figure 2.4: *Pairwise combination and distance calculation of hypoxic/normoxic pairs*

The Series GSE17188 is characterised by two MDA231 breast cancer sub cell lines (LM2 and SCP2) including 12 samples, 8 hypoxic and 4 normoxic, whose GEO sample identifiers (starting with GSM) are reported in the *Sample_id* column. These unique identifiers are automatically assigned when samples are uploaded on the GEO archive (a). The dataset was split by cell type, as reported in Chapter 2.4.1, and analysed separately. For example, LM2 cell lines data were divided into hypoxic and normoxic samples (b). Hypoxic samples

were compared to normoxic samples defining a list of all possible pairwise combinations. Finally, for each pairwise combination, the geometric distance was calculated.

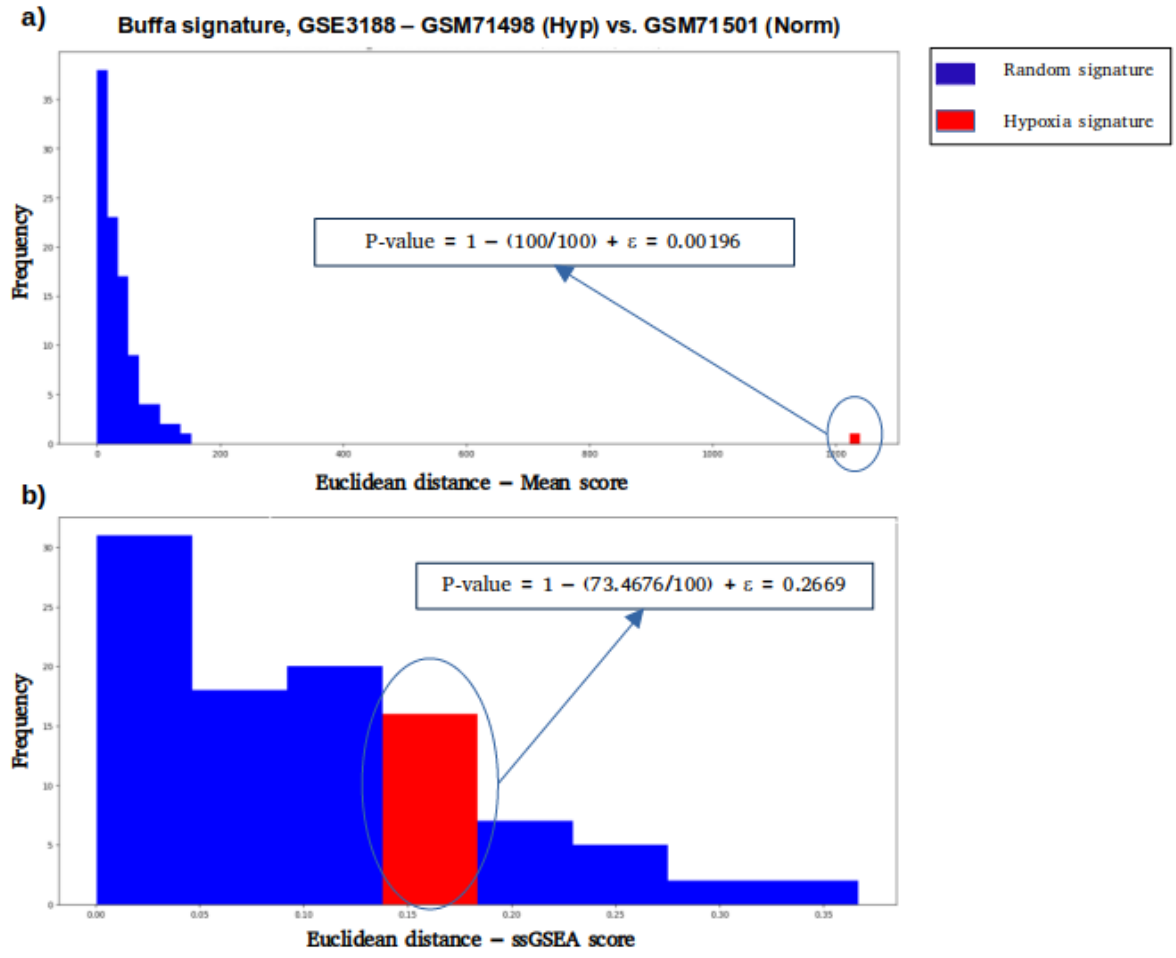


Figure 2.5: *Distribution of the distances calculated using hypoxia signatures and 500 random gene sets for a pairwise combination (GSM71498 vs. GSM71501) from the Series GSE3188 using Buffa signature with mean (a) and ssGSEA (b) scores.*

Y axis represents the number of gene sets falling in a given quantile. X axis represents the Euclidean distance calculated between the two samples. Blue bars only contain random gene sets, red bars denote the quantile where the hypoxia signature belongs to. A p-value of 0 means that no RGS outperforms the original signature.

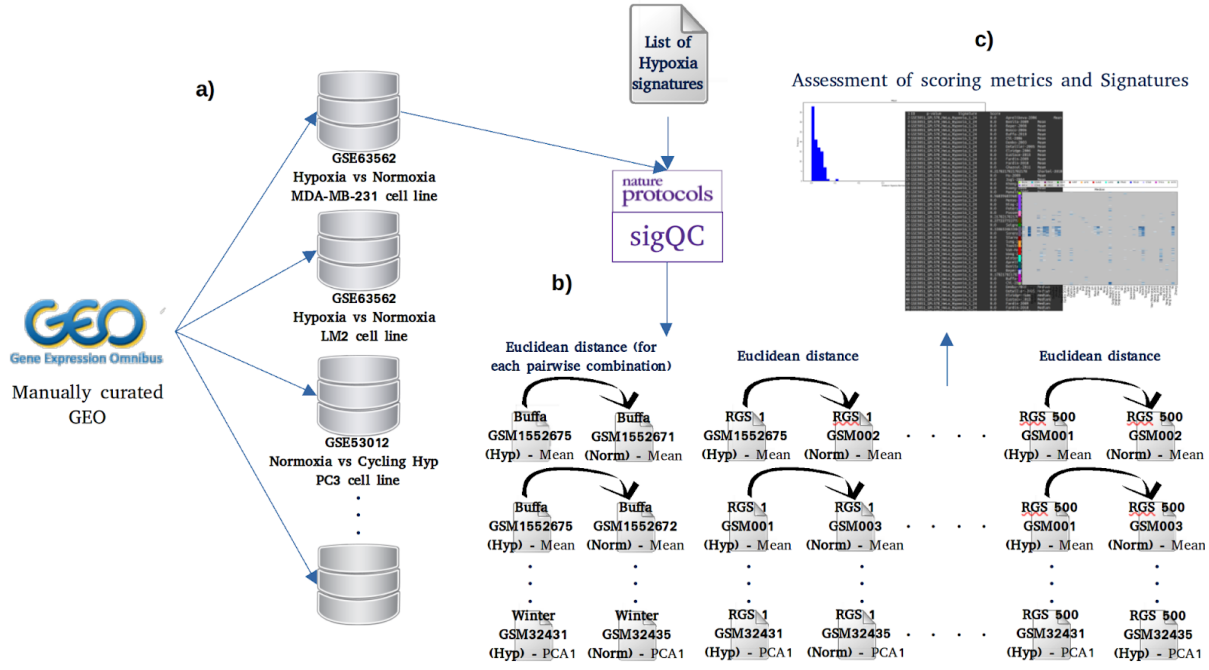


Figure 2.6: *A summary of the data workflow presented in this study.*

After a manual curation of GEO Series and the division into multiple subsets according to experimental conditions and cell line type (a), the four hypoxia scores (mean, median, GSVA and ssGSEA) have been calculated using sigQC on each hypoxia signature. For each score and hypoxia signature, the Euclidean distance was calculated on each hypoxic/normoxic pairwise combination in random gene sets versus hypoxia signatures (b). Finally, the assessment of the scoring methods (c) was obtained by calculating the accuracy of classification of a signature via a comparison of the p-values on each experimental condition.

2.6 Evaluating hypoxia signatures and scores in clinical data

The evaluation of hypoxia signatures in clinical data is more challenging compared to cell lines. Unlike cell line experiments, where it is possible to induce a controlled hypoxic response (e.g. culturing in hypoxic chambers), this cannot be done in clinical samples. Consequently, it is not possible to evaluate the performance of a signature following the same procedure that was used for cell line experiments.

To attempt to overcome this limitation, perhaps a good compromise in evaluating the efficacy of hypoxia signatures in clinical tumour samples consists of comparing their performance against a set of normal tissues. The TCGA project has profiled tumour-adjacent normal tissues (NAT) in a number of patients with different cancer types. Interrogation of these samples has shown NAT to have a transcriptomic profile that is highly correlated with healthy normal tissues[85], so using NAT in lieu of a normoxic control appears a potentially worthwhile strategy as it is unlikely healthy tissue would be more hypoxic than cancer samples.

Another major issue in the field is that no large datasets exist including multiple markers of hypoxia such as oxygen electrode measurements, immunohistochemistry and F-MISO PET. Even if this were in existence, which approach should be considered the most accurate is an area of considerable debate. Thus, one is left in a situation of inferring hypoxic status from existing samples and so multiple evaluation steps are needed to define the most promising signature for clinical use. I defined three steps to assess hypoxia signatures to identify the most promising signature and score for use in clinical samples:

1. **Hypoxia scores in cancer samples vs. NAT**: hypoxia scores from cancer biopsies should not, on average, be lower than NAT

   ∴ Evaluation Step 1: Do hypoxia signatures scores tend to be higher in tumours?

2. **Hypoxia scores vs. scores from RGS**: the difference observed between hypoxia signature scores from cancer samples and NAT should be more significant than the equivalent difference observed from RGS.
   ∴ *Evaluation Step 2: Do hypoxia signatures scores tend to outperform random gene signatures of the same length when comparing NAT and tumour tissues?*

3. **The prognostic power of hypoxia signatures**: a valid hypoxia signature is associated with poor prognosis across several cancer types[86].
   ∴ *Evaluation Step 3: Do hypoxia signature scores from tumours confer prognostic information?*

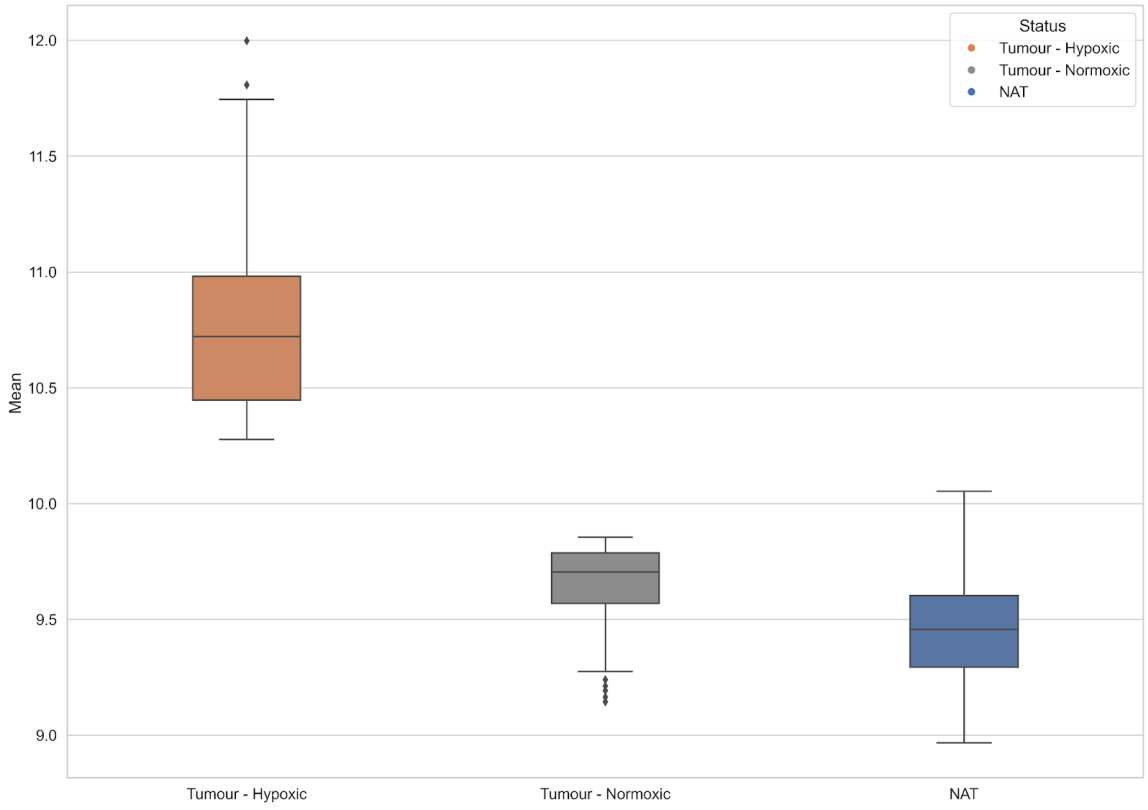
2.6.1 Evaluation Step 1: Do hypoxia signatures scores tend to be higher in tumours?

One would expect tumour tissue to be, on average, more hypoxic than normal tissue due to metabolic demand outstripping aberrant blood vessel supply. Normoxic tumours, in fact, should have hypoxia scores similar to NAT (*Fig. 2.7a*). Here the difference across solid tumours in TCGA for each signature was examined. The distributions of hypoxia signature scores in the cancer tissues with NAT were compared using the nonparametric Mann-Whitney U (example shown in *Fig. 2.7b*). Finally, by assuming that the average hypoxia score in the cancer samples will exceed that in NAT (but not the other way around), an alternative hypothesis has been defined as follows:

*Let **T(u)** and **N(u)** be the cumulative distribution functions of the distributions underlying x and y, respectively. Then the distribution underlying x (Tumour hypoxia scores) is stochastically greater than the distribution underlying y (NAT hypoxia scores), i.e. **T(u) > N(u)** for all u.*

I.e. if the hypoxia signature scores in the cancer group are on average lower than NAT then the assumption is violated and the difference is called non-significant.

a)


Status
Tumour - Hypoxic
Tumour - Normoxic
NAT
Mean
12.0
11.5
11.0
10.5
10.0
9.5
9.0
Tumour - Hypoxic
Tumour - Normoxic
NAT

b)

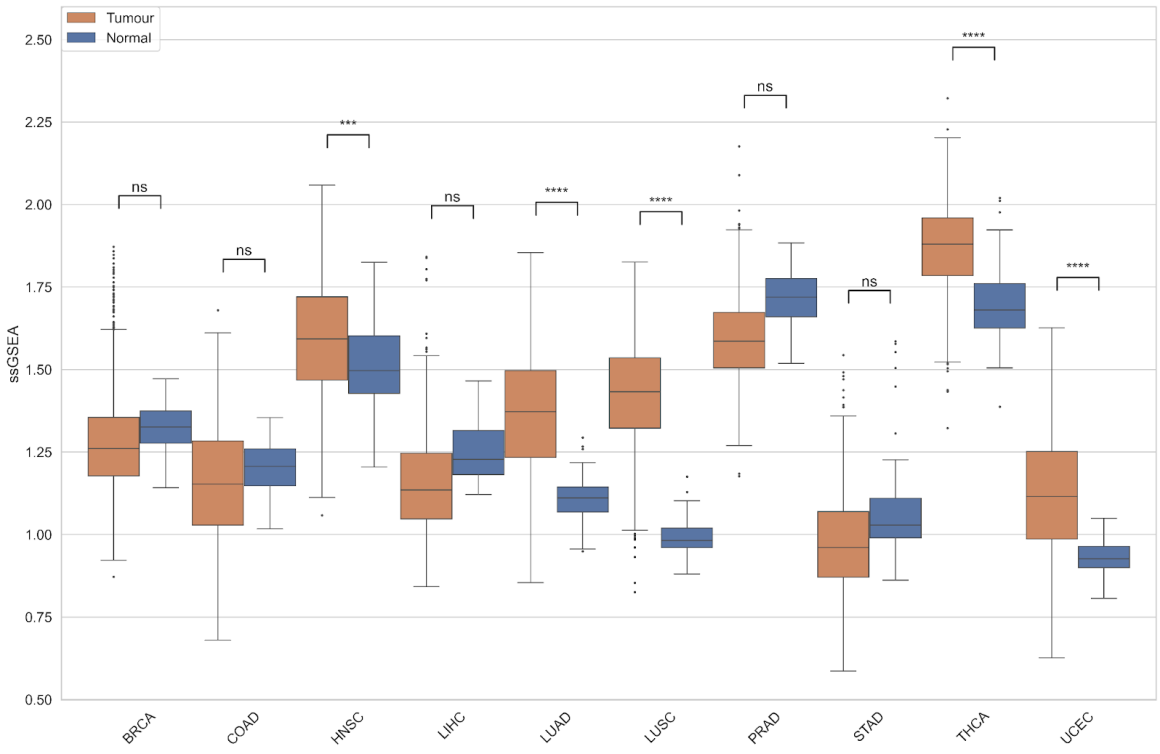

Tumour
Normal
ssGSEA
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
0.50
ns
ns
***
ns
****
****
ns
ns
****
****
BRCA
COAD
HNSC
LIHC
LUAD
LUSC
PRAD
STAD
THCA
UCEC

Figure 2.7: *An example of Evaluation Step 1*

Hypoxia scores in normoxic tumours should have similar values to those in NAT. Boxplot *(a)* shows the difference between hypoxic and normoxic samples (defined as the highest and lowest quantiles respectively) versus NAT using the Buffa signature mean score. The boxplot in *(b)* represents the distribution of Sorensen ssGSEA scores across the 10 TCGA cancer types included in this study. Blue and orange boxes describe normal and cancer tissues respectively. P-values are calculated using the nonparametric Mann-Whitney-U. Differences are considered as non-significant if the distribution of NAT hypoxia scores is higher than cancer tissues, as in Liver Hepatocellular Carcinoma (LIHC) and Prostate Adenocarcinoma (PRAD).

2.6.2 Evaluation Step 2: Do hypoxia signatures scores tend to outperform random gene signatures of the same length when comparing NAT and tumour tissues?

Chapter 2.5 demonstrates the importance of comparing hypoxia signatures with RGS. The same concepts can be applied to clinical samples. However, instead of evaluating pairs of samples, here the focus will be on testing if significant differences can be observed between the distributions of hypoxia scores in tumour vs. NAT. Three main analyses are involved in this evaluation step. For each signature, score and cancer type:

1. The non-parametric Mann-Whitney U test on the original hypoxia signature is used to calculate a p-value, as explained in the Evaluation Step 1 (Chapter 2.6.1)
2. 500 RGS are generated and the p-value between the distribution of tumour hypoxia scores vs NAT hypoxia scores is calculated with Mann-Whitney U
3. Finally, a Signature Performance Index (SPI, expressed in percentage) was derived by calculating how many times the p-value from the original hypoxia signature is lower than any of the 500 RGS p-values. For example, if 490 RGS out of 500 have a higher p-value than the original hypoxia signature, the SPI will be equal to 98%

It is noteworthy to mention that the sample size of both cancer and adjacent normal tissues will be the same in each of the 500 RGS. This approach enables the comparison of 500 p-values from the simulations with the p-values from the original signature to derive the SPI and determine if the hypoxia signature is more “significant” than any RGS.

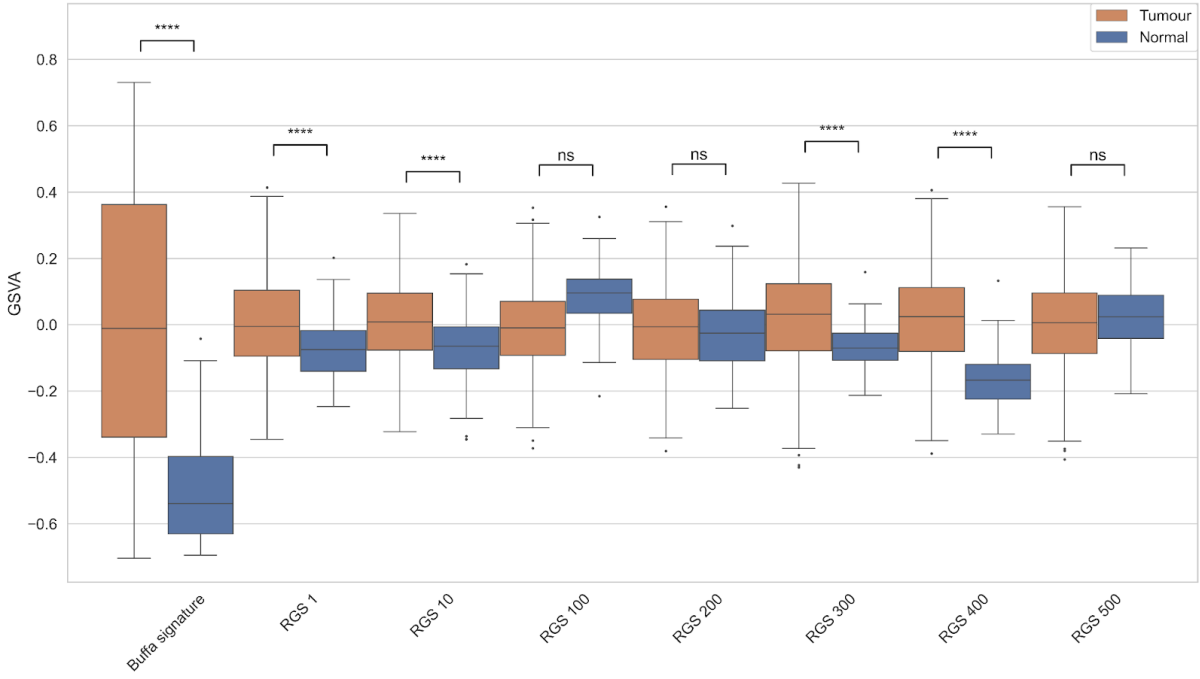


Figure 2.8: *An example of Evaluation Step 2*

The plot represents the distribution of the Buffa GVSA score in TCGA breast cancer (BRCA) and the performance of seven random gene sets (RGS 1, RGS 10 etc). Boxplot and statistical conventions as in *Fig. 2.7*. The Buffa signature has the lowest p (4.744e-41).

2.6.3 Evaluation Step 3: Do hypoxia signature scores from tumours contain prognostic information?

Hypoxia is a key characteristic of solid tumours and previous studies have shown its correlation with poor prognosis of cancer patients[87–89]. Thus, this step consists of evaluating the prognostic power of hypoxia signatures that performed well on the two previous steps in the TCGA dataset. Prognostic performance was investigated by using both two commonly used survival analysis techniques, the Kaplan-Meier (KM) estimator and the Cox Proportional Hazard (CPH) model.

The KM estimator is a non-parametric statistic used to derive the survival function from lifetime data. Traditionally, the mean or median values of the hypoxia scores on clinical samples are used to categorise the samples into “Low” and “High” hypoxic groups. This evaluation step investigates whether comparing tumour samples against NAT might also be worthwhile. Survival analysis of the following groups was also carried out:

- Those samples *above and below the maximum NAT* sample hypoxia score (max normal)
- Those samples *above and below the mean NAT* sample hypoxia score (mean normal)
- Those samples *above and below the median NAT* sample hypoxia score (median normal)

The log-rank test was used to compare the survival distribution between samples. This approach, in addition to the CPH, was included to test whether selecting a heuristic threshold to determine hypoxic and normoxic samples is valid. The KM estimator was calculated using the Python module lifelines v.0.26.4.

CPH analyses were also carried out using Python module lifelines v.0.26.4. The proportional hazards assumption was tested using Schoenfeld residuals. Observation time

was ended by date of death or end of monitoring period, from five years since diagnosis. Hypoxia scores were investigated as continuous variables in multiply-adjusted models. Models were adjusted for the following categorical variables: age (10 year age groups), stage and histological type. Sex was used as a strata in the model.

Chapter 3: Identifying and evaluating different hypoxia signature scores

3.1 Identifying hypoxia signatures

A key step in this work was to identify hypoxia gene expression signatures from the literature. I then go on to establish which have the widest utility and applicability. Hypoxia signatures were identified by interrogating Pubmed, Scopus and Web of Science using the search terms ‘hypoxia signature’ or ‘hypoxia’ AND/+ ‘signature’ as per Harris et al. 2015. This search was discussed with the Bodelian library team and felt to be appropriate.

This systematic approach identified 53 signatures which tended to be reported as gene symbols. This was 21 more than previously identified by our group (*Table 3.1*). The signatures ranged from 759 genes long (Starmans[74]) to four genes long (Mo[90]). Their mean and median lengths were 72 and 32 respectively. 22 signatures were derived from clinical samples, leaving 31 derived from solely *in vitro* approaches. Within these subdivisions, signatures derived from clinical samples ranged from 4 to 166 in length with a median number of 19 genes (to the nearest gene). Whereas *in vitro* signatures tended to be longer with a median number of genes as 42. The median *in vitro* signature length in this work was smaller than that seen by Harris et al in 2015 (42 as against 85) and in signatures derived from clinical samples it was 19 vs 61. All the signatures identified and their gene symbols are found in the *Appendix 1.*

Surprisingly, perhaps, no gene was found in all 53 signatures. This could reflect their origin in terms of the derived tissues’ response to hypoxia (different cell lines/ tumour types) or this might reflect differences in experimental conditions. Nonetheless, several genes were seen in a number of different signatures and this is summarised in *Table 3.2*. The top 20 most frequently occurring genes from this analysis are found in the left hand column whereas the right hand column details the results of Harris et al.

The comparison between the top 20 genes from the hypoxia signatures in this work and from Harris et al. yields a number of similarities and some striking differences. The

major difference seen is that VEGFA is the gene most frequently observed across these 53 signatures but does not feature in the top 20 genes seen previously. This is the result of a "trivial" change in nomenclature, as in the previous work several signatures listed VEGF and others listed VEGFA. After correcting for this artefact, VEGFA is now found to be the most common gene affected in hypoxia, participating in 26 signatures (*Fig. 3.1a*). Overall, some signatures share a good number of genes with other hypoxia signatures whereas others do not share any gene. This depends on several factors, such as number of genes in the signature or the cancer types they have developed for (*Fig. 3.1b*).

VEGFA is a growth factor that promotes angiogenesis. As such, it is produced in response to low oxygen tensions in tissues, under the influence of the HRE/HIF-1 system. It acts particularly on endothelial cells, promoting the formation of capillaries. It is a homodimer, with eight conserved cysteine residues organised into a characteristic 'cystine knot' structure. Its target, VEFGR (particularly VEFGR2) is a tyrosine kinase receptor situated on endothelial cells and, typically for a growth factor, stimulates cell growth and division through a series of mitogenic kinases (PI3 kinase and Akt, protein kinase C, MAPK). VEGFA also increases blood vessel permeability through its suppressive effects on cell adhesion molecule production (ICAM-1, VCAM1) and was originally identified as a vascular permeability factor. This combination of increased endothelial cell proliferation, and loosening of cell-cell contacts, is instrumental in driving capillary growth[91].

VEGFA is of particular interest in the growth of tumours. Solid tumours in particular become hypoxic due to insufficient blood supply, which is overcome by their production of VEGFA, promoting angiogenesis locally. VEGFA antagonists have been explored as anti-cancer agents. Interestingly, VEGFA also suppresses the anti-tumour immune response by suppressing the maturation of dendritic cells and CD8+ T cells, while switching the T cell population in tumours to the more passive Treg lymphocytes[92].

The second most commonly featured gene is NDRG1. NDRG1 (N-myc downstream regulated gene 1) was the gene most frequently found in the analysis of previous signatures.

It has a hypoxia response element in its promoter and has been shown to be upregulated by the HIF system in hypoxia, while decreasing in mutant VHL systems. Defects in NDRG1 are associated with demyelination neuropathies (e.g. Charcot-Marie-Tooth Type 4D syndrome)[93].

High levels of NDRG1 are found in several types of tumour, where it appears to be induced as a tumour suppressor protein linked to the TP53 system. Its expression is inhibited by the oncogene N-myc. As such it promotes apoptosis and promotes DNA repair via the O6-methylguanine-DNA methyltransferase (MGMT). The precise function of NDRG1 is unknown, but its pleiotropic effects, and range of interactions, suggest it may have some role as a scaffold protein; it also stabilises the spindle checkpoint (by interaction with tubulin) and interacts with cadherin (inhibiting metastasis). This complex series of interactions may explain why it is found to be protective in some cancers (prostate, colon and pancreatic cancer), while associated with poor prognosis in others (bladder, hepatocarcinoma)[94].

The next most commonly featured gene is PGK1, ranked fourth previously and third in this analysis. Perhaps surprisingly, PGK1 is a glycolytic enzyme, catalysing the first ATP-generating step in glycolysis, and thus a central enzyme in a general housekeeping function in cells. Like other glycolytic enzymes (e.g. LDH-A, PDK1), its levels rise under the influence of HIF1α as glycolysis becomes more important to energy production as oxygen levels decline. Unusually, however, PGK1 is secreted into the bloodstream and has been claimed to promote angiogenesis. In addition, PGK1 will stimulate the production of HIF1α by binding to its HRE, giving rise to a potential positive feedback system. Possibly as a result of one of these effects, high levels of PGK1 are associated with a variety of tumours (endometrial, breast and gastric cancers). The very different functions of PGK1 as both enzyme and transcription factor have led to designating PGK1 a ‘moonlighting’ enzyme.

P4HA1 is the fourth commonest gene in this analysis and was in the top five previously (third). P4HA1 is the catalytic subunit of the tetrameric prolyl hydroxylase (A2B2), participating in the hydroxylation of collagen and the maturation of this structural protein. Its level rises under the influence of the HIF system, and it has been suggested that levels of

this enzyme alone provides a reliable hypoxia signal, at least in head and neck squamous cell carcinomas[95]. As implied above, P4HA1 is commonly raised in cancers, possibly as collagen is needed to provide a structural framework for tumour growth. There is an interesting interplay between the various prolyl hydroxylases and oxygen sensing. Hydroxylation of HIF1α involves a proline hydroxylase, PHD, and the resultant hydroxylated HIF1α is degraded in the cytoplasm, leading to the absence of expression of the hypoxia stimulated genes. Increased levels of P4HA1, however, leads to the stabilisation of HIF1α and expression of these genes. This appears not to be due to effects at the gene level - there is no change in HIF synthesis - but paradoxically the P4HA1 hydroxylase inhibiting HIF1α hydroxylation. This has been suggested to be due to increased P4HA1 outcompeting the PHD for the 2-oxoglutarate that both enzymes require as a cofactor. As an additional twist to this system, the result is that P4HA1 will increase HIF1α levels, which in turn will promote P4HA1 levels, another potential positive feedback system[96].

The fifth most frequent gene across all the hypoxia signatures is DDIT4. DDIT4 (DNA damage induced transcript 4) - also known as REDD1 - is another protein known to be raised in hypoxia, and contains a HRE in its promoter region (induced by the HIF system). However, its levels are also induced to rise by other stresses such as radiation and energy depletion, possibly by interacting with the TP53 system.

DDIT4 is a heterodimer and acts by inhibiting mTOR (the master controller of cell growth) by activating the TSC complex - hence the link to p53. This means it will repress cell growth and promote autophagy. DDIT4 is short lived (minutes), being subject to ubiquitinylation and proteosome destruction like HIF-1α and p53, suggesting it can respond rapidly to increasing or decreasing stress. DDIT4 tends to be overexpressed in tumour cells (so consistently that it may be used as a cancer marker in some tissues). Its inhibitory effects on mTOR, suppressing cell growth would suggest it has tumour suppressor activity, but paradoxically - like many of these proteins - it also can act as an oncogene in ovarian cancer cell lines. High levels of DDIT4 are associated with resistance to cancer therapies,

although whether this is due to the association with hypoxia is unknown. A recent review of this protein is given by Tirado-Hurtado et al.[97].

Pubmed ID	Signature	Cell line	Cell/Tissue origin*	# gene symbols	Experimental conditions	Ref.
10706099	Koong 2000	FaDu SiHa	Pharynx (transformed keratinocytes) Cervix (transformed keratinocytes)	10	0.05%O2, 24h	[48]
12947397	Denko 2003	NCK, NDK, NCF, SiHa, C33a, FaDu	Cervix (keratinocytes and stromal fibroblasts)	80 → 72 (?)	Various as detailed in [48,98–100]	[49]
15093745	Jogi 2004	SK-N-BE(2), SH-SY5Y, SK-N-F1, IMR-32, LA-N-2, LA-N-5, SK-N-RA	Brain (neuroblastoma)	107 → 103	1%O2, 72h	[50]
15100389	Ning 2004	HAECs	Aortic endothelial cells	104 → 99	1%O2, 8-24h	[51]
15374877	Manalo 2005	ECs	Coronary artery endothelial cells	107 → 105	1%O2, 24h	[52]
15833863	Wang 2005	HEK293T	Kidney	56 → 55	1%O2, 16h	[53]

15994966	Detwiller 2005	HT1080 10T1/2 SVR HEK293	Fibrosarcoma Fibroblast (mus musculus) Pancreas (mus musculus) Kidney	27	1%O2, 48h	[54]
16417408	Chi 2006	ECs SMCs HMECs RPTECs	Coronary artery endothelial cells Smooth muscle cells Mammalian epithelial cells Renal proximal tubule epithelial cells	111	0.1-2%O2, 1-24h	[58]
16507782	Mense 2006	HFAs	Fetal astrocytes	111 → 94	1%O2, 24h	[56]
16565084	Elvidge 2006	MCF7	Breast	181 → 173	1%O2, 16h DMOG, 16h	[59]
16595741	Peters 2006	HPAECs	Pulmonary artery endothelial cells	159 → 158	1%O2, 8-24h	[60]
16740701	Aprelikova 2006	MCF7	Breast	236 → 230	0.5%O2, 8h	[57]
16849508	Bosco 2006	PBMC	Peripheral blood monocytes	177 → 173	1%O2, 16h	[55]
17187782	Shi 2007	LX-2	Hepatic stellate cells	32 → 31	1%O2, 8-24h	[62]

17320280	Sung 2007	CNE-2, C666-1, HONE-1, HK1	Head and neck (nasopharyngeal carcinoma)	90	0.1%O2, 16h	[63]
17409455	**Winter 2007**	Clinical samples	Head and neck (squamous cell carcinoma)	99 → 97		[41]
17532074	Seigneuric 2007 (common)	HMECs from Chi 2006	Mammary epithelial cell	14	0-2%O2, 1-24h	[61]
17532074	Seigneuric 2007 (early0)	HMECs from Chi 2006	Mammary epithelial cell	71 → 68	0-0.02%O2, 1-6h	[61]
17532074	Seigneuric 2007 (early2)	HMECs from Chi 2006	Mammary epithelial cell	34 → 31	2%O2, 12-24h	[61]
18984585	Beyer 2008	HeLa HEK293 786-0	Cervix Kidney Renal cancer	159 → 158	0.2-1%, 24h	[64]
19291283	**Hu 2009**	Clinical samples	Breast	13		[68]
19491311	Benita 2009	DLD-1 HCT116, SW480, Lovo Panc-1 HeLa MCF7	Colorectal Colon Pancreas Cervix Breast	57 → 54	1%O2, 18h	[66]

19832978	Fardin 2009	GI-LI-N, ACN, GI-ME-N, IMR-32, LAN-1, SK-N-BE(2)C, SK-N-F1, SK-N-SH	Brain (neuroblastoma)	8	1%O2, 18h	[67]
19884889	Lendhal 2009	HeLa, P493-6, HCT116, Hep3B, MCF7, RCC4, SK-N-BE(2)C, (E-MEXP-836)	Cervix, Burkitt's lymphoma, Colon, Liver, Breast, Kidney (VHL mutated), Brain (neuroblastoma)	23	Different conditions	[101]
20087356	**Buffa 2010**	Clinical samples	Head and Neck (squamous cell carcinoma), Breast	51		[71]
20416888	**Ghorbel 2010**	Clinical samples	Cyanotic tetralogy of Fallot (TOF)	166 → 158		[72]
20429727	Sorensen 2010	SiHa FaDuDD UTSCC5, UTSCC14, UTSCC15	Cervix Head and neck (squamous cell carcinomas)	27 → 26	0-5% O2, 24h	[70]
20592013	Van Malenstein 2010	HepG2	Liver	4	2%O2, 72h	[65]
20652058	Fardin 2010	GI-LI-N, ACN, GI-ME-N, IMR-32, LAN-1, SK-N-BE(2)C, SK-N-F1, SK-N-SH	Brain (neuroblastoma)	35	1%O2, 18h	[69]
21325071	**Ghazoui 2011**	Clinical samples	Breast	70 → 68		[73]

21846821	Toustrup 2011	UTSCC5, UTSCC14, UTSCC15 FaDu SiHa	Head and neck Cervix	15	O2 < 2.5 mm Hg (electrode)	[42]
22356756	Starmans 2012	DU145 HT29 MCF7	Prostate Colon Breast	759 → 756	0%O2, 1-24h	[74]
22890239	**Halle 2012**	HeLa, SiHa, CaSki, Clinical samples	Cervix	31	0.2%, 24h	[102]
23820108	**Eustace 2013**	Clinical samples	Laryngeal cancer Bladder cancer	26 → 25		[43]
25216520	**Boidot 2014 (Continuous Hypoxia)**	MCF-7, MDA-MB-231, T47D, A549, Widr, HCT116 WTP53, HCT116 -/-P53, HT29, Colo-205, LoVo, HCT15, SiHa, PC3, U373, HepG2, Hep3B, PLC/PRF/5, SK-HEP-1, A498, HT1080, Clinical samples	Breast Colon Prostate Colorectal Liver Fibrosarcoma	98 → 93 (~50 based on heatmap)	1%, 24h	[103]
25216520	**Boidot 2014 (Cyclic Hypoxia)**	MCF-7, MDA-MB-231, T47D, A549, Widr, HCT116 WTP53, HCT116 -/-P53, HT29, Colo-205, LoVo, HCT15, SiHa, PC3, U373, HepG2, Hep3B, PLC/PRF/5, SK-HEP-1, A498, HT1080, Clinical samples	Breast Colon Prostate Colorectal Liver Fibrosarcoma	96 → 90 (~50 based on heatmap)	Cycling Hypoxia - 30 min 1% O2 + 30 min Normoxia, 24h	[103]

25461803	**Ragnum 2015**	22Rv1, LNCaP, PC-3, DU 145, Clinical samples	Prostate	32	0.2%, 24h	[104]
27012812	**Fjeldbo 2016**	Clinical samples	Cervix	6		[105]
28324887	**Suh 2017**	Clinical samples	Head and neck	21 (5)		[106]
28400426	**Yang 2017**	Clinical samples	Bladder	24		[47]
30037853	Ye 2018	MCF-7, MCF10A, MCF12A, MDA-MB-157, MDA-MB-175, MDA-MB-231 , MDA-MB-436, MDA-MB-468, SKBR3, SUM1315MO2, SUM185PE, SUM229, SUM149PT, SUM159PT, SUM225CWN, T47D, ZR-75-1	Breast	42	1%, 24h	[107]
29729848	**Yang 2018 (Prostate)**	PNT2-C2, LNCaP, DU-145, PC-3, Clinical samples	Prostate	28 (14)	1%, 24h	[108]
29423096	**Yang 2018 (Sarcoma)**	HT1080, SKUT1, sNF96.2, 93T449, SW684, SW872, SW982, Clinical samples	Soft tissue sarcoma	24	1%, 24h	[109]

30257451	Trong 2018	NCH551b, NCH612, NCH620, NCH645, NCH421k, NCH601, NCH644, NCH660h	Brain (glioma)	5 (2)	1.5%, 72h	[110]
30973670	Chen 2019	A549, HCC827	Lung (adenocarcinoma)	17	1%, 72h	[111]
31572060	**Zou 2019**	Clinical samples	Colorectal	14 (9)		[112]
32887635	**Zhang 2020**	Huh-7, HepG2, Clinical samples	Liver	3	0-1%, 24 h	[113]
32724434	**Wang 2020**	Clinical samples	Breast	14 (7)		[114]
33133157	**Shou 2020**	Clinical samples	Skin (melanoma)	7 (3)		[115]
32500034	**Lin 2020**	Clinical samples	Brain (glioma)	5		[116]
32655624	**Mo 2020**	Clinical samples	Lung (adenocarcinoma)	4		[90]

32655701	**Sun 2020**	Clinical samples	Early-stage lung (adenocarcinoma)	16 (11)		[117]
32887267	Tardon 2020	Ge835, Ge898, Ge904, LN18, and LN229	Brain (glioblastoma multiforme)	19	1%, 48h	[118]

Table 3.1: *Published hypoxia signatures assessed in this study*

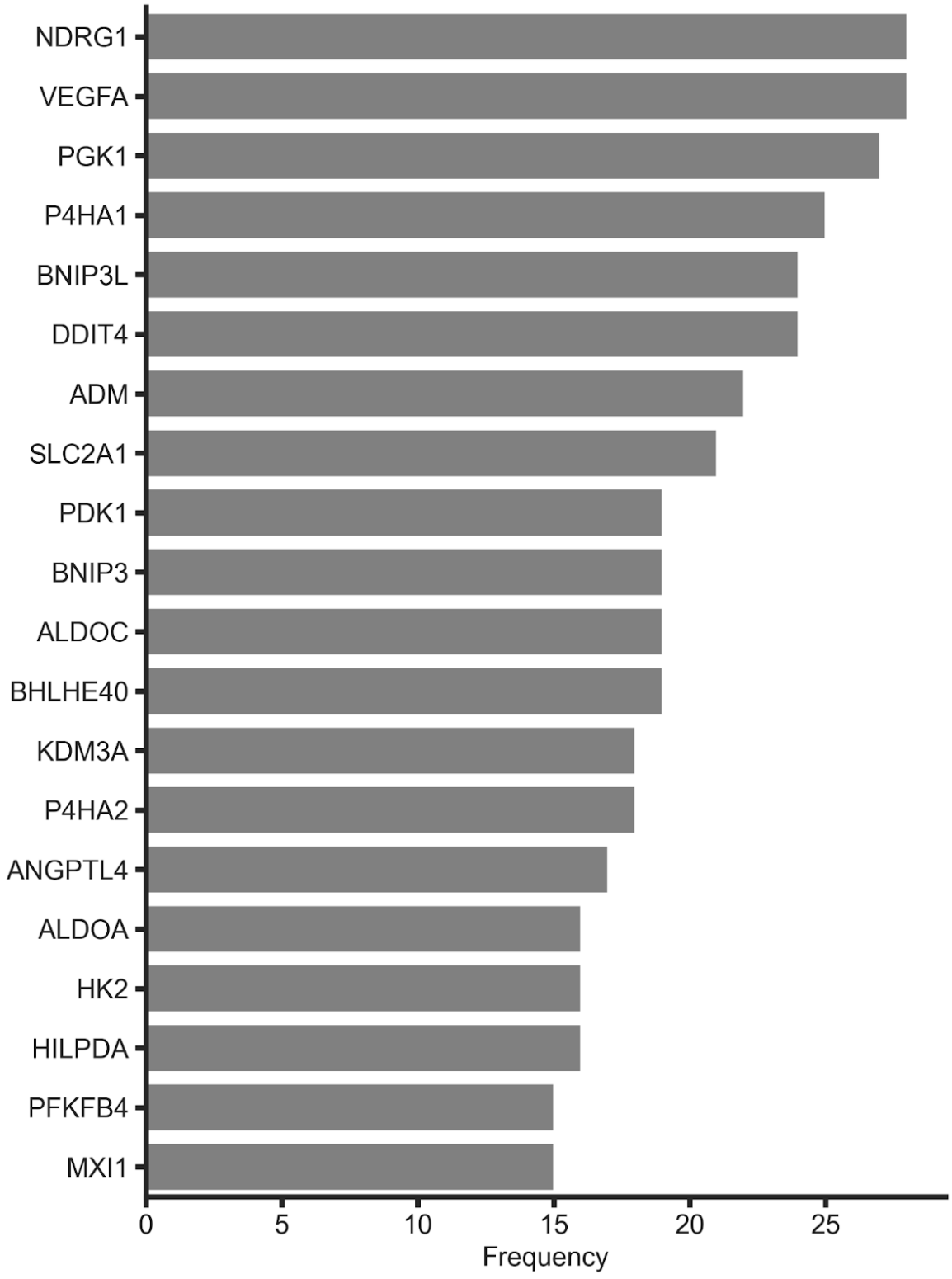
Published hypoxia signatures, identified by PubMed ID, first author name and year of publication. Table indicates originating group, source of tissue, associated malignancy and hypoxia conditions tested. The number of genes identified in the signature is given in column 5, the → symbol indicates the number of genes that survived the reannotation step (Chapter 2.2). Downregulated genes (if any) are reported in parentheses and a question mark indicates whether it was possible to identify the up- and down-regulated genes in the signature from the manuscript.

Top 20 most frequent genes in hypoxia signatures (Di Giovannantonio 2022)		Most frequent genes in hypoxia signatures using best annotation available (probeset IDs from Harris et al. 2015)	
VEGFA	Vascular Endothelial Growth Factor A	NDRG1	N-myc downstream regulated 1
NDRG1	N-myc downstream regulated 1	P4HA1	prolyl 4-hydroxylase, alpha polypeptide I
PGK1	phosphoglycerate kinase 1	BNIP3L	BCL2/adenovirus E1B 19kDa interacting protein 3-like
P4HA1	prolyl 4-hydroxylase, alpha polypeptide 1	PGK1	phosphoglycerate kinase 1
DDIT4	DNA Damage Inducible Transcript 4	ADM	adrenomedullin
BNIP3L	BCL2/adenovirus E1B 19kDa interacting protein 3-like	BNIP3	BCL2/adenovirus E1B 19kDa interacting protein 3
PDK1	Pyruvate Dehydrogenase Kinase 1	BNIP3P1	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 1
BNIP3	BCL2/adenovirus E1B 19kDa interacting protein 3	ALDOC	aldolase C, fructose-bisphosphate
ALDOC	aldolase C, fructose-bisphosphate	PLOD2	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2
SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1	P4HA2	prolyl 4-hydroxylase, alpha polypeptide 2
P4HA2	Prolyl 4-Hydroxylase Subunit Alpha 2	DDIT4	DNA-damage-inducible transcript 4
ADM	adrenomedullin	MXI1	MAX interactor 1, dimerization protein
HILPDA	Hypoxia Inducible Lipid Droplet Associated	ERO1L	ERO1L ERO1-like (S. cerevisiae)
FAM162A	Family With Sequence Similarity 162 Member A	SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1
BHLHE40	Basic Helix-Loop-Helix Family Member E40	ANGPTL4	angiopoietin-like 4
ANGPTL4	angiopoietin-like 4	-	-
KDM3A	Lysine Demethylase 3A	-	-
ERO1A	Endoplasmic Reticulum Oxidoreductase 1 Alpha	-	-
MXI1	MAX Interactor 1, Dimerization Protein	-	-
HK2	Hexokinase 2	-	-

Table 3.2: *The most frequently appearing genes across 53 published hypoxia signatures.*

The table is ordered in descending order of frequency of genes found in hypoxia signatures (e.g. NDRG1 was the second most commonly found gene in signatures investigated by Di Giovannontio et al. but most commonly found in those investigated by Harris et al.) The left hand side of the table was generated during the systematic hypoxia signature identification stage of this thesis (see methods for search query). The gene symbol and brief description of the gene is given. The right hand side of the table is a comparison with the study of Harris et al who identified 32 hypoxia signatures in 2015. Matched colours indicate genes found in both analyses, and highlight the rank of each gene in each study to easily contrast how genes in common to many signatures have changed in rank between the two studies.

a)


NDRG1
VEGFA
PGK1
P4HA1
BNIP3L
DDIT4
ADM
SLC2A1
PDK1
BNIP3
ALDOC
BHLHE40
KDM3A
P4HA2
ANGPTL4
ALDOA
HK2
HILPDA
PFKFB4
MXI1
0
5
10
15
20
25
Frequency

b)

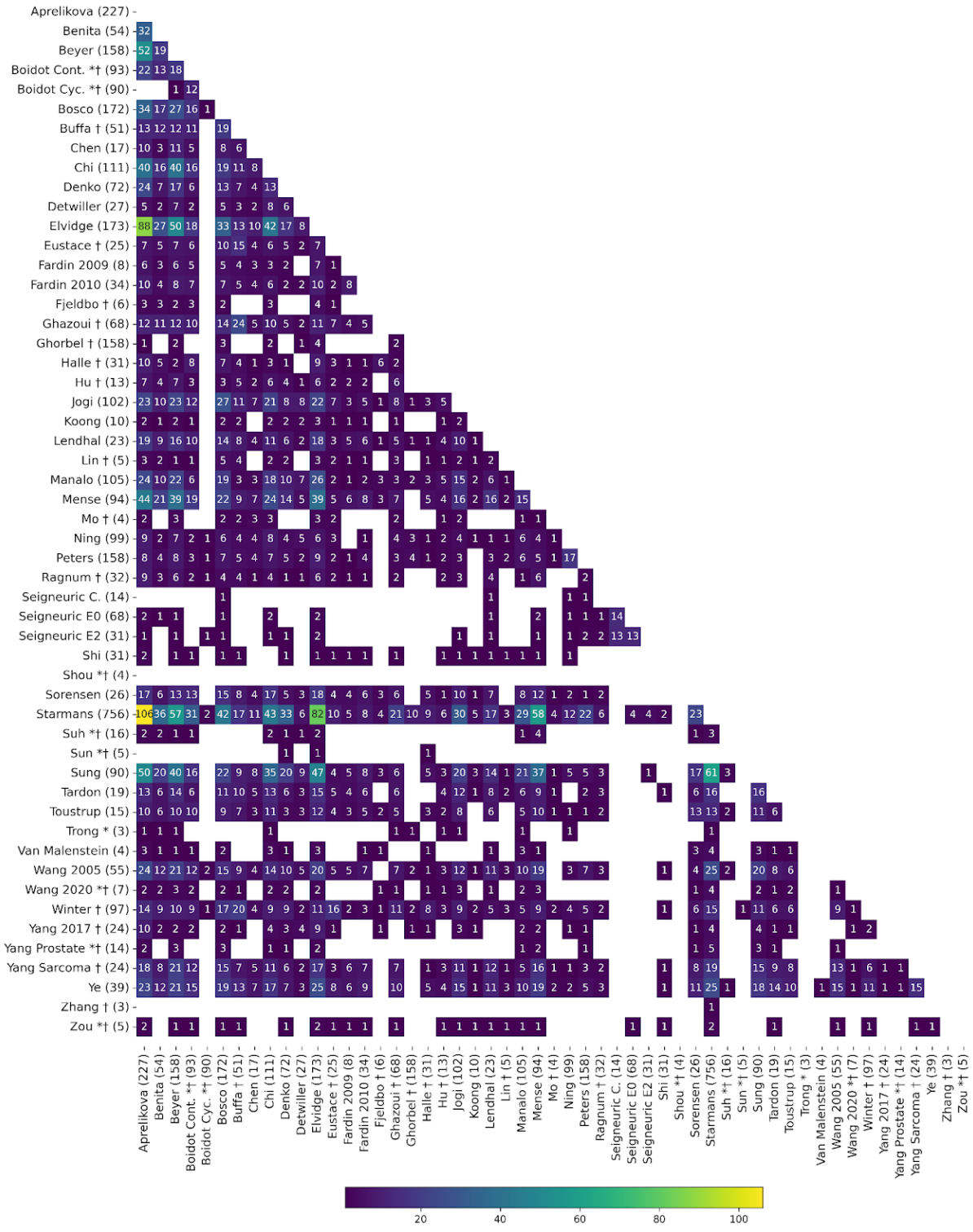

Aprelikova (227)
Benita (54)
Beyer (158)
Boidot Cont. *† (93)
Boidot Cyc. *† (90)
Bosco (172)
Buffa † (51)
Chen (17)
Chi (111)
Denko (72)
Detwiller (27)
Elvidge (173)
Eustace † (25)
Fardin 2009 (8)
Fardin 2010 (34)
Fjeldbo † (6)
Ghazoui † (68)
Ghorbel † (158)
Halle † (31)
Hu † (13)
Jogi (102)
Koong (10)
Lendhal (23)
Lin † (5)
Manalo (105)
Mense (94)
Mo † (4)
Ning (99)
Peters (158)
Ragnum † (32)
Seigneuric C. (14)
Seigneuric E0 (68)
Seigneuric E2 (31)
Shi (31)
Shou *† (4)
Sorensen (26)
Starmans (756)
Suh *† (16)
Sun *† (5)
Sung (90)
Tardon (19)
Toustrup (15)
Trong * (3)
Van Malenstein (4)
Wang 2005 (55)
Wang 2020 *† (7)
Winter † (97)
Yang 2017 † (24)
Yang Prostate *† (14)
Yang Sarcoma † (24)
Ye (39)
Zhang † (3)
Zou *† (5)
20
40
60
80
100

Figure 3.1: *a) Frequency of 20 most common genes amongst the 53 published hypoxia signatures, as identified in Table 3.2 (left hand column). b) Number of genes overlapping across all the signatures analysed in the study.*

The commonly found genes across signatures (a) are arranged in descending order from the most commonly found at the top (VEGFA) to the 20th most common (HK2). Horizontal scale indicates the number of signatures where those genes are found. The number of overlapping genes across all signatures included in the study is shown in b). Both x- and y-axis are the signatures and the total number of genes in the signatures are shown in parenthesis. Masked squares (plain white) mean there is no overlap between the two signatures.

This analysis reveals other genes not identified previously as common to hypoxia signatures. Within the top 20 most frequently occuring in the analysis, these are HILPDA, FAM162A, BHLHE40, KDM3A, ERO1A (although reannoted from ERO1L), MXI1 and HK2. Perhaps the most intriguing of these is HILPDA (Hypoxia-Inducible Lipid Droplet Associated protein) which has been shown to promote cancer progression through hypoxia dependent and independent pathways[119] but recently has been linked to evasion of anti-tumour immunity. Indeed, tumour infiltrating macrophages (TAMs) infiltration has been found to be significantly increased in tissues with high HILPDA expression and HILPDA positively correlates with immunosuppressive genes, such as PD-L1[120]. This potential immunosuppressive effect of HILPDA may occur through lipid accumulation in macrophages as has been seen in adipose tissue within mouse models of obesity[121]. This effect in part may help to give an explanation for poor prognosis associated with hypoxia when targeted with immunotherapy.

3.2 Hypoxia signatures in different technologies

Our scientific understanding increases as technology evolves. Platforms for measuring gene expression are no exception. The first technique applied historically, the polymerase chain reaction, is excellent at measuring the gene expression of a single target but can be labour intensive in multi-hypothesis/hypothesis-free studies when reporting the whole transcriptome is the aim, as required for 'gene signatures'. In the late 1990s and early 2000s, microarrays made measurement of the transcriptome vastly easier. Microarrays comprise a set of sequences arranged in a grid pattern representing the entire genome or a set of genes of interest. As in the polymerase chain reaction, measurement in these arrays involves first converting mRNA to cDNA to improve stability. The DNA molecules are then cut into fragments by restriction endonucleases and fluorescent markers attached (e.g. Cyanine 3 [Cy3]). The fluorescent labelled DNA fragments then bind to their complementary sequences on the microarray chip. Fluorescence emission is measured by passing a laser over the chip, with the higher emission indicating the greater abundance of cDNA and thus initial mRNA[122]. Initial hypoxia signatures were identified using this technology.

However, within this approach, results have been seen to differ between competing manufacturers (e.g. Affymetrix and Illumina) and as biological understanding evolved, different platforms have adapted at different rates. This makes interpretation of signatures potentially complex as many papers report the gene symbols annotated from the array and not the initial probeset IDs. Indeed, our group suggested a minimum of three annotations should be submitted for all microarray studies to allow accurate re-annotation (probeset ID, ensembl ID and gene symbol)[13].

The next major step in mRNA measuring technology comes from RNA sequencing (RNASeq). As with microarray technology, mRNA is converted to cDNA but the cDNA is then sequenced and mapped against a reference genome. On top of providing gene expression levels, RNASeq provides an additional level of detail by measuring alternative levels of

splice variants and non-coding RNAs. This high-throughput sequencing method is expected to supersede microarrays.

The combination of different methodological approaches (microarrays and RNASeq) and different platforms within each presents a challenge when interpreting signatures and combining datasets. To standardise signatures across different datasets, the HGNC Multi-symbol checker was used to update symbols from all the published hypoxia signatures to the approved most recent alias (see Methods). This approach allowed consistency in gene ontology.

Even with this approach, certain genes from published signatures were missing from (and hence undetectable in) certain platforms and experiments. This is illustrated in *Fig. 3.2* and *Fig. 3.3* where the percentage of missing genes in each signature (signatures on y axis) across the hypoxia experimental datasets found in GEO (x axis) is shown. In *Fig. 3.2*, one can see that in GPL3423, an Affymetrix custom microarray platform, a large percentage of genes expected in the signatures are missing in a large proportion of signatures, some even with 100% missing (Fjeldbo and Zhang). But even in a non-custom Affymetrix platform (GPL16686) a large percentage of genes are missing (e.g. 77% in Fardin 2010 and 75% in Mo). The challenge of missing genes also presents itself in some RNASeq datasets (*Fig. 3.3*). For example, GPL16791 (Illumina HiSeq 2500) is used in the GEO series GSE109367 and also four other series (Table 3.3). However, even with the same technology, the percentage of missed genes in GSE109367 is much higher (see *Fig. 3.3*); this is likely because of different approaches in downstream analysis.

Therefore, to be able to interpret signature performance amongst different technologies and different experiments, platforms (bottom axis) from the subsequent analysis where ≥12 out of 53 hypoxia signatures (≥20%) had ≥20% of genes that are not detectable on that specific platform ID were removed. This excluded four microarray platform IDs (GPL91, GPL17692, GPL16686 and GPL3423 - right hand side of *Fig. 3.2*) and two RNASeq experiments (GSE1419417 using GPL23227 and GSE109367 using GPL16791 -

right hand side of *Fig. 3.3*) from the analysis. *Table 3.3* also highlights the diversity of platforms used in hypoxia experiments on GEO.

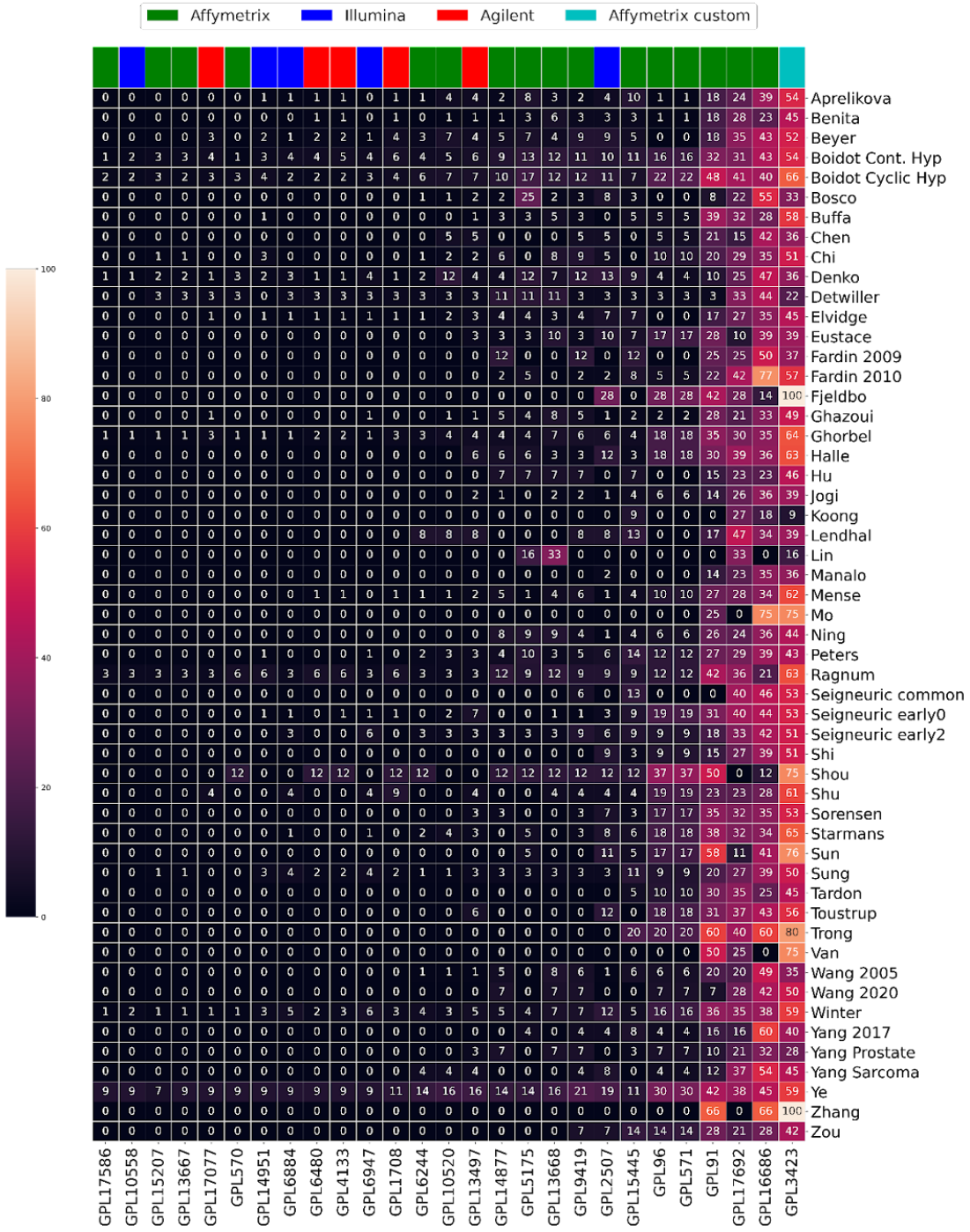


Figure 3.2: *Percentage of genes missing in each signature in hypoxia-related microarray experiments*

Summary of percentage missing genes in each signature for all the microarray platforms in GEO hypoxia experiments. Top row denotes type of manufacturer and if a custom platform.

Signature names are found on the y-axis and GEO platform IDs are on the x-axis. The lighter the colour of the square, the higher the percentage of missing genes for that signature in the platform in question.

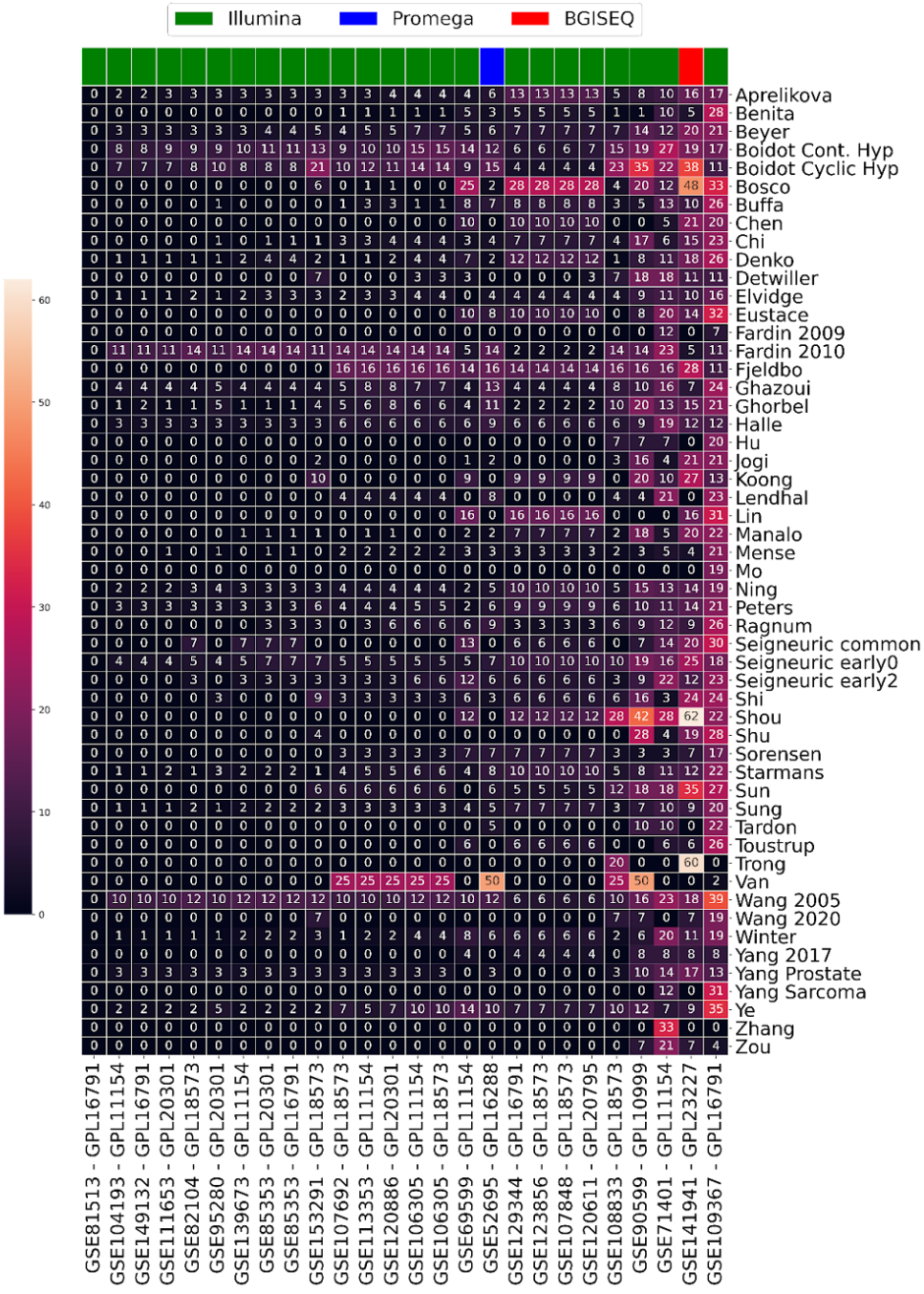


Figure 3.3: *Percentage of genes missing in each signature in hypoxia-related RNASeq experiments*

Summary of percentage missing genes in each signature for all the RNASeq platforms in GEO hypoxia experiments. Top row denotes the type of manufacturer. Signature names are found on the y-axis and GEO platform (GPL) and series (GSE) IDs are on the x-axis. The

lighter the colour of the square, the higher the percentage of missing genes for that signature in the platform and series in question.

Company	Technology type	Platform	Platform ID	Number of GEO series	GEO Series IDs
Affymetrix	Microarray	[HG_U95A] Affymetrix Human Genome U95A Array	GPL91	1	GSE1056
Affymetrix	Microarray	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	10	GSE3188; GSE53012; GSE39042; GSE9350; GSE17353; GSE58049; GSE3045; GSE3051; GSE35973; GSE4086
Affymetrix	Microarray	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array [cdf: HGU133Plus2_Hs_ENTREZG.cdf]	GPL13668	1	GSE29641
Affymetrix	Microarray	[HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array	GPL571	2	GSE29406; GSE19197
Affymetrix	Microarray	[HG-U133A] Affymetrix Human Genome U133A Array	GPL96	1	GSE3188
Affymetrix	Microarray	[HG-U219] Affymetrix Human Genome U219 Array	GPL13667	1	GSE60729
Affymetrix	Microarray	[HTA-2_0] Affymetrix Human Transcriptome Array 2.0 [transcript (gene) version]	GPL17586	2	GSE117775; GSE147516
Affymetrix	Microarray	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [HuEx-1_0-st-v2,mainR3,A20071112,EP.cdf]	GPL10520	1	GSE57613
Affymetrix	Microarray	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version]	GPL5175	3	GSE66894; GSE37340; GSE27523
Affymetrix	Microarray	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	GPL6244	8	GSE38061; GSE30979; GSE42791; GSE71280; GSE124524; GSE42416; GSE111259; GSE65168
Affymetrix	Microarray	[HuGene-2_0-st] Affymetrix Human Gene 2.0 ST Array [transcript (gene) version]	GPL16686	1	GSE99766
Affymetrix	Microarray	[HuGene-2_1-st] Affymetrix Human Gene 2.1 ST Array [transcript (gene) version]	GPL17692	1	GSE81416
Affymetrix	Microarray	[PrimeView] Affymetrix Human Gene Expression Array	GPL15207	1	GSE67549
Affymetrix	Microarray	Affymetrix GeneChip Human Genome U133 Plus 2.0 Array [CDF: Hs133P_Hs_REFSEQ_10]	GPL9419	1	GSE18494
Affymetrix	Microarray	Affymetrix Human Genome U133 Plus 2.0 Array [Brainarray Version 13, HGU133Plus2_Hs_ENTREZG]	GPL14877	1	GSE41491
Affymetrix	Microarray	Affymetrix Human Genome U133	GPL15445	1	GSE75101

		Plus 2.0 Array [CDF:gahgu133plus2_2.2.0]			
Affymetrix Custom	Microarray	Stanford Denko EOS Human 35K Genechip v1.1	GPL3423	1	GSE4186
Agilent	Microarray	Agilent-012391 Whole Human Genome Oligo Microarray G4112A (Feature Number version)	GPL1708	1	GSE33438
Agilent	Microarray	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Feature Number version)	GPL4133	2	GSE55563
Agilent	Microarray	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name version)	GPL6480	1	GSE33521
Agilent	Microarray	Agilent-026652 Whole Human Genome Microarray 4x44K v2 (Probe Name version)	GPL13497	1	GSE78245
Agilent	Microarray	Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 (Probe Name version)	GPL17077	2	GSE117036; GSE117041
Illumina	Microarray	Illumina HumanHT-12 V3.0 expression beadchip	GPL6947	2	GSE45301; GSE15530
Illumina	Microarray	Illumina HumanHT-12 V4.0 expression beadchip	GPL10558	13	GSE72723; GSE63562; GSE59729; GSE80657; GSE89891; GSE41666; GSE125177; GSE111246; GSE118683; GSE107300; GSE42868; GSE45362; GSE147384
Illumina	Microarray	Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip	GPL14951	2	GSE55211; GSE55212
Illumina	Microarray	Illumina HumanWG-6 v3.0 expression beadchip	GPL6884	6	GSE7272; GSE36562; GSE55212; GSE61799; GSE30019; GSE47533
Illumina	Microarray	Sentrix Human-6 Expression BeadChip	GPL2507	1	GSE3188
BGISEQ	RNASeq	BGISEQ-500 (Homo sapiens)	GPL23227	1	GSE141941
Illumina	RNASeq	HiSeq X Ten (Homo sapiens)	GPL20795	1	GSE120611
Illumina	RNASeq	Illumina Genome Analyzer IIx (Homo sapiens)	GPL10999	1	GSE90599
Illumina	RNASeq	Illumina HiSeq 2000 (Homo sapiens)	GPL11154	6	GSE104193; GSE106305; GSE69599; GSE113353; GSE71401; GSE139673
Illumina	RNASeq	Illumina HiSeq 2500 (Homo sapiens)	GPL16791	5	GSE109367; GSE81513; GSE129344; GSE85353; GSE149132
Illumina	RNASeq	Illumina HiSeq 4000 (Homo sapiens)	GPL20301	5	GSE120886; GSE111653; GSE85353; GSE149132; GSE95280

Illumina	RNASeq	Illumina NextSeq 500 (Homo sapiens)	GPL18573	7	GSE106305; GSE108833; GSE107848; GSE107692; GSE82104; GSE123856; GSE153291
Promega	RNASeq	AB 5500xl Genetic Analyzer (Homo sapiens)	GPL16288	1	GSE52695

Table 3.3: *Platforms used in hypoxia gene expression experiments in the Gene Expression Omnibus (GEO)*

The frequency platforms and technologies used in hypoxia-related experiments in GEO. Light pink highlights microarrays and purple RNASeq. The most common platform used was the Illumina HumanHT-12 V4.0 expression beadchip in 13 GEO series.

3.3 Summarising hypoxia signatures into scores

Upon identifying the 53 published hypoxia signatures, converting them into approved symbols, and identifying useful datasets, the study moved to evaluate the most effective methods for summarising these signatures into a convenient score that could potentially be used to denote hypoxia by scientists and clinicians alike.. Several approaches could be used. From the literature, six promising ways to convert the signatures into scores were identified.

- *Mean:* sum the expression values for each of the genes in the signature and divide by the number of genes in the signature.
- *Median:* arrange all gene expression values in numerical order from lowest to highest. The median value separates the upper half of the data values from the lower half.
- *PCA1:* PCA1 is a value derived from Principal Component Analysis (PCA). PCA is a common dimensionality-reduction method used in bioinformatics. It allows us to reduce the dimensionality of large datasets, such as gene-expression data, by summarising all this information into one or multiple scores. PCA1 is the first principal component derived through this method and is the score that represents the largest amount of variance given by genes in the signature within the dataset.
- *GSVA*[45]*:* GSVA is a type of Gene Set Enrichment (GSE) analysis which estimates variation of pathway activity in an unsupervised manner. The initial step of GSVA is to evaluate the expression of a gene (high vs low) within the dataset whilst normalising each expression value to a common scale. The classical maximum deviation method then provides a score that penalises deviations that are large in both tails, thus emphasising genes that are concordantly activated in one direction only. However, this comes at the

theoretical disadvantage that if genes in the same pathway are expressed in opposing directions (some over, some under expressed) then the score will be minimal.

- *PLAGE*[80]*:* PLAGE is an unsupervised, single sample enrichment method based on PCA. The expression of each pathway gene set is standardised over the samples to give estimates of each pathway activity. These estimates are then used as coefficients of the first right-singular vector of the singular value decomposition of the gene set.
- *ssGSEA*[123]*:* ssGSEA is an extension of the Gene Set Enrichment Analysis (GSEA) which can be applied to individual samples rather than a whole dataset.

Mean, median, PC1, GSVA, PLAGE and ssGSEA scores of hypoxia signatures should change with hypoxia, and therefore, one could perhaps use these as an indicator of cell line and tissue hypoxia. However, this has never been systematically tested. To investigate this, the six scores of a pioneering signature from our laboratory (Buffa 2010)[71] were calculated on a publicly available dataset of MCF-7 cells (GSE153291, ER+ breast cancer cells) from Wu et al.[124]. In this experiment, MCF-7 breast cancer cells were incubated either in normoxic conditions or subjected to hypoxia (1% oxygen) for 24 hours. There were three normoxic (GSM4639878, GSM4639879, GSM4639880) and three hypoxic (GSM4639881, GSM4639882, GSM4639883) replicates. Here, the performance of the Buffa signature in correctly identifying hypoxic cells was examined, using scores from each of the six aforementioned approaches. The performance of the signature was finally evaluated using the signature assessment package sigQC, which calculated the six scores in question[81].

3.4 Buffa signature summary scores applied to dataset GSE153291

The median score of the gene expression values has been proposed to indicate the degree of hypoxia. This analysis examines the efficacy of not just the median, but also the mean, PCA1, GSVA, PLAGE, and ssGSEA scores of the Buffa signature in identifying hypoxic MCF-7 cells in a laboratory experiment. The performance of the Buffa signature in this dataset was also evaluated using sigQC[81].

Laboratory data comes from the freely accessible Gene Expression Omnibus (GEO) portal. Dataset GSE153291 was generated as part of a study looking at long intergenic non-coding RNAs (lincRNAs) in hypoxia, specifically lincNORS which appears to be a non-coding regulator of the cellular sterol homeostasis[124]. To assess hypoxia signature efficacy is a novel use of this dataset, and highlights the advantage of open source dataset sharing approaches such as GEO. Hypoxia-regulated genes (coding and long noncoding) were identified by transcriptomic profiling of MCF-7 cells under hypoxia (1% O2, 24 hrs) versus normoxia (21% O2, 24 hrs) using RNASeq (Illumina NextSeq 500). There were three plates of MCF-7 cells in hypoxia and three in normoxia.

The six hypoxia scores (median, mean, PCA1, GSVA, PLAGE and ssGSEA) for the Buffa signature were calculated for the six samples in GSE153291 and, these are summarised for hypoxic and normoxic samples in *Table 3.4*. As expected, the mean and median scores using the Buffa signature in the hypoxic samples were markedly different to the normoxic samples. This can be compared with the scores derived from comparing a random gene set in hypoxia and normoxia used as a control, where there was no consistent difference (*Table 3.5*). The same pattern was seen with ssGSEA.

Clear differences were also seen in PC1 scores, where the hypoxic samples gave negative values whereas for normoxic they were all positive, and using PLAGE, where the

sign difference was reversed. In contrast, GSVA scores gave a mixed picture with one of the hypoxic samples (GSM4639881) giving a similar score to the normoxic samples, hinting that the GSVA score might not be the best discriminator here. This is formally tested later.

On assessing the performance of the Buffa signature itself, some reassuring features in the summary-scoring metrics and correlations when comparing against a random gene set of the same length were found (*Fig. 3.4 and Fig. 3.5*). Firstly, the correlation matrices (*Fig. 3.4a* and *3.5a*) from the Buffa signature show strong positive correlations between mean, median, ssGSEA and GSVA scores, which is expected if the signature is representative of a common feature, such as tumour hypoxia. This is not observed in the random gene set. A radar plot (*Fig. 3.4b* and *3.5b*) is useful to assess overall signature quality, combining a number of key metrics of quality and interpretability fully explained in Dharwan et al[81]. A good summary measure, as described by Dhawan et al., can be derived from the area of the radar plot[81]; the greater area inside the red line, the better performing the signature within the given dataset. The area inside the red line for the radar plot derived using the Buffa signature (*Fig. 3.4b*) is clearly much larger than that using the random signature (*Fig. 3.5b*).

The scatterplots in Panel C describe the correlation of all tested scores (mean, median, PCA1, ssGSEA, GSVA, and PLAGE) with dots representing the individual samples. Using the Buffa signature, the samples cluster to different ends of the plots, whereas the random gene set the separation is much less marked, indicating mixing of hypoxic and normoxic samples (*Fig. 3.4c* and *3.5c*). The PCA1 score derived using the Buffa signature appears to be describing much more of the variance of the dataset than using the random gene signature (*Fig. 3.4d* and *3.5d*), suggesting that the genes within the Buffa signature are more relevant to determining the feature in question (hypoxic state of cell lines).

After evaluating their performance using sigQC[81], the study transitioned to a novel approach aimed at assessing the reliability of these scores in differentiating between hypoxic and normoxic cells within this dataset. This approach compares the significance of hypoxia gene signatures versus random gene set stimulations (for further detail see Chapter 2.5).

Sample	Mean	Median	PCA1	GSVA	ssGSEA	PLAGE
GSM4639881	229.299	9.549	-1989.220	-0.123	7.388	0.428
GSM4639882	236.586	10.040	-2437.146	0.044	7.422	0.427
GSM4639883	269.736	11.202	-2952.792	0.477	7.520	0.361
GSM4639878	88.722	6.890	2560.898	-0.434	6.542	-0.357
GSM4639879	90.358	7.245	2437.901	-0.625	6.520	-0.381
GSM4639880	98.164	7.819	2380.358	-0.010	6.593	-0.479

Table 3.4: *Scores derived using the Buffa hypoxia signature in hypoxic and normoxic MCF7 cells (GSE153291)*

Hypoxia scores using six scoring methods and the Buffa hypoxia signature. Red rows denote hypoxic samples whereas blue rows denote normoxic samples.

Sample	Mean	Median	PCA1	GSVA	ssGSEA	PLAGE
GSM4639881	2.555	0.155	-19.139	0.045	0.589	0.514
GSM4639882	2.494	0.191	-9.64	0.066	0.519	0.443
GSM4639883	2.06	0.336	20.891	0.174	0.942	0.238
GSM4639878	3.202	0.167	-14.089	-0.146	0.182	-0.375
GSM4639879	3.174	0.246	-4.539	0.121	0.861	-0.453
GSM4639880	2.573	0.292	26.517	-0.140	-0.057	-0.367

Table 3.5: Scores given from a *random gene set same length as Buffa hypoxia signature in hypoxic and normoxic MCF7 cells (GSE153291)*

Hypoxia scores using six scoring methods and a random gene set the same length as the Buffa hypoxia signature. Red rows denote hypoxic samples whereas blue rows denote normoxic samples.

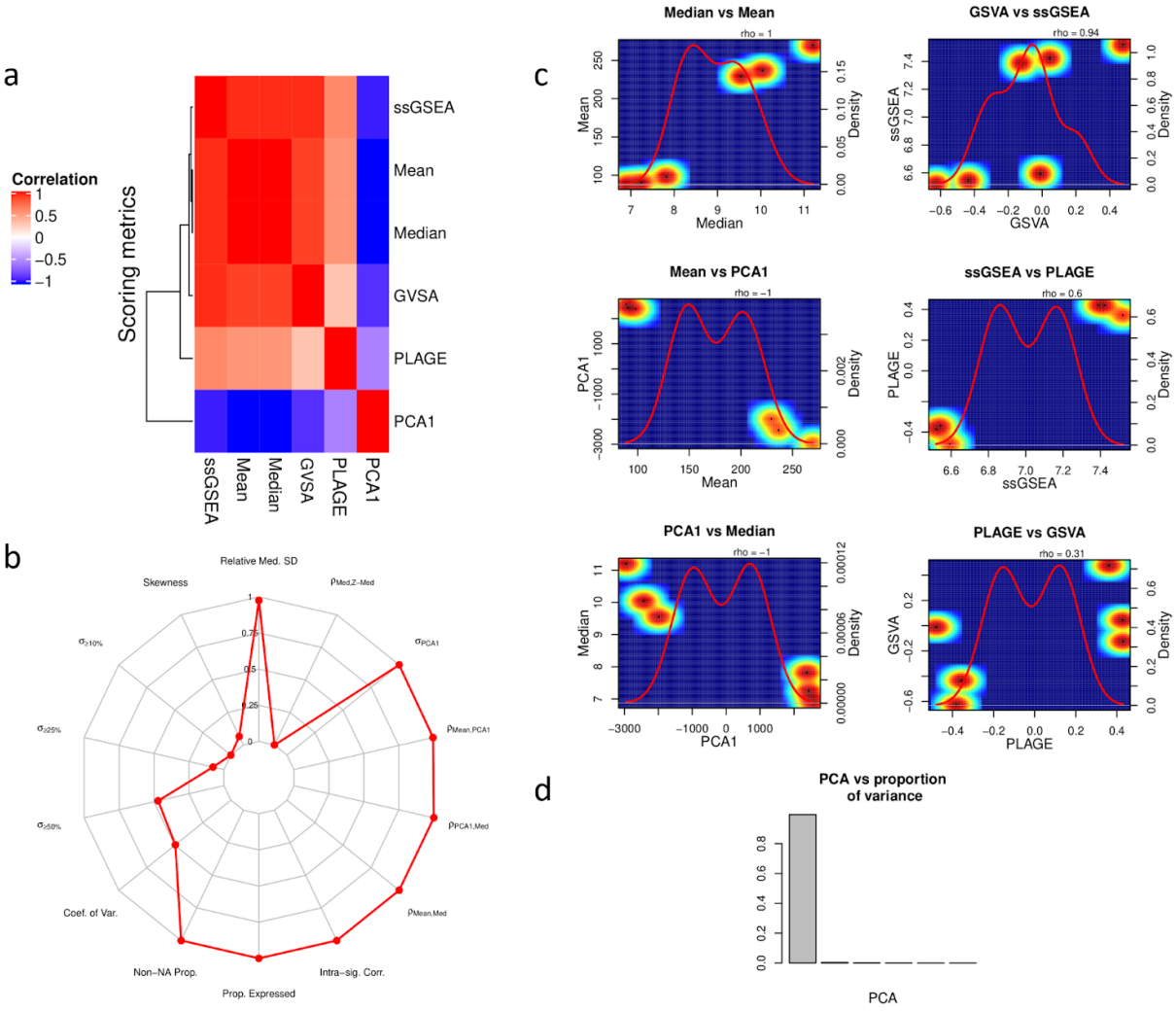


Figure 3.4: *The performance of the Buffa hypoxia signature in GSE153291*

Panel a) depicts the correlation of the scores in the individual samples with the other scores. The samples on the diagonal show bright red as there is a perfect correlation (e.g. GSVA vs. GSVA, correlation coefficient of 1). Panel b) shows the performance of the signature based on several measures described fully in sigQC. Notably, the proportion of expressed genes in the signature is very high and there is a high level of non-NA values in the signature genes. The absolute correlation coefficient of mean:median of signature genes, PCA1:median and PCA1:mean are high and the intra-signature correlation is high suggesting it is compact and functioning well. Panel c) shows the intercorrelation of all tested scores (mean, median,

PCA1, ssGSEA, GSVA, and PLAGE). The solid red lines represent the distribution of the samples in the dataset and are related with the y-axis named density; the solid grey line represents the horizontal axis beneath density plots. The coloured dots represent the individual samples. Panel d) highlights that PCA1 (first bar) accounts for well over 80% of the variance within the dataset.

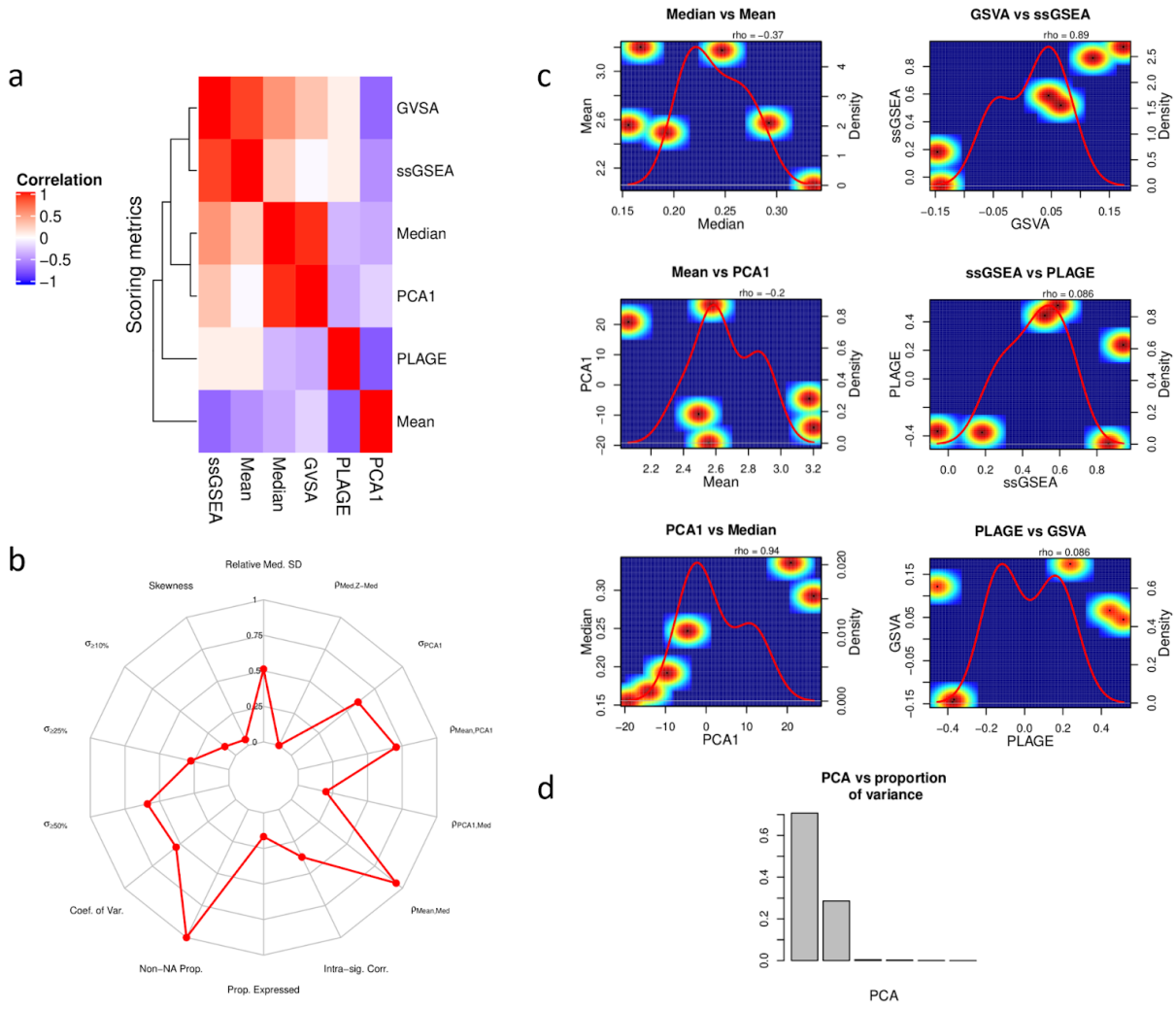


Figure 3.5: *The performance of a single random set of genes of same length as the Buffa Hypoxia signature (51 genes) in GSE153291*

Panel a) depicts the correlation of the scores in the individual samples with the other scores. The samples on the diagonal show bright red as there is a perfect correlation (e.g. GSVA vs. GSVA, correlation coefficient of 1). Panel b) shows the performance of the signature based on several measures described fully in sigQC. The non-NA proportion is high but other important measures such as proportion of expressed gene and intra-signature correlation are both low. Panel c) shows the intercorrelation of all tested scores (mean, median, PCA1, ssGSEA, GSVA, and PLAGE). The solid red lines represent the distribution of the samples in

the dataset and are related with the y-axis named density; the solid grey line represents the horizontal axis beneath density plots. The coloured dots represent the individual samples. Panel d) highlights that PCA1 (first bar) accounts for ~70% of the variance within the dataset.

The new approach investigates the performance of the different scores using the Buffa signature compared to RGS of the same length. This is done by calculating the Euclidean distance between hypoxic and normoxic sample scores in both hypoxia signatures and RGS, and comparing their distribution (see Chapter 2.5). By analysing the distance of the Buffa signature vs simulations of 500 random gene sets it is possible to get an idea of its performance. However, only mean, median, GVSA and ssGSEA can be used in this approach, as discussed in Chapter 2.5. *Fig. 3.6* shows the performance of the Buffa signature scores in differentiating between the first pair of normoxic (GSM4639878) and hypoxic (GSM4639881) samples. Here the median and scores for the Buffa signature lie some distance from the peak of the frequency distribution, indicating that they clearly perform differently to the random gene sets. In contrast, using the GSVA and ssGSEA, the Buffa signature score is found within the body of the distribution indicating that the GSVA and ssGSEA scores perform similarly to random gene sets.

Fig 3.7 shows the performance of the Buffa signature scores in correctly identifying hypoxic vs normoxic samples within GSE153191 (p value calculation and explanation described in Chapter 2.5). The median score gives the most significant difference in the discrimination compared to RGS (all p values <0.005). In contrast, the ssGSEA score for the Buffa hypoxic signature did not perform significantly differently than scores from the RGS in correctly classifying hypoxic samples, and GSVA outperformed the RGS in only three of the nine pairwise combinations. These results suggest that the median and mean Buffa signature hypoxia scores are “best” performing in this dataset, whereas ssGSEA is not a good classifier of hypoxic state. This example highlights the importance of which summary score is chosen for hypoxia signatures. It also raises the related question of whether all hypoxia signatures should be summarised using the same score, or whether the ‘best’ score may vary between signatures or systems (e.g. perhaps ssGSEA is a more appropriate score and performs better using a different hypoxia signature on this dataset, or on a dataset from different cell line). This is investigated further below.

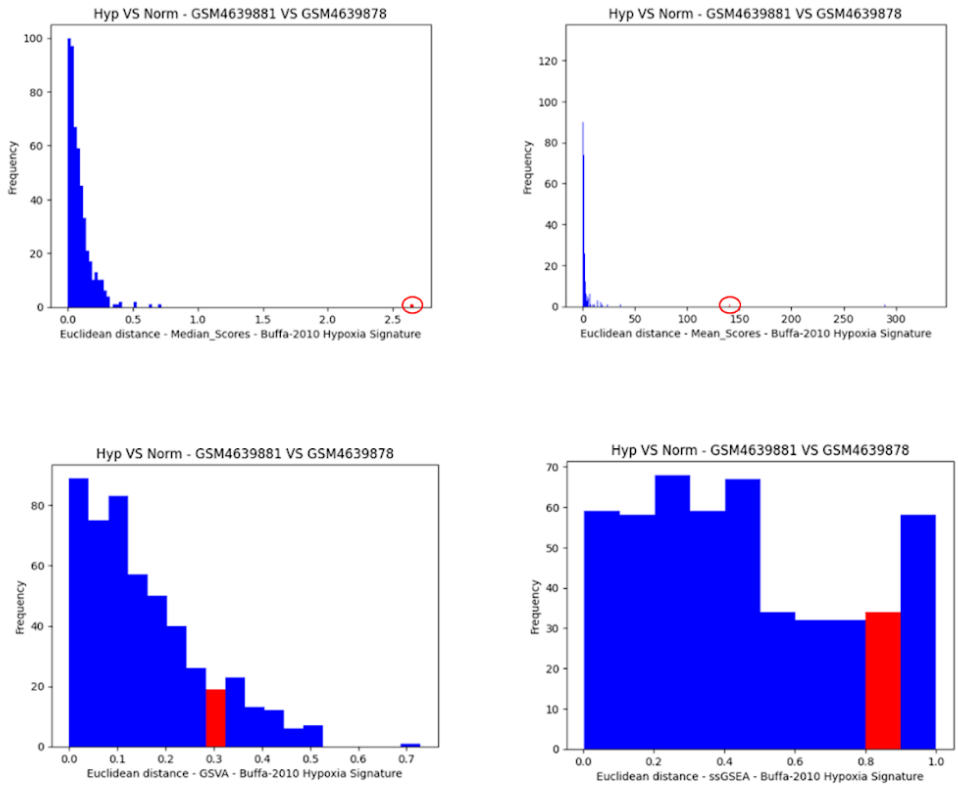


Figure 3.6: *Distribution of the distances calculated using the Buffa hypoxia signature in comparison to 500 RGS of the same length in GSE153291 for a pairwise combination of one hypoxic and one normoxic sample (GSM4639881 vs. GSM4639879)*

The y-axis represents the number of gene sets falling into a given quantile of the distribution whereas the x-axis represents the Euclidean distance calculated between the two samples (GSM4639881 vs. GSM4639878). Blue bars only contain RGS distances, red bars denote the quantile where the original hypoxia signature distance belongs to. If the red bar is not in the tail of the distribution, the distance between the pair of samples derived from the original signature performs similarly to RGS. Alternatively if the red bar is in the tail of the distribution (e.g. last quantile on the right side of the distribution, as in the two upper plots), the original signature is outperforming all RGS, maximising the distance observed between the pair.

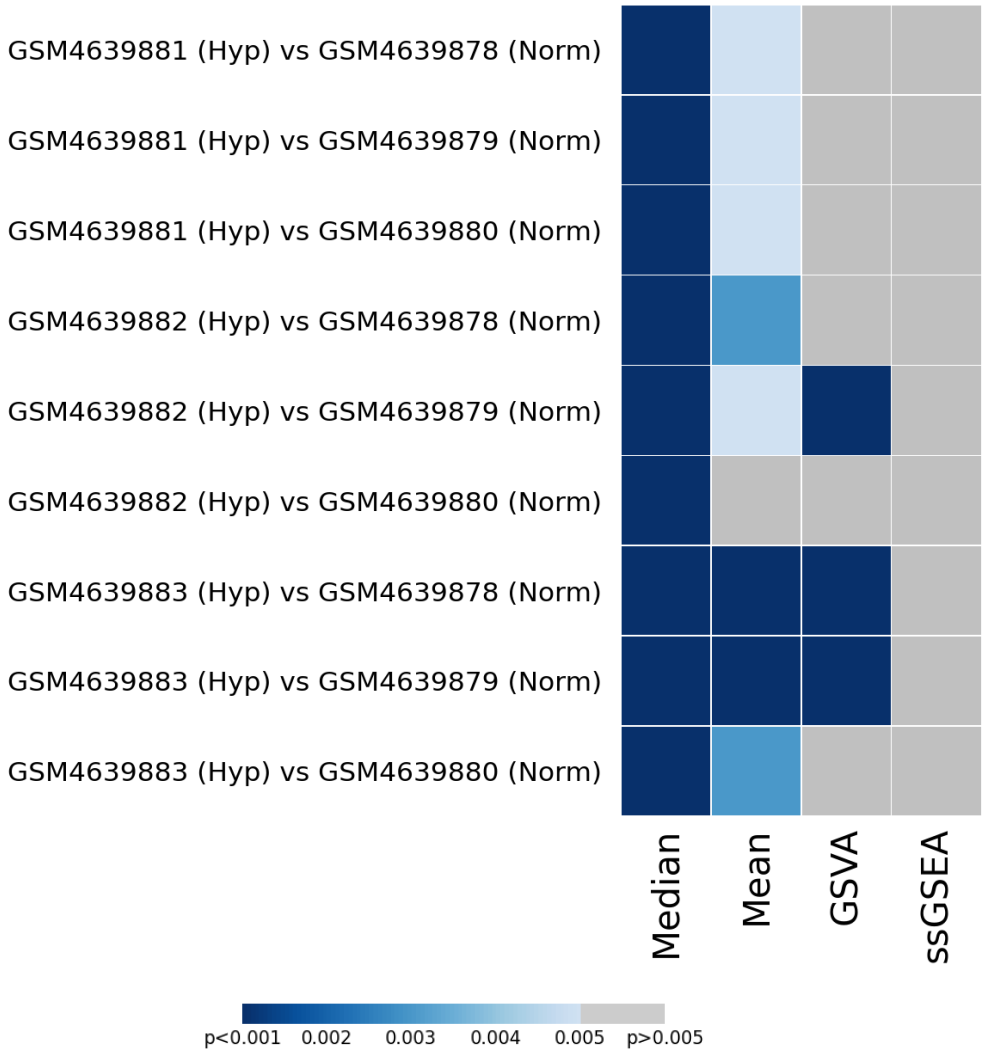


Figure 3.7: *The performance of different scores summarising the Buffa Hypoxia signature in comparison to 500 random gene sets of the same length in GSE153291*

Each square represents a pairwise comparison of scores between a hypoxic and normoxic sample, using the Buffa hypoxic signature. The intensity of colour indicates the significance level (p statistic) of the difference in scores (i.e. the ability of the signature to distinguish between hypoxic and normoxic samples). Comparison of medians allows distinction with a p value of <0.005. In contrast, ssGSEA score would not allow the two states to be distinguished in the systems used here (p>0.005).

3.5 Hypoxia signature performance in GSE153291

Following the assessment of the Buffa 2010 signature in the GSE153291 dataset using the four summary scores (above), the study moved to a systematic evaluation of all 53 published signatures using the four summary scores within GSE153291. These 212 combinations of signatures and scores were evaluated in two main ways: a) assessing performance in relation to RGS of the same length and b) assessing accuracy at correctly determining hypoxic samples. Both components are discussed in detail in Chapter 2 and comparison with RGS is outlined through an example in Chapter 3.4. In brief, classification accuracy was calculated by assessing performance of each signature against RGS. The number of pairwise combinations performing significantly differently to random gene sets ($p < 0.005$) was calculated and then accuracy was defined by dividing by the total number of pairwise combinations in the analysis. For instance, if the Buffa signature using the median score significantly differed to random gene set simulations ($p < 0.005$) in 15 of the 16 pairwise combinations, classification accuracy would be 93.75% ($15 \div 16 \times 100$).

Fig. 3.7 demonstrates the performance of all signatures within GSE153291. In this dataset, depending on which scoring method was used, there were large differences in signature performance (*Fig. 3.8* and *Table 3.6*). One marked difference in performance comes depending on which score is used for the comparison - as indicated above for the Buffa signature. Using the Boidot continuous hypoxia signature, the median score also correctly classifies the samples significantly better than the RGS (100% accuracy) whereas the ssGSEA score does not differ significantly from the RGS across all pairwise combinations (0% accuracy). In general, the median score performed the best at identifying hypoxia over the greatest number of signatures (*Fig. 3.8*, top panel), followed by GSVA, Mean and ssGSEA.

Comparing the validity of individual signatures, a number had high accuracy (discriminatory ability) in a number of different scores. Indeed, 25 signatures, including many

derived from cell lines other than breast tumours (e.g Halle, which originated from cervical cancer cell lines), reached 100% accuracy in identifying hypoxic samples using the median scores (*Table 3.6*). That several signatures give 100% accuracy in this dataset is expected (since they all purport to identify hypoxia) and reassuring given the relatively small number of hypoxic and normoxic samples (three replicates of each and nine pairwise combinations of samples). However, this analysis also highlights the importance of choosing the correct scoring system, with other scores having much lower accuracy (e.g. ssGSEA, *Table 3.6*).

Twelve signatures (Fardin, Ghorbel, Ning, Seigneuric common, Seigneuric early0, Shi, Shou, Suh, Sun, Trong, Zhang and Zou) did not perform significantly differently to the random gene sets using any of the four summary scores across all pairwise combinations, suggesting these signatures are not a good indicator of hypoxic status in GSE153291. Ten of these twelve signatures were not derived from breast cell lines, which might explain their poor performance. The Seigneuric signatures were, however, deduced from studies on mammary epithelial cells. In this case, the signatures were taken from cells during early exposure to hypoxia (1, 3 and 6 hours), not 24 hours as in this dataset, which may point to some time dependence of the transcriptome. Nonetheless, the wide applicability of several signatures suggest a fair degree of conformity.

The main point to highlight here is the difference in performance between signatures and scores. The results established the optimum summary score for comparison (median) and the best performing signatures in this breast cancer cell line dataset. Will the median always be the best summary score to use using all freely available data in GEO? Will the signature(s) that are most indicative of hypoxia in GE153291 remain so in other cell lines/cancer types? The following chapters aim to answer these important questions, starting with individual tissue types and building towards the best signature across all conditions.

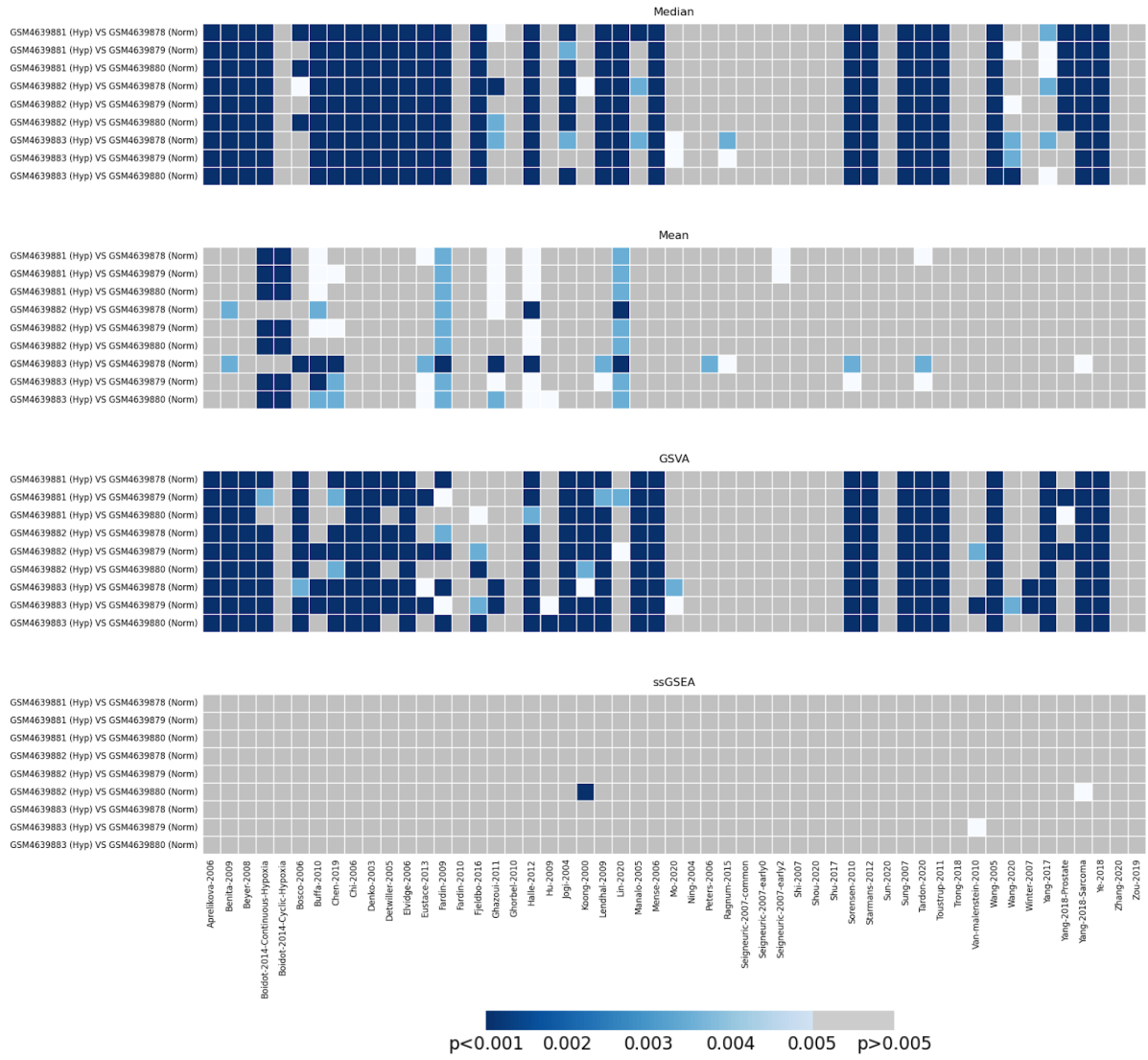


Figure 3.8: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in GSE153291.*

Comparison of the scoring methods of the 53 hypoxia signatures across one experiment (GSE153291) comparing gene expression in MCF-7 cells in normoxia (21% O2, 24 hrs) and hypoxia (1% O2, 24 hrs). The darker the shade of blue the more accurate the signatures and specific scores are at differentiating between hypoxic and non-hypoxic samples in GSE153291 comparing random gene sets of the same length. Grey means the score and signature did not significantly outperform random gene sets (p-value > 0.005). Median (top

panel), shows the best classification across the greatest number of experiments compared to mean, GSVA and ssGSEA.

	Median	Mean	GSVA	ssGSEA
Aprelikova	100.00	0.00	100.00	0.00
Benita	100.00	22.22	100.00	0.00
Beyer	100.00	0.00	100.00	0.00
Boidot Cont. *†	100.00	77.78	88.89	0.00
Boidot Cyc. *†	0.00	77.78	0.00	0.00
Bosco	44.44	11.11	100.00	0.00
Buffa †	100.00	88.89	33.33	0.00
Chen	100.00	55.56	88.89	0.00
Chi	100.00	0.00	100.00	0.00
Denko	100.00	0.00	100.00	0.00
Detwiller	100.00	0.00	66.67	0.00
Elvidge	100.00	0.00	100.00	0.00
Eustace †	100.00	44.44	44.44	0.00
Fardin 2009	100.00	100.00	77.78	0.00
Fardin 2010	0.00	0.00	0.00	0.00
Fjeldbo †	100.00	0.00	55.56	0.00
Ghazoui †	44.44	77.78	22.22	0.00
Ghorbel †	0.00	0.00	0.00	0.00
Halle †	100.00	100.00	100.00	0.00
Hu †	0.00	11.11	22.22	0.00
Jogi	88.89	0.00	100.00	0.00
Koong	11.11	0.00	100.00	11.11
Lendhal	100.00	22.22	100.00	0.00
Lin †	100.00	100.00	22.22	0.00
Manalo	33.33	0.00	100.00	0.00
Mense	100.00	0.00	100.00	0.00
Mo †	22.22	0.00	22.22	0.00
Ning	0.00	0.00	0.00	0.00
Peters	0.00	11.11	0.00	0.00
Ragnum †	22.22	11.11	0.00	0.00
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	0.00	0.00	0.00	0.00
Seigneuric E2	0.00	22.22	0.00	0.00
Shi	0.00	0.00	0.00	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	0.00	0.00	0.00	0.00
Sorensen	100.00	22.22	100.00	0.00
Starmans	100.00	0.00	100.00	0.00
Sun *†	0.00	0.00	0.00	0.00
Sung	100.00	0.00	100.00	0.00
Tardon	100.00	33.33	100.00	0.00
Toustrup	100.00	0.00	100.00	0.00
Trong *	0.00	0.00	0.00	0.00
Van Malenstein	0.00	0.00	22.22	11.11
Wang 2005	100.00	0.00	100.00	0.00
Wang 2020 *†	55.56	0.00	11.11	0.00
Winter †	0.00	0.00	22.22	0.00
Yang 2017 †	66.67	0.00	100.00	0.00
Yang Prostate *†	66.67	0.00	33.33	0.00
Yang Sarcoma †	100.00	11.11	100.00	11.11
Ye	100.00	0.00	100.00	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	0.00	0.00	0.00

Table 3.6: *Percentage accuracy of determining hypoxic samples from normoxic samples in GSE153291 for the 53 signatures across four hypoxia scores.*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue) within each score (column). Red cells highlight the highest accuracy in the whole table. The star symbol (*) denotes signatures made of both upregulated and

downregulated genes. The obelisk symbol (†) indicates signatures that have been derived including clinical samples.

Chapter 4: Evaluation of hypoxia gene expression signature scores across the Gene Expression Omnibus

Since HIF-1α dictates a key part of the cellular and transcriptional response to hypoxia in all tissues, it is possible that hypoxia signatures derived from individual tissues might apply across different tissue types. This provides a rationale as to why many studies have opted to just use a single hypoxia signature (and summary score) irrespective of tissue. As shown above, some hypoxia signatures do apply across several tissue types.

However, hypoxia signatures have been developed from different tissue types, using different technologies and approaches (e.g. clinical samples vs cell lines), it is possible that different signatures may indicate hypoxic status differently in various tissue types. Cancers in different tissues do display different receptor statuses (e.g. oestrogen receptor status in breast cancer), mutational profiles, and have different prognoses in the clinic, so one might expect different responses to the hypoxic state. It may also be the case that signatures claimed to identify hypoxia do not actually do so reliably, perhaps being derived from a too small sample size or due to some other experimental artefacts. For example, as shown above, seven out of the published 53 hypoxia signatures do not identify hypoxia in cultured mammary cells. Accordingly, the generality and efficacy of published hypoxia signatures in identifying hypoxia was investigated, not only by testing the signatures identified in one tissue type (e.g. simply looking at prostate cancer derived signatures in prostate cancer) but by looking universally in an unbiased fashion, e.g. testing whether signatures derived from prostate cancer were effective in indicating hypoxia across all cancer cell lines tested irrespective of their tissue of origin. Moreover, the analysis is not biassed towards a specific summary score in the prior application of the signatures, as all four scores are investigated throughout the study. As it is not correct to assume just because the median score of the Buffa signature gave 100% accuracy in the GSE153291 dataset does not necessarily ensure it will perform as effectively in other breast cancer cell line experiments and/or in different cancer types, perhaps it might be inferior to another score, like the mean. Thus, each tissue type (where sufficient samples exist, >30 pairwise combinations) will be discussed singularly,

before identifying the signatures and summary scores that perform best across all tissue and cancer types tested, in the search for an effective universal hypoxia signature.

4.1 Performance of hypoxia signatures in breast cancer cell lines

After showing marked differences in performance of the four summary scores of the 53 hypoxia signatures in one MCF-7 experiment (GSE153291), I was intrigued to see if similar results would be observed in other breast cancer cell line experiments. To investigate this, experiments involving the comparison of gene expression profiles from hypoxic and normoxic cells were first identified in the Gene Expression Omnibus (as described in Methods). From these, breast cancer cell lines were selected. This yielded the 34 cell lines shown in *Table 4.1*: comprising 19 triple negative (ER-, PR-, HER2-) lines, three ER-, PR-, HER2+ lines, one ER+, PR-, HER2- line, one ER-, PR+/-, HER2- line, four ER+, PR+, HER2- lines, one ER+, PR+, HER2+ line and five transformed “non-cancerous” mammary tissue cell lines. Although triple negative cell lines were the most frequently identified, ER+, PR+, HER2- had the highest number of pairwise combinations that could be investigated, 256 [mainly MCF-7] compared to 89 triple negative samples [mainly MDA-MB-231]).

I assessed the performance of the 53 published signatures at defining hypoxia in these 34 cell lines in 27 different datasets (GSE3188, GSE63562, GSE18494, GSE111653, GSE104193, GSE89891, GSE41491, GSE29641, GSE108833, GSE99766, GSE111246, GSE39042, GSE29406, GSE61799, GSE15530, GSE33438, GSE107692, GSE71401, GSE124524, GSE47533, GSE123856, GSE42416, GSE111259, GSE147516, GSE153291, GSE85353, GSE149132) to identify the most effective signature at identifying hypoxia across breast cancer cell lines. Here, ssGSEA was the worst performing score over all datasets tested (*Fig. 4.1*), with the best performing signatures in this score achieving only 1.97% accuracy (Chi, *Table 4.2*). The median score was the best performing across the signatures (*Fig. 4.1*, top panel), and applying this to Sorensen, the analysis was able to differentiate between hypoxic and normoxic samples with 97.75% accuracy (*Table 4.2*). Interestingly, Sorensen was derived from non-breast cell lines (cervical and head & neck).

The best performing signatures derived from breast cancer cell lines had median scores that had similar accuracy: Lendhal (not solely breast [see *Table 3.1*], 96.90%), Elvidge (MCF-7, 96.34%) and Aprelikova (MCF-7, 95.21%). This suggests that hypoxia signatures are worth investigating as markers of hypoxia independent of cell line of origin. For instance, one might not expect *a priori* to find the best performing signature in breast cancer cell lines to have been identified from cervical and head & neck cell lines.

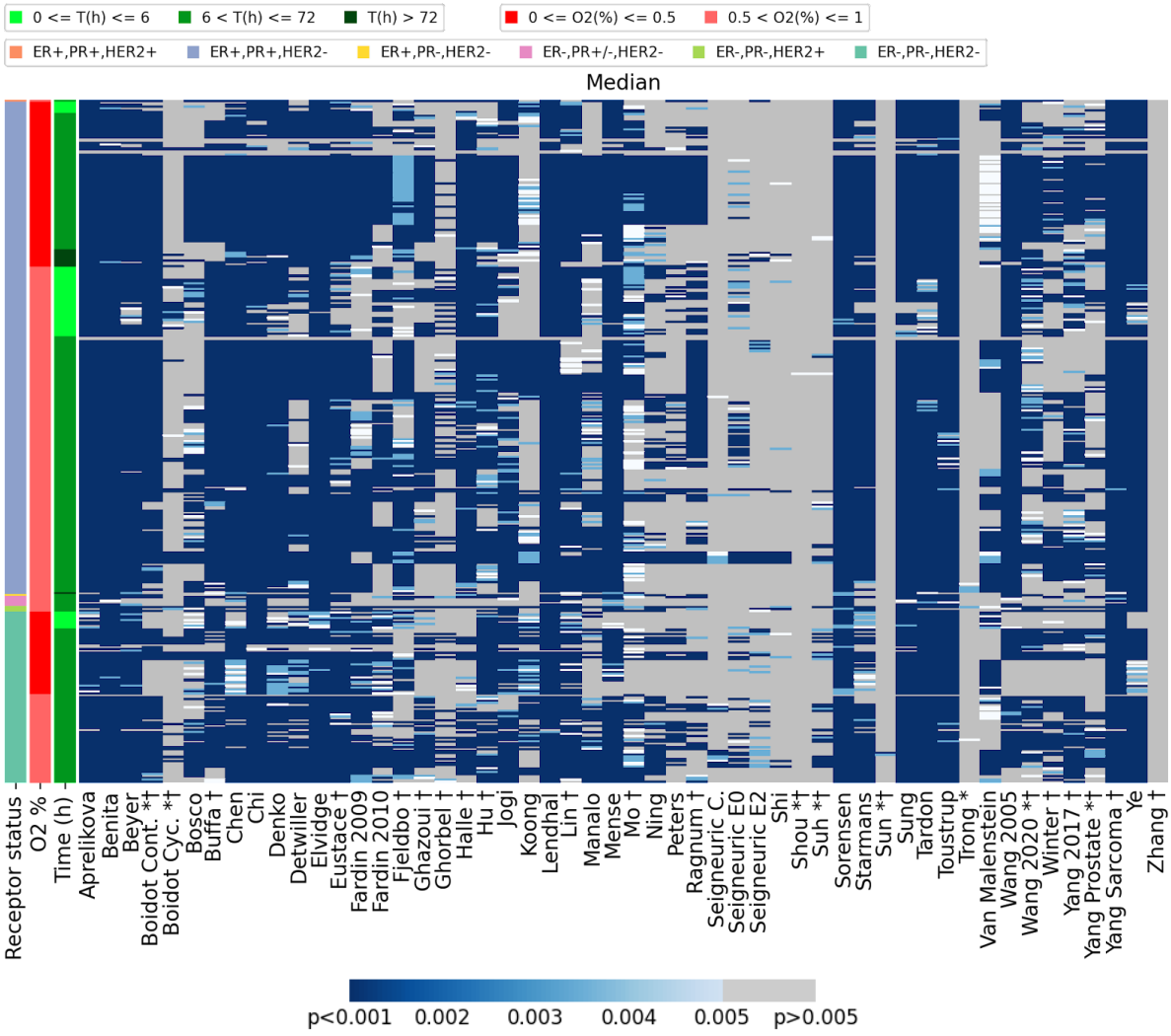
The worst performing signatures across all scores were Sun (from lung cancer samples) and Shou (from melanoma samples). This is shown by the vertical grey bars in all panels of *Fig. 4.1* and also in *Table 4.2* with the maximum accuracy achieved by these signatures being <2%.

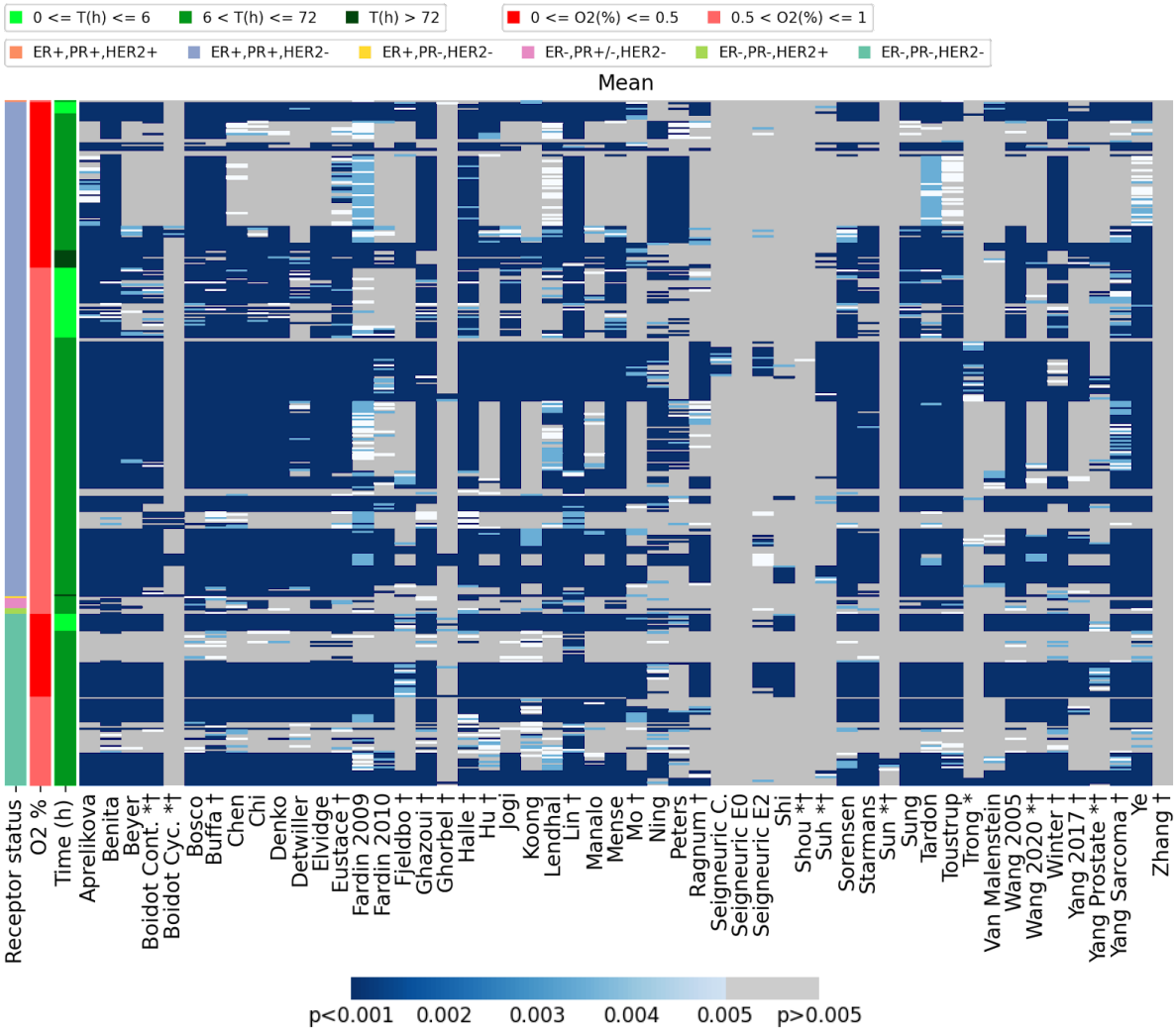
Cell line	Receptor Status	Number of Hypoxic samples	Number of Normoxic samples	Pairs
MCF-7	ER+, PR+, HER2-	79	48	243
MDA-MB-231	ER-, PR-, HER2-	20	14	56
T-47D	ER+, PR+, HER2-	5	5	11
UFH-001	ER-, PR-, HER2-	3	3	9
HCC1806	ER-, PR-, HER2-	3	3	5
ZR-75-1	ER-, PR+/-, HER2-	5	2	5
MDA231-LM2	ER-, PR-, HER2-	2	2	4
BT20	ER-, PR-, HER2-	1	1	1
BT474	ER+, PR+, HER2+	1	1	1
BT549	ER-, PR-, HER2-	1	1	1
CAMA1	ER+, PR+, HER2-	1	1	1
DU4475	ER-, PR-, HER2-	1	1	1
HCC1428	ER+, PR+, HER2-	1	1	1
HCC1569	ER-, PR-, HER2+	1	1	1
HCC1937	ER-, PR-, HER2-	1	1	1
HCC38	ER-, PR-, HER2-	1	1	1
HS578T	ER-, PR-, HER2-	1	1	1
MDA-MB-157	ER-, PR-, HER2-	1	1	1
MDA-MB-175	ER+, PR-, HER2-	1	1	1
MDA-MB-231-PSOC	ER-, PR-, HER2-	1	1	1
MDA-MB-436	ER-, PR-, HER2-	1	1	1
MDA-MB-468	ER-, PR-, HER2-	1	1	1
SKBR3	ER-, PR-, HER2+	1	1	1
SUM1315	ER-, PR-, HER2-	1	1	1

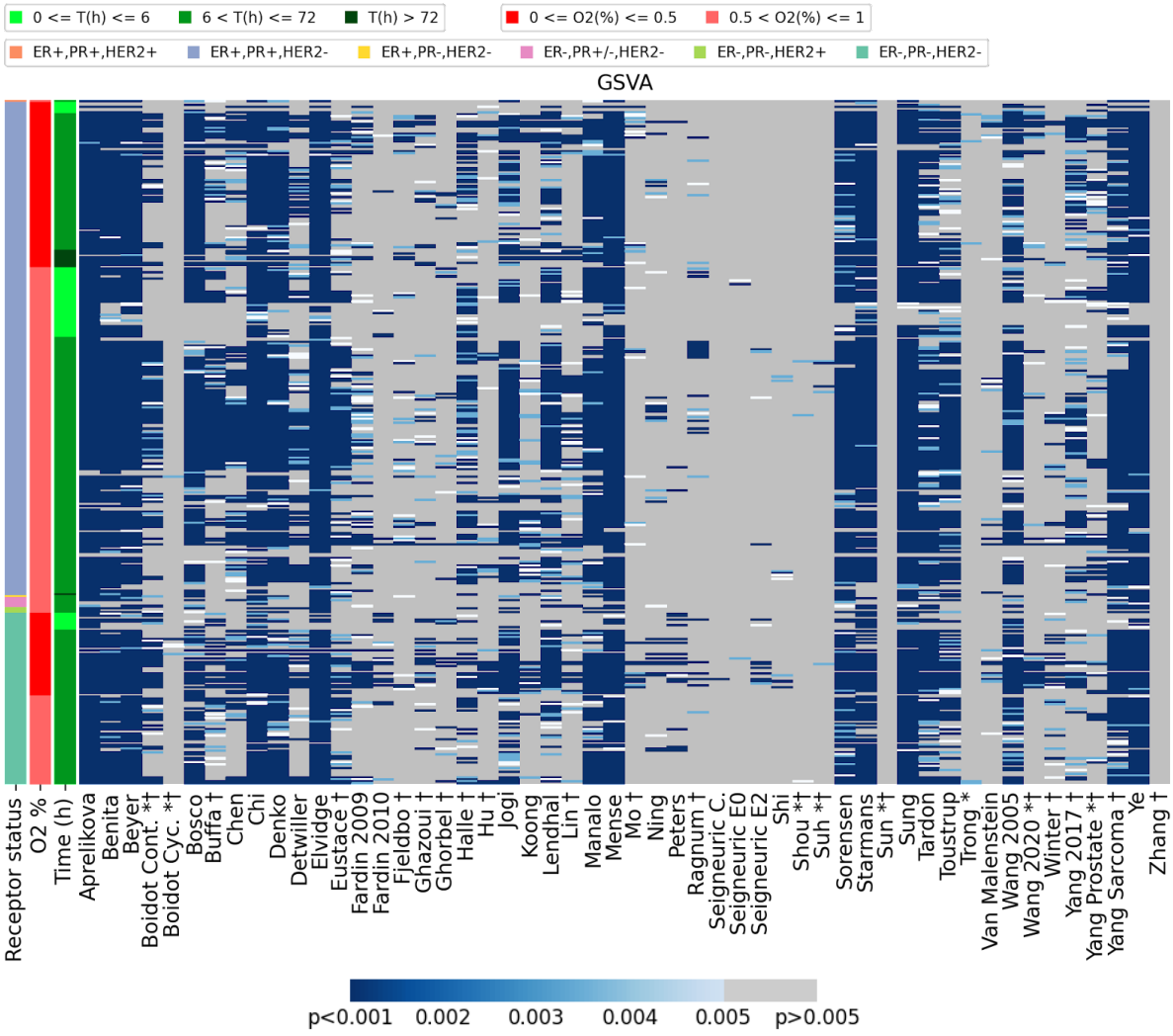
SUM149	ER-, PR-, HER2-	1	1	1
SUM159	ER-, PR-, HER2-	1	1	1
SUM185	ER-, PR-, HER2-	1	1	1
SUM225CWN	ER-, PR-, HER2+	1	1	1
SUM229	ER-, PR-, HER2-	1	1	1
Total		**139**	**99**	**355**

Table 4.1: *Breast cancer cell lines in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO. There is a preponderance of MCF-7 pairwise combinations that can be made in the analysis (243) and thus ER+, PR+, HER2- cells (256). There are 19 triple negative cell lines in the analysis with 89 pairwise combinations. The ZR-75-1 cell line has intermediate PR status but is predominantly PR+. To reflect this, the Receptor Status column reports PR+/-.


0 <= T(h) <= 6
6 < T(h) <= 72
T(h) > 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
ER+,PR+,HER2+
ER+,PR+,HER2-
ER+,PR-,HER2-
ER-,PR+/-,HER2-
ER-,PR-,HER2+
ER-,PR-,HER2-
Median
Receptor status
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= T(h) <= 6
6 < T(h) <= 72
T(h) > 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
ER+,PR+,HER2+
ER+,PR+,HER2-
ER+,PR-,HER2-
ER-,PR+/-,HER2-
ER-,PR-,HER2+
ER-,PR-,HER2-
Mean
Receptor status
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= T(h) <= 6
6 < T(h) <= 72
T(h) > 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
ER+,PR+,HER2+
ER+,PR+,HER2-
ER+,PR-,HER2-
ER-,PR+/-,HER2-
ER-,PR-,HER2+
ER-,PR-,HER2-
GSVA
Receptor status
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

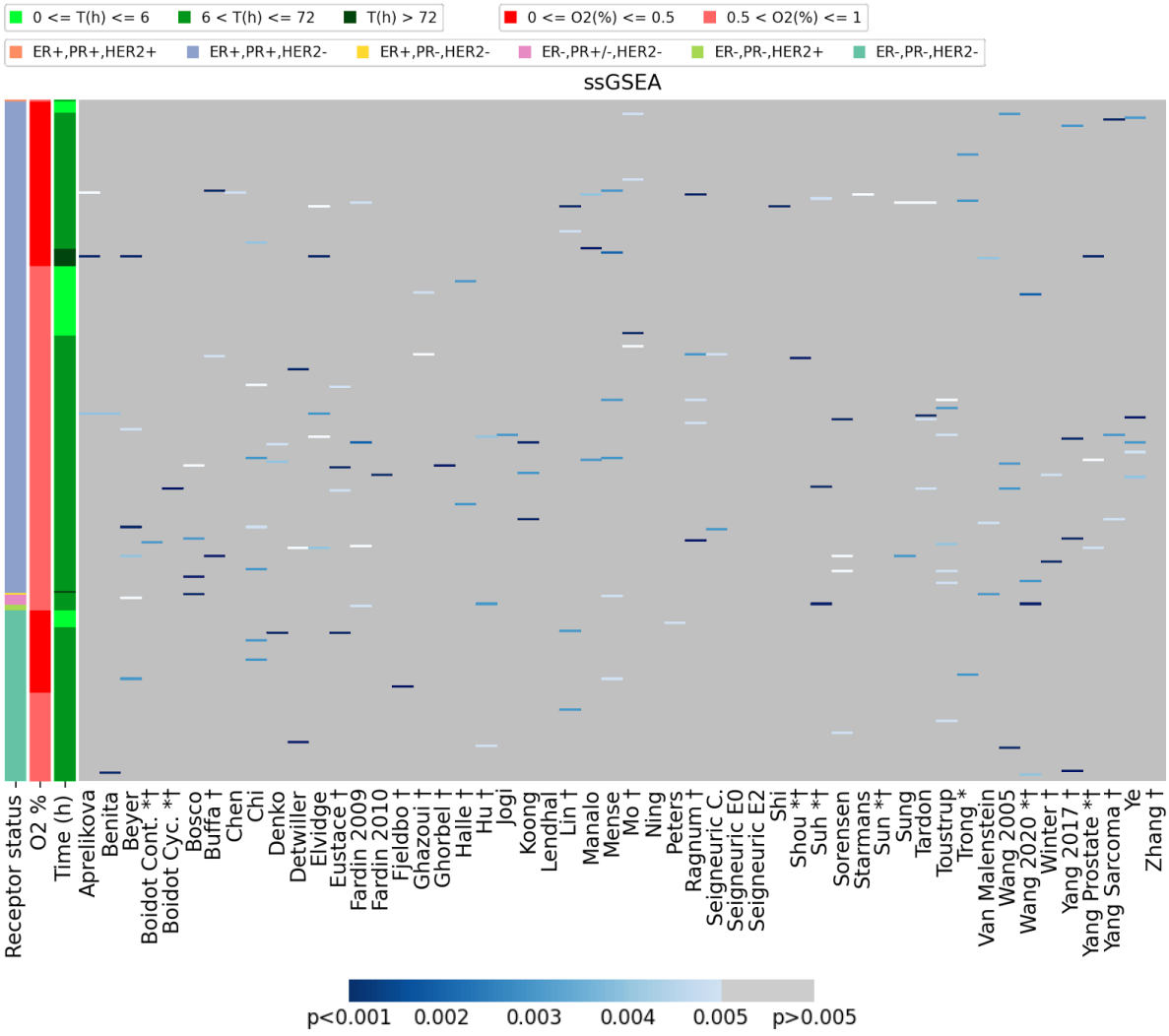


Figure 4.1: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in breast cancer*

Comparison of the scoring methods of the 53 hypoxia signatures across all breast cancer hypoxia experiments identified. Signatures are shown on the x axis, the star symbol (*) denotes signatures made of both upregulated and downregulated genes. The obelisk symbol (†) indicates signatures that have been derived including clinical samples. The darker the shade of blue the more accurate the signatures and specific scores are at differentiating between hypoxic and non-hypoxic samples comparing random gene sets of the same length ($p<0.005$). Grey means the score and signature did not significantly outperform random

gene sets (p>=0.005). The legend above the pictorial figure shows several features for the samples analysed (legend titles are reported at the start of the x-axis). Median (top panel), shows the best classification across the greatest number of experiments compared to mean, GSVA and ssGSEA.

	Median	Mean	GSVA	ssGSEA
Aprelikova	95.21	78.03	94.08	0.85
Benita	96.90	87.32	86.76	0.56
Beyer	92.96	65.35	89.86	1.69
Boidot Cont. *†	83.38	65.07	46.20	0.28
Boidot Cyc. *†	13.80	5.35	1.41	0.28
Bosco	54.93	85.92	77.46	1.13
Buffa †	83.94	88.45	43.38	0.85
Chen	91.83	76.34	45.63	0.28
Chi	93.80	70.42	88.73	1.97
Denko	88.45	69.86	80.28	0.85
Detwiller	71.27	58.31	39.15	0.85
Elvidge	96.34	68.45	93.24	1.41
Eustace †	87.04	80.56	49.30	1.13
Fardin 2009	75.21	77.75	29.01	1.13
Fardin 2010	57.18	46.20	7.32	0.28
Fjeldbo †	68.45	34.65	13.52	0.28
Ghazoui †	56.06	81.69	15.49	0.56
Ghorbel †	23.38	4.79	7.89	0.28
Halle †	79.72	85.07	51.55	0.56
Hu †	79.72	48.45	13.52	0.85
Jogi	78.87	67.61	65.92	0.28
Koong	45.92	43.10	21.97	0.85
Lendhal	96.90	71.27	63.38	0.00
Lin †	88.45	94.08	24.23	1.13
Manalo	53.52	47.61	76.06	0.85
Mense	92.11	69.01	91.55	1.69
Mo †	65.35	34.93	9.01	1.13
Ning	40.85	56.62	9.01	0.00
Peters	29.58	36.90	5.92	0.28
Ragnum †	50.14	41.13	8.73	1.41
Seigneuric C.	5.92	3.94	0.28	0.56
Seigneuric E0	21.41	0.00	0.85	0.00
Seigneuric E2	11.55	14.37	3.66	0.00
Shi	5.07	11.27	2.54	0.28
Shou *†	0.28	0.28	0.85	0.28
Suh *†	11.27	15.77	1.41	0.85
Sorensen	97.75	71.55	84.79	1.13
Starmans	83.38	61.41	84.79	0.28
Sun *†	0.56	1.41	0.00	0.00
Sung	94.37	70.70	90.99	0.56
Tardon	93.80	80.28	75.77	1.13
Toustrup	97.18	77.18	66.20	1.97
Trong *	1.69	10.42	1.41	0.85
Van Malenstein	50.14	33.24	7.89	0.85
Wang 2005	86.76	72.11	73.52	1.13
Wang 2020 *†	60.85	43.38	11.27	1.13
Winter †	44.79	77.18	13.80	0.56
Yang 2017 †	61.69	34.93	41.97	1.13
Yang Prostate *†	39.44	24.51	24.51	0.85
Yang Sarcoma †	97.46	70.42	81.97	0.85
Ye	91.55	81.97	90.70	1.41
Zhang †	0.00	0.00	0.00	0.00
Zou *†	16.06	19.72	3.10	0.85

Table 4.2: *Percentage accuracy of determining hypoxic samples from normoxic samples in breast cancer for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Red cells highlight the best performing signature(s) and score modality/modalities; here Sorensen using the median score performs best (97.75% accuracy). The star symbol (*) denotes signatures made of both upregulated and downregulated genes. The obelisk symbol (†) indicates signatures that have been derived including clinical samples.

4.2 Performance of hypoxia signatures in lung cancer cell lines

Lung cancer cell lines/tissue with gene expression data in hypoxia and normoxia were identified from the Gene Expression Omnibus (as in Methods). This yielded a dataset of *ex vivo* samples (non-small cell lung cancer) and three datasets from cell lines, A549, H460 and PC-9 (*Table 4.3*). Clinical samples originating from the GSE30979 dataset made up the majority of analysis with 100 pairwise combinations possible. In this study (GSE30979), Leithner et al. collected non-small cell lung cancer (NSCLC) fragments from patients prior to surgery and grew them *ex vivo* in ambient oxygen or 1% oxygen[125]. Gene expression was measured using Affymetrix GeneChip 1.0 ST microarrays. The cell line that made up most of the rest of the comparisons was A549, a lung adenocarcinoma explant cell line, which added 14 pairwise combinations to the analysis.

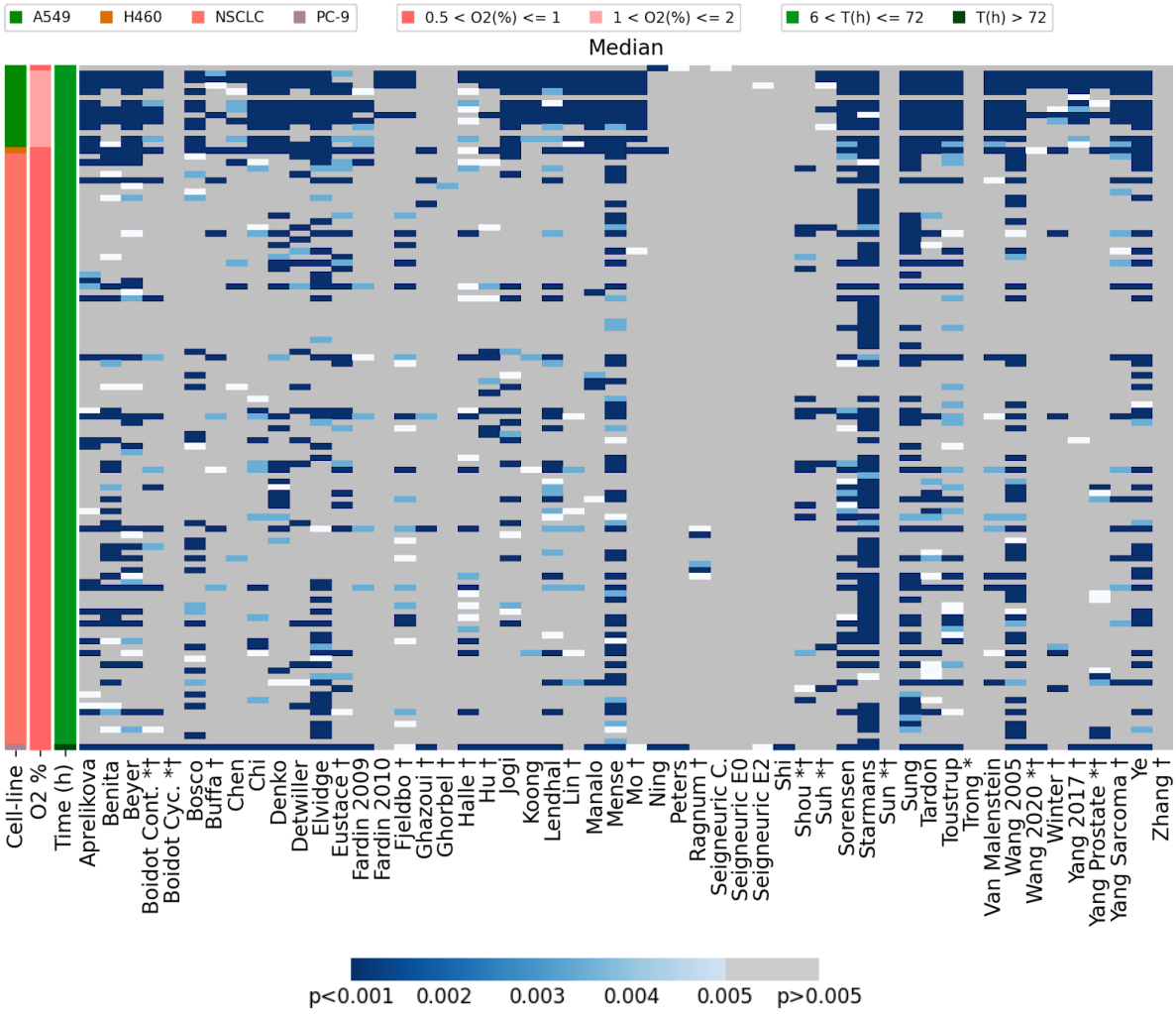
I assessed the performance of the 53 signatures at defining the hypoxic state correctly in these three cell lines and clinical samples (GSE30979). Mean and GVSA scores here appeared the most useful (*Fig. 4.3*), while in contrast to the observations in breast cancer lines (Chapter 4.1) the median score was less accurate (*Fig. 4.3* & *Table 4.4*). The best performing signatures for distinguishing hypoxic from normoxic cells were Starmans and Mense with mean scores giving 86.21% accuracy (highlighted red in *Table 4.4*). The Starmans signature was derived from one prostate, one colon and one breast cancer cell line whereas Mense was derived from fetal astrocytes (*Table 3.1*). Interestingly, neither were derived from lung cancer cell lines or tissue.

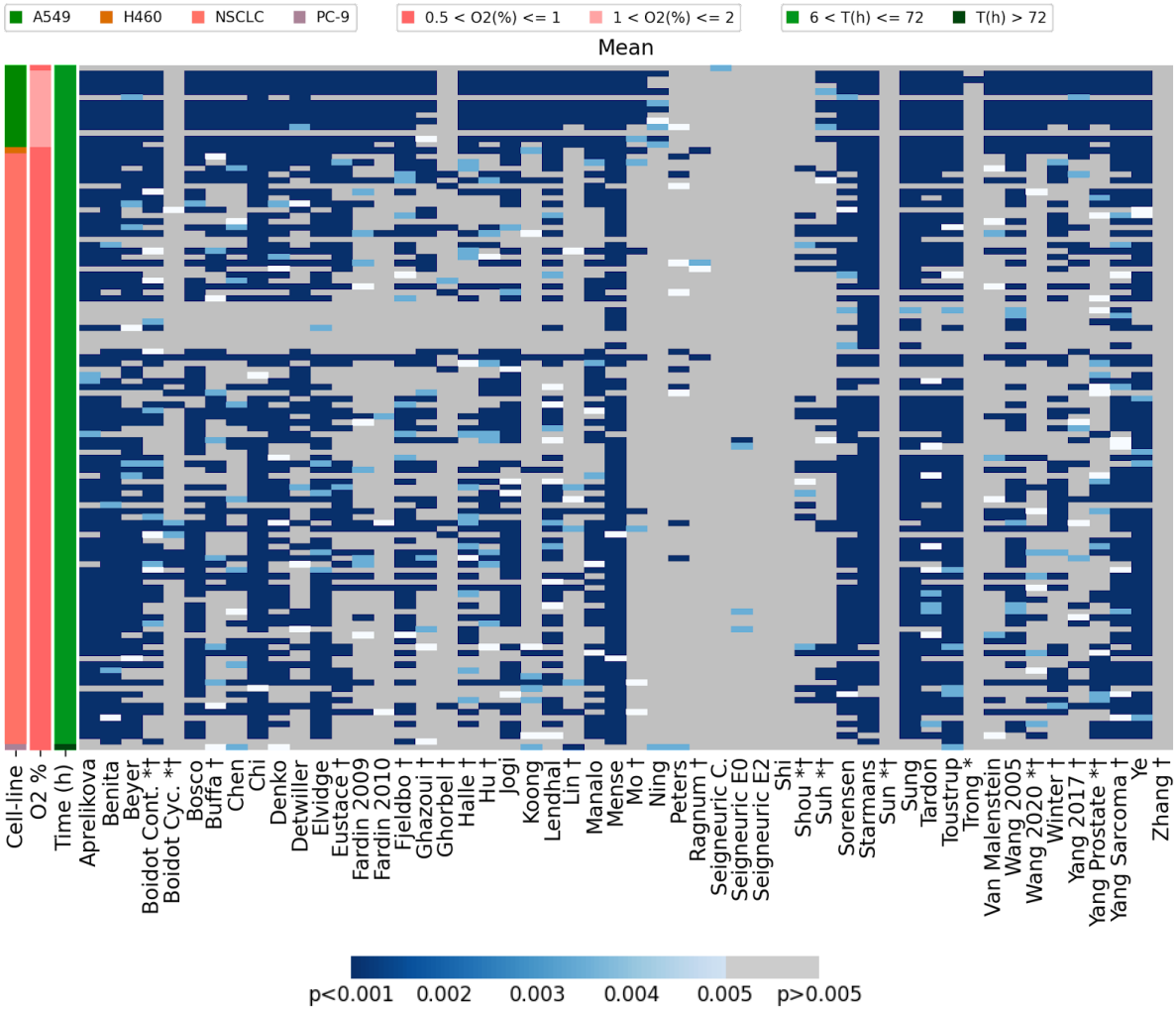
In these datasets, ssGSEA was the worst performing score. The worst performing signature across all scores was the signature developed by Sun, with 0% accuracy on the four scores (*Table 4.4*). Interestingly, this signature was indeed developed on lung tissue, using early-stage lung cancer samples[117]. These workers used ssGSEA scores and transcriptome profiling data of stage I–II lung cancer adenocarcinoma patients to develop their signature, but did not directly measure hypoxia.

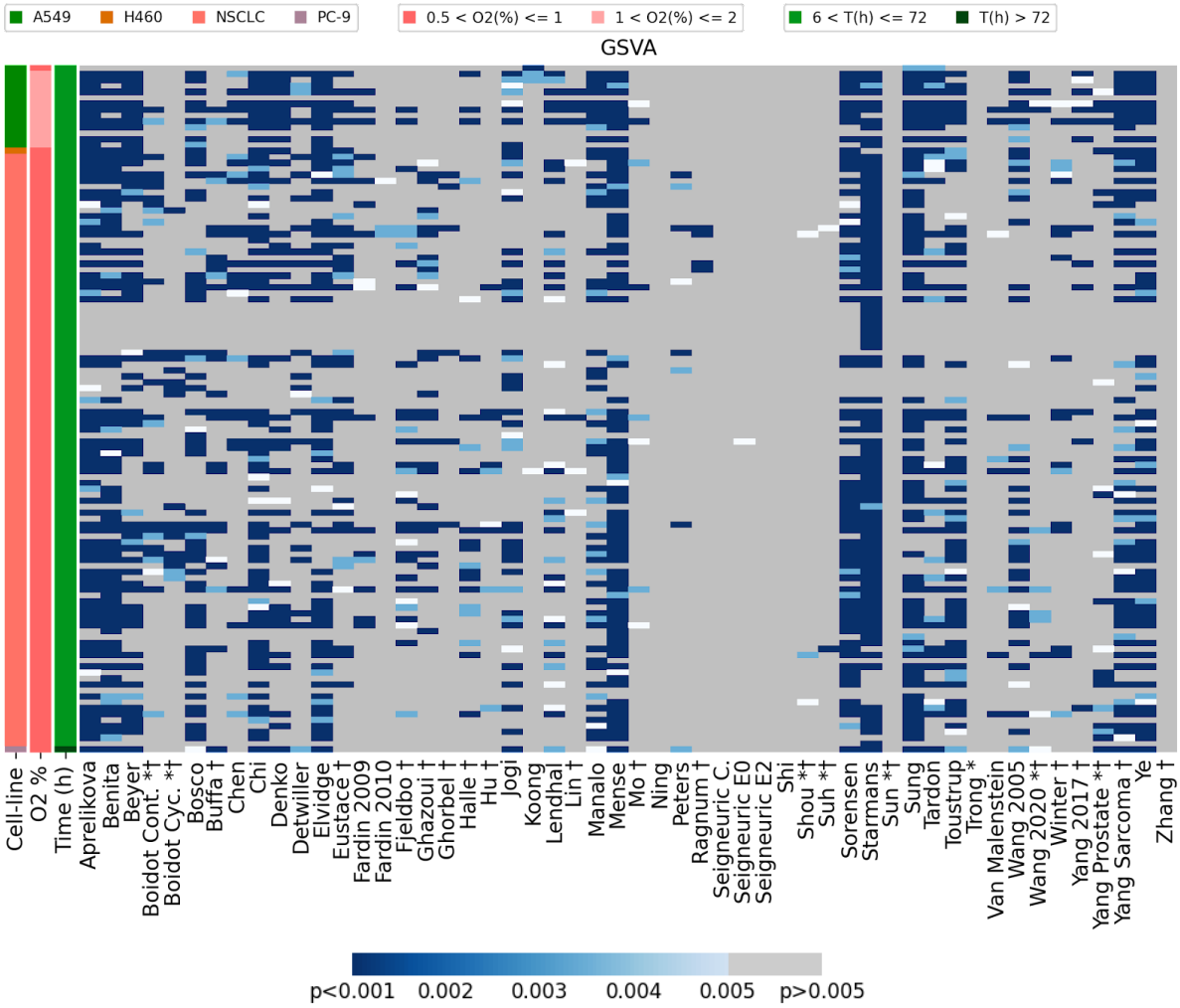
Cell line/samples	Number of Hypoxic samples	Number of Normoxic samples	Pairs
NSCLC *ex vivo* clinical samples	10	10	100
A549	6	6	14
H460	1	1	1
PC-9	1	1	1
Total	**18**	**18**	**116**

Table 4.3: *Lung cancer cell lines in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using lung cancer cell lines. 14 pairs are achieved in the A549 line as three hypoxia and three normoxia samples are taken from one dataset (GSE117036 = 9 pairs), two hypoxia and two normoxia samples are taken from another (GSE117041 = 4 pairs) and one hypoxia and one normoxia samples are taken from another (GSE42416 = 1 pair).


A549
H460
NSCLC
PC-9
0.5 < O2(%) <= 1
1 < O2(%) <= 2
6 < T(h) <= 72
T(h) > 72
Median
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


A549
H460
NSCLC
PC-9
0.5 < O2(%) <= 1
1 < O2(%) <= 2
6 < T(h) <= 72
T(h) > 72
Mean
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


A549
H460
NSCLC
PC-9
0.5 < O2(%) <= 1
1 < O2(%) <= 2
6 < T(h) <= 72
T(h) > 72
GSVA
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

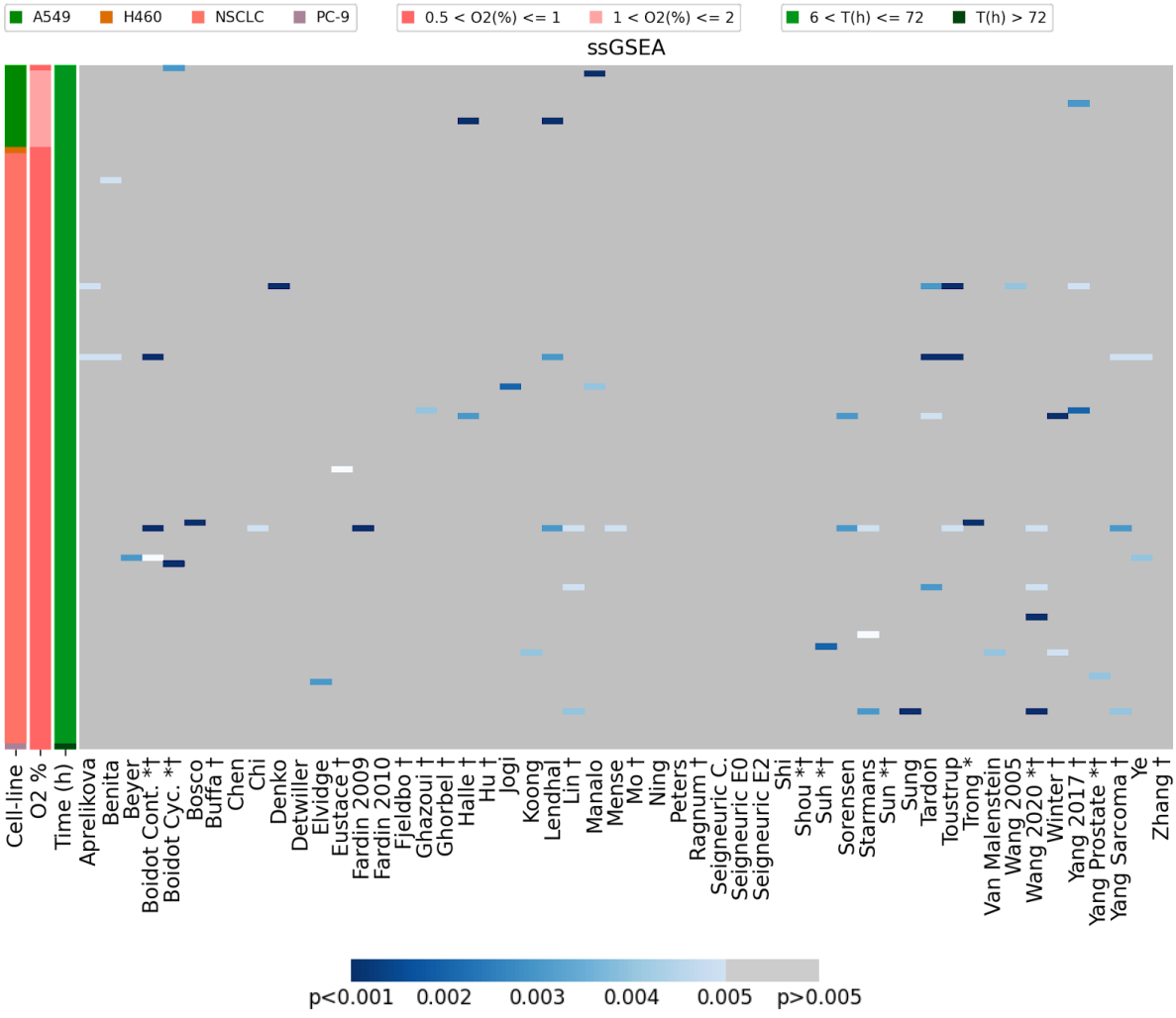


Figure 4.2: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in lung cancer cell lines*

Comparison of the scoring methods of the 53 hypoxia signatures across all lung cancer hypoxia experiments identified. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	30.17	77.59	68.10	1.72
Benita	37.93	81.03	66.38	1.72
Beyer	37.93	70.69	52.59	0.86
Boidot Cont. *†	18.10	60.34	23.28	2.59
Boidot Cyc. *†	0.86	10.34	13.79	1.72
Bosco	31.03	68.10	50.86	0.86
Buffa †	11.21	44.83	21.55	0.00
Chen	12.07	45.69	21.55	0.00
Chi	25.86	81.03	62.07	0.86
Denko	30.17	61.21	33.62	0.86
Detwiller	23.28	45.69	21.55	0.00
Elvidge	46.55	72.41	57.76	0.86
Eustace †	28.45	49.14	33.62	0.86
Fardin 2009	12.93	30.17	13.79	0.86
Fardin 2010	3.45	16.38	2.59	0.00
Fjeldbo †	22.41	57.76	25.86	0.00
Ghazoui †	5.17	37.93	23.28	0.86
Ghorbel †	0.86	9.48	5.17	0.00
Halle †	25.00	46.55	18.10	1.72
Hu †	14.66	43.97	7.76	0.00
Jogi	27.59	63.79	40.52	0.86
Koong	11.21	19.83	3.45	0.86
Lendhal	32.76	60.34	31.90	2.59
Lin †	15.52	18.10	6.90	2.59
Manalo	16.38	60.34	47.41	1.72
Mense	56.90	86.21	73.28	0.86
Mo †	11.21	14.66	6.90	0.00
Ning	2.59	7.76	0.00	0.00
Peters	1.72	11.21	6.90	0.00
Ragnum †	4.31	3.45	3.45	0.00
Seigneuric C.	0.86	0.86	0.00	0.00
Seigneuric E0	0.00	3.45	0.86	0.00
Seigneuric E2	1.72	0.00	0.00	0.00
Shi	0.86	0.00	0.00	0.00
Shou *†	12.93	17.24	2.59	0.00
Suh *†	8.62	25.86	1.72	0.86
Sorensen	40.52	76.72	63.79	1.72
Starmans	74.14	86.21	80.17	2.59
Sun *†	0.00	0.00	0.00	0.00
Sung	51.72	81.03	66.38	0.86
Tardon	33.62	69.83	35.34	3.45
Toustrup	39.66	77.59	52.59	2.59
Trong *	0.00	0.86	0.00	0.86
Van Malenstein	18.97	30.17	7.76	0.86
Wang 2005	50.00	58.62	37.93	0.86
Wang 2020 *†	5.17	25.86	8.62	3.45
Winter †	10.34	46.55	16.38	1.72
Yang 2017 †	11.21	31.03	11.21	2.59
Yang Prostate *†	14.66	43.97	18.10	0.86
Yang Sarcoma †	24.14	73.28	55.17	2.59
Ye	43.10	81.03	59.48	1.72
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	5.17	0.86	0.00

Table 4.4: *Percentage accuracy of determining hypoxic samples from normoxic samples in lung cancer cell lines for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

4.3 Performance of hypoxia signatures in colorectal cancer cell lines

Colorectal cancer cell lines with gene expression data in hypoxia and normoxia were identified from the Gene Expression Omnibus (as in Methods). This yielded seven cell lines: HCT116, HT29, DKO3, COLO-205, HCT-15, LoVo and WiDr. By far the most common cell line used was HCT116, a human colorectal carcinoma cell line that harbours a mutation in codon 13 of the ras proto-oncogene (*Table 4.5*).

I assessed the performance of the 53 signatures at defining hypoxia in these seven cell lines in 11 different experiments (GSE38061, GSE4186, GSE41666, GSE41491, GSE29641, GSE81513, GSE58049, GSE90599, GSE35973, GSE42416, GSE109318). The best performing overall score in these colorectal cancer cell lines was the median (*Fig. 4.3*, top panel), whilst again, ssGSEA appeared the worst performing score (*Fig. 4.3*, bottom panel). The best performing signatures were Chen (82.50% using median) and Sung (82.50%, using GSVA) and other accuracies are detailed in Table 4.6. These two signatures have been developed from lung adenocarcinoma (A549, HCC827) and head and neck cancer cell lines (CNE-2, C666-1, HONE-1, HK1) respectively. Unexpectedly, none of the signatures derived from experiments including colorectal cell lines (Benita, Lendhal, Starmans and Zou) have shown the highest accuracy in the classification of hypoxic samples, however, the median score for Benita was able to reach 75%.

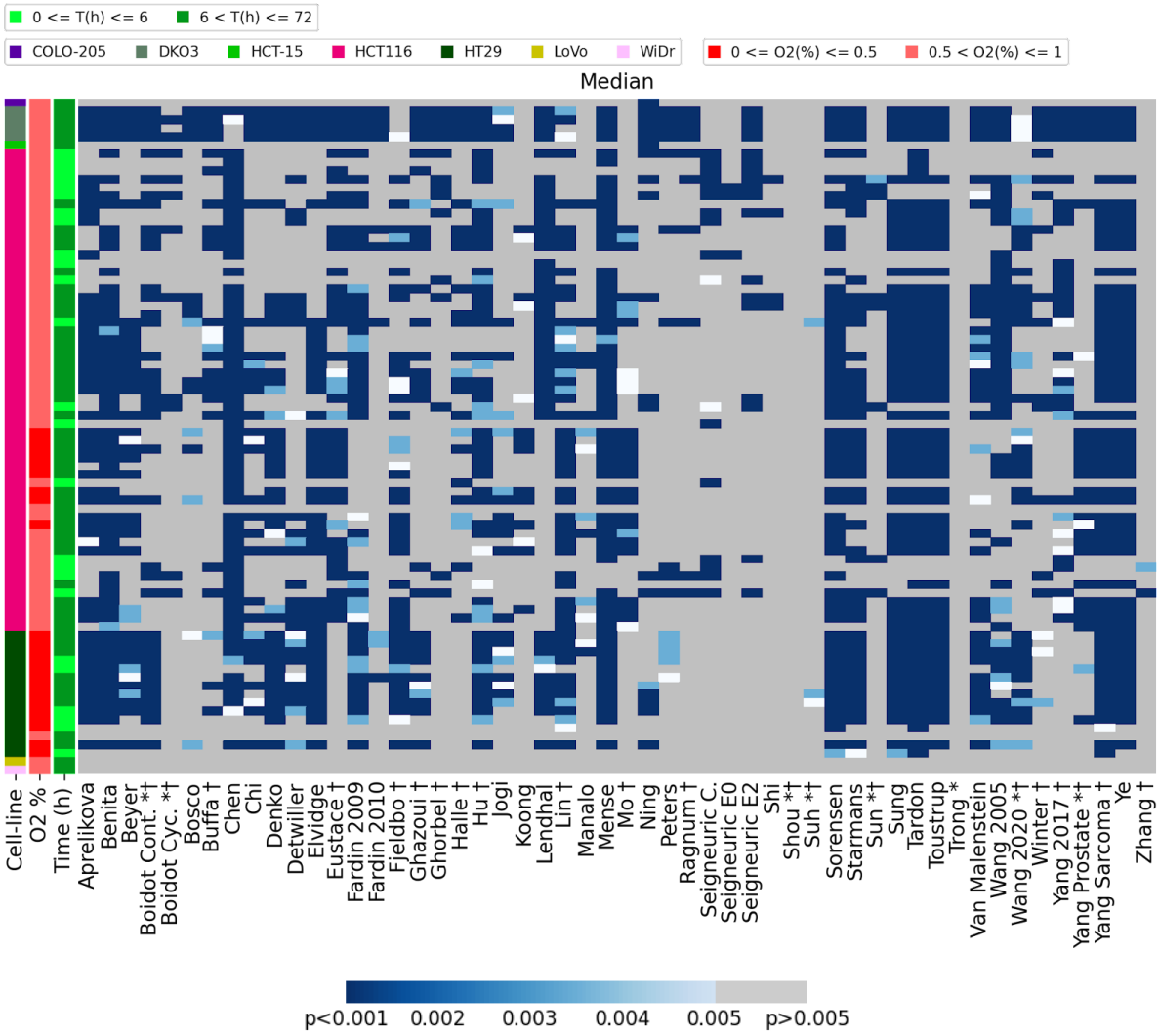
The worst performing signature across all the scores was again Shou, which was unable to correctly classify any pairwise combination in the analysis (0% accuracy). However it is worth noting that almost 30% of the hypoxic samples of the pairwise combinations in the analysis have been exposed to low oxygen tensions (0-1 % O2) for a relatively short time between 1 and 6 hours. These tend not to be correctly classified as hypoxic even using the highest performing signatures and scores (*Fig. 4.3*). This suggests that the early hypoxic response in colorectal cancer lines is different compared to >6 hours

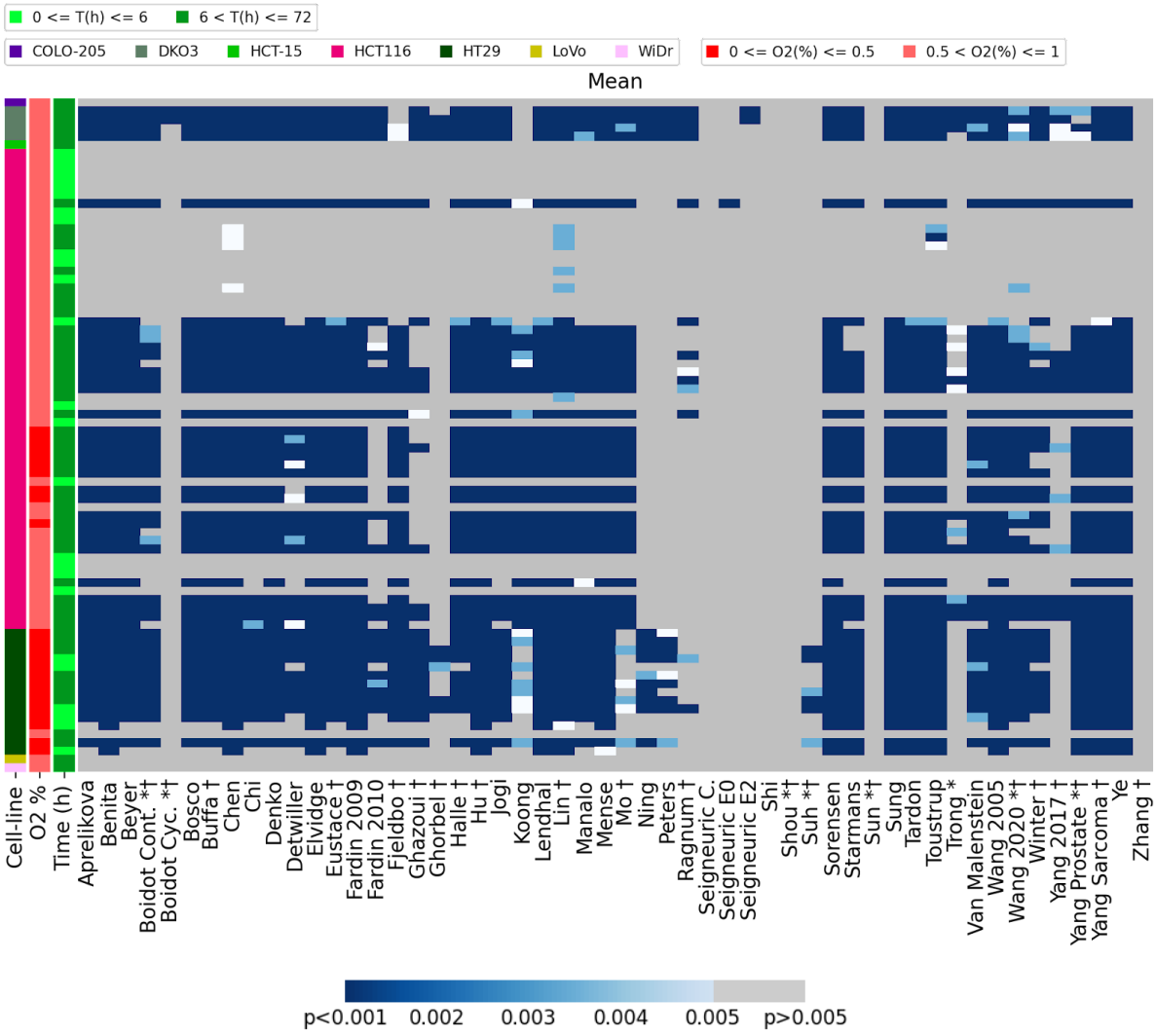
and perhaps this is the reason why signatures struggle to correctly differentiate between normoxic and hypoxic samples. However, this does not appear to be the case in breast cancer (Chapter 3) but I will continue to monitor this potential pattern in other cancer types.

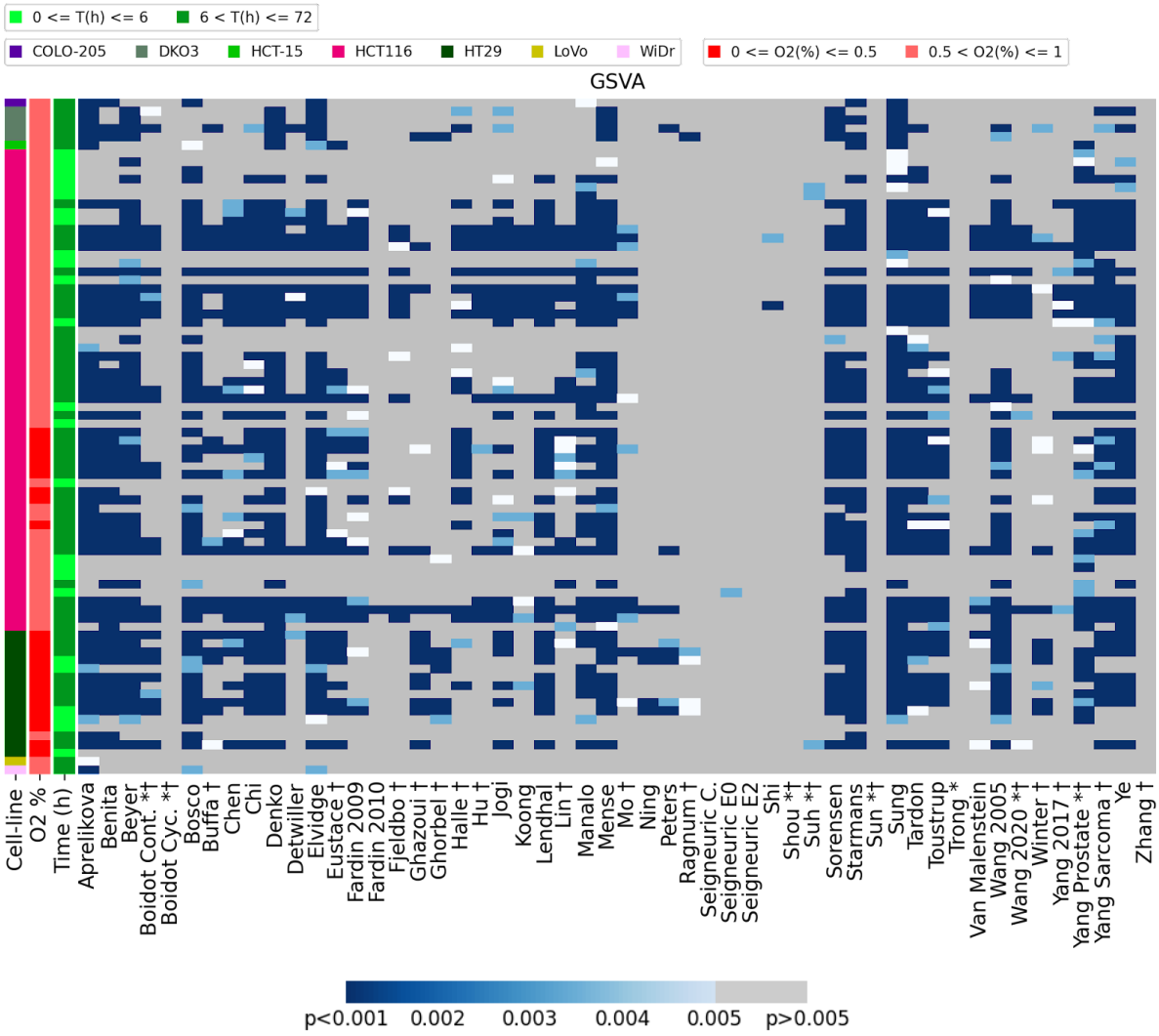
Cell line	Number of Hypoxic samples	Number of Normoxic samples	Pairs
HCT116	19	16	57
HT29	15	3	15
DKO3	2	2	4
COLO-205	1	1	1
HCT-15	1	1	1
LoVo	1	1	1
WiDr	1	1	1
Total	**40**	**25**	**80**

Table 4.5: *Colorectal cancer cell lines in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using colorectal cancer cell lines.


0 <= T(h) <= 6
6 < T(h) <= 72
COLO-205
DKO3
HCT-15
HCT116
HT29
LoVo
WiDr
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
Median
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= T(h) <= 6
6 < T(h) <= 72
COLO-205
DKO3
HCT-15
HCT116
HT29
LoVo
WiDr
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
Mean
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont.*†
Boidot Cyc.*†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou*†
Suh*†
Sorensen
Starmans
Sun*†
Sung
Tardon
Toustrup
Trong*
Van Malenstein
Wang 2005
Wang 2020*†
Winter †
Yang 2017 †
Yang Prostate*†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= T(h) <= 6
6 < T(h) <= 72
COLO-205
DKO3
HCT-15
HCT116
HT29
LoVo
WiDr
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
GSVA
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

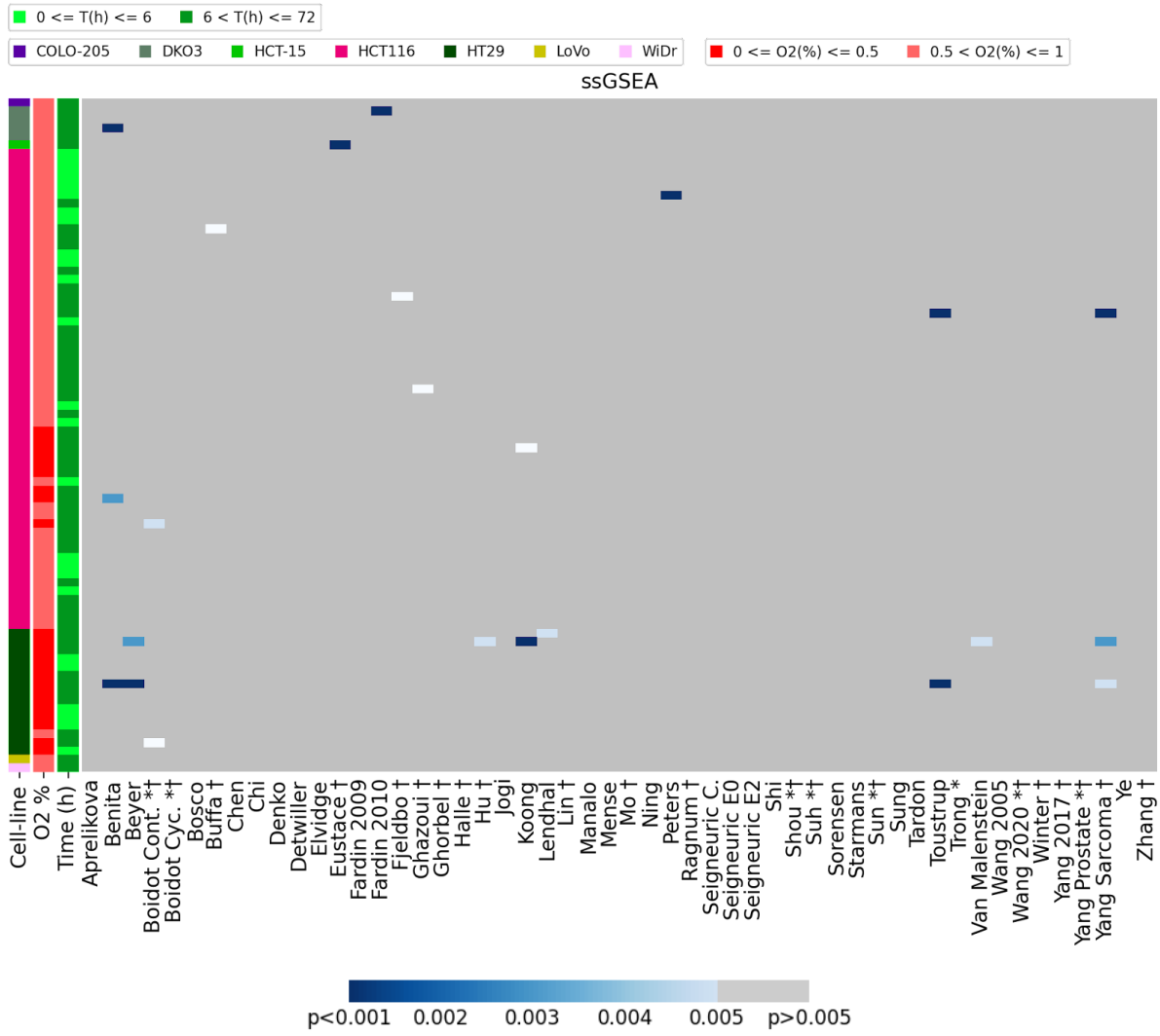


Figure 4.3: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in colorectal cancer*

Comparison of the scoring methods of the 53 hypoxia signatures across all colorectal cancer hypoxia experiments identified. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	63.75	56.25	70.00	0.00
Benita	75.00	58.75	60.00	3.75
Beyer	47.50	56.25	68.75	2.50
Boidot Cont. *†	48.75	50.00	32.50	2.50
Boidot Cyc. *†	13.75	2.50	0.00	0.00
Bosco	18.75	56.25	66.25	0.00
Buffa †	31.25	56.25	27.50	1.25
Chen	82.50	63.75	38.75	0.00
Chi	36.25	55.00	52.50	0.00
Denko	50.00	56.25	66.25	0.00
Detwiller	38.75	50.00	20.00	0.00
Elvidge	55.00	58.75	70.00	0.00
Eustace †	52.50	56.25	40.00	1.25
Fardin 2009	61.25	58.75	36.25	0.00
Fardin 2010	12.50	32.50	1.25	1.25
Fjeldbo †	55.00	52.50	18.75	1.25
Ghazoui †	36.25	32.50	18.75	1.25
Ghorbel †	17.50	10.00	11.25	0.00
Halle †	30.00	53.75	33.75	0.00
Hu †	68.75	58.75	17.50	1.25
Jogi	30.00	53.75	42.50	0.00
Koong	17.50	45.00	17.50	2.50
Lendhal	55.00	58.75	48.75	1.25
Lin †	66.25	66.25	28.75	0.00
Manalo	16.25	55.00	56.25	0.00
Mense	80.00	57.50	67.50	0.00
Mo †	33.75	46.25	18.75	0.00
Ning	23.75	17.50	6.25	0.00
Peters	18.75	16.25	10.00	1.25
Ragnum †	12.50	16.25	6.25	0.00
Seigneuric C.	21.25	0.00	0.00	0.00
Seigneuric E0	3.75	1.25	1.25	0.00
Seigneuric E2	22.50	2.50	0.00	0.00
Shi	5.00	0.00	2.50	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	3.75	7.50	3.75	0.00
Sorensen	75.00	58.75	60.00	0.00
Starmans	62.50	52.50	68.75	0.00
Sun *†	10.00	0.00	0.00	0.00
Sung	71.25	58.75	82.50	0.00
Tardon	77.50	58.75	61.25	0.00
Toustrup	71.25	62.50	52.50	2.50
Trong *	0.00	17.50	0.00	0.00
Van Malenstein	48.75	52.50	20.00	1.25
Wang 2005	60.00	58.75	62.50	0.00
Wang 2020 *†	47.50	52.50	15.00	0.00
Winter †	22.50	42.50	18.75	0.00
Yang 2017 †	41.25	21.25	13.75	0.00
Yang Prostate *†	33.75	56.25	61.25	0.00
Yang Sarcoma †	73.75	58.75	60.00	3.75
Ye	71.25	58.75	65.00	0.00
Zhang †	2.50	0.00	0.00	0.00
Zou *†	35.00	0.00	6.25	2.50

Table 4.6: *Percentage accuracy of determining hypoxic samples from normoxic samples in colorectal cancer for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

4.4 Performance of hypoxia signatures in liver cancer cell lines

Liver cancer cell lines with gene expression data in hypoxia and normoxia were identified from the Gene Expression Omnibus (as in Methods). This identified six liver cancer cell lines: HepG2, Huh-7, Hep3B, PLC-PRF-5, SK-HEP-1 and SMMC-7721. The commonest lines used are HepG2 and Huh-7 (*Table 4.7*). The cell line with the largest number of combinations of pairs was HepG2 (46) which originates from a 15-year-old caucasian boy.

I assessed the performance of the 53 signatures at defining hypoxia in these six liver cancer cell lines in 9 different experiments (GSE120886, GSE18494, GSE55212, GSE59729, GSE41666, GSE42416, GSE1056, GSE120611, GSE57613).

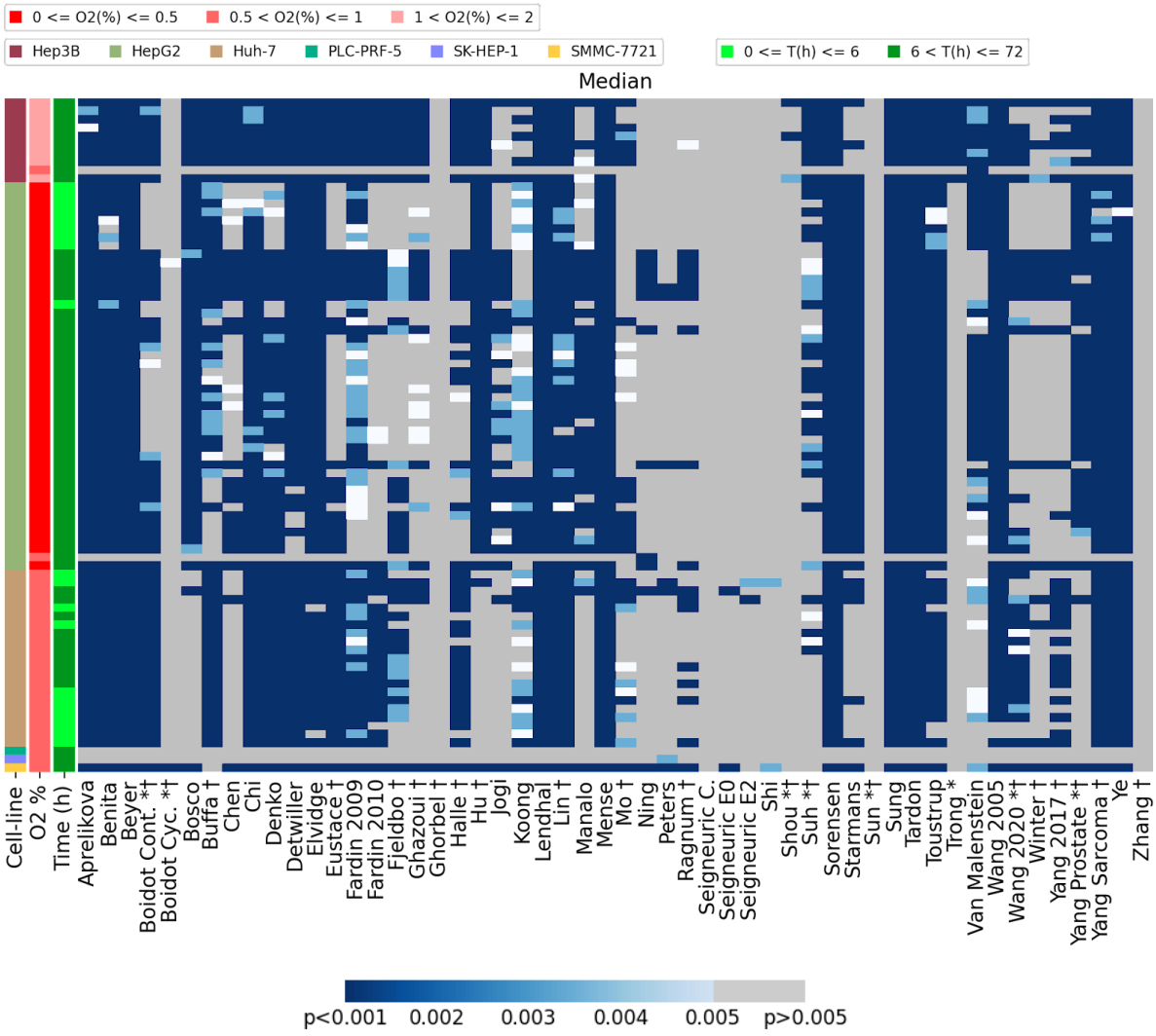
In distinguishing hypoxic cell lines, the median, mean and GSVA scores all performed well. Using the median score, the highest performing signatures (Aprelikova, Beyer, Lendhal, Mense, Sorensen, Sung, Tardon and Ye) achieved 95% accuracy (highlighted red in *Table 4.8*), while Toustrup also reached 95% accuracy on the mean score. Looking at the highest performing signatures, these were developed from a range of tissues (*Table 3.1*) and the Lendhal signature was the only one that included any liver cancer cell line in its derivation (Hep3B). The signature that was developed solely from the liver cancer cell line HepG2, Van Malenstein, reached only 63.75% accuracy in identifying hypoxia samples (using median, *Fig. 4.4*, top panel). This again underlines the benefit of unbiased testing of all signatures, independently of tissue origin, across all tissue types.

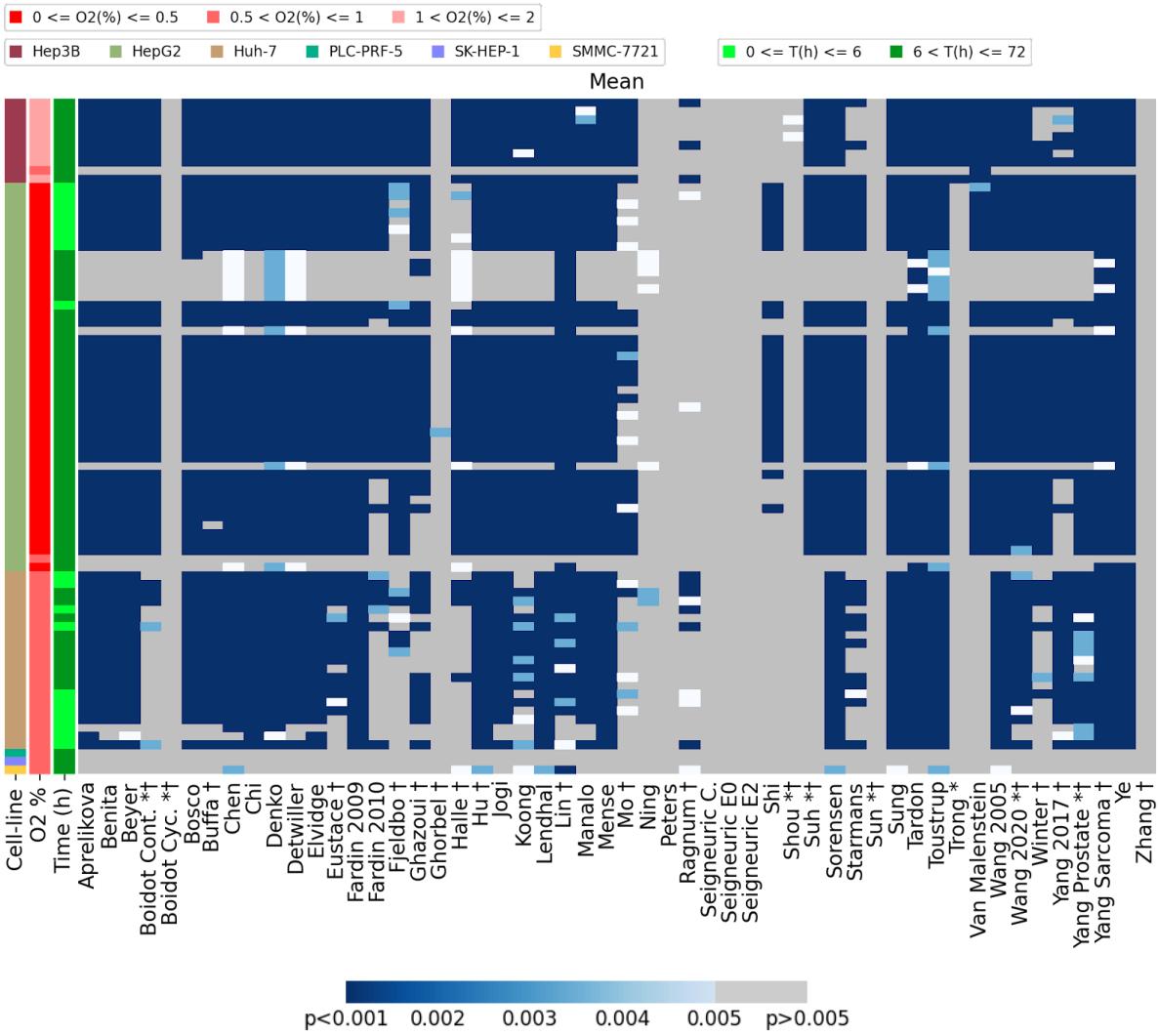
Again, the ssGSEA score performs poorly (*Fig. 4.4*, final panel), with the highest percentage accuracy using any signature reaching only 2.5% (*Table 4.8*). The worst performing signatures were Shou, Sun and Zhang (vertical grey columns in almost all heatmaps in *Fig. 4.4*), with Shou being the best performing of these, achieving only 6.25% accuracy using GVSA (*Table 4.8*).

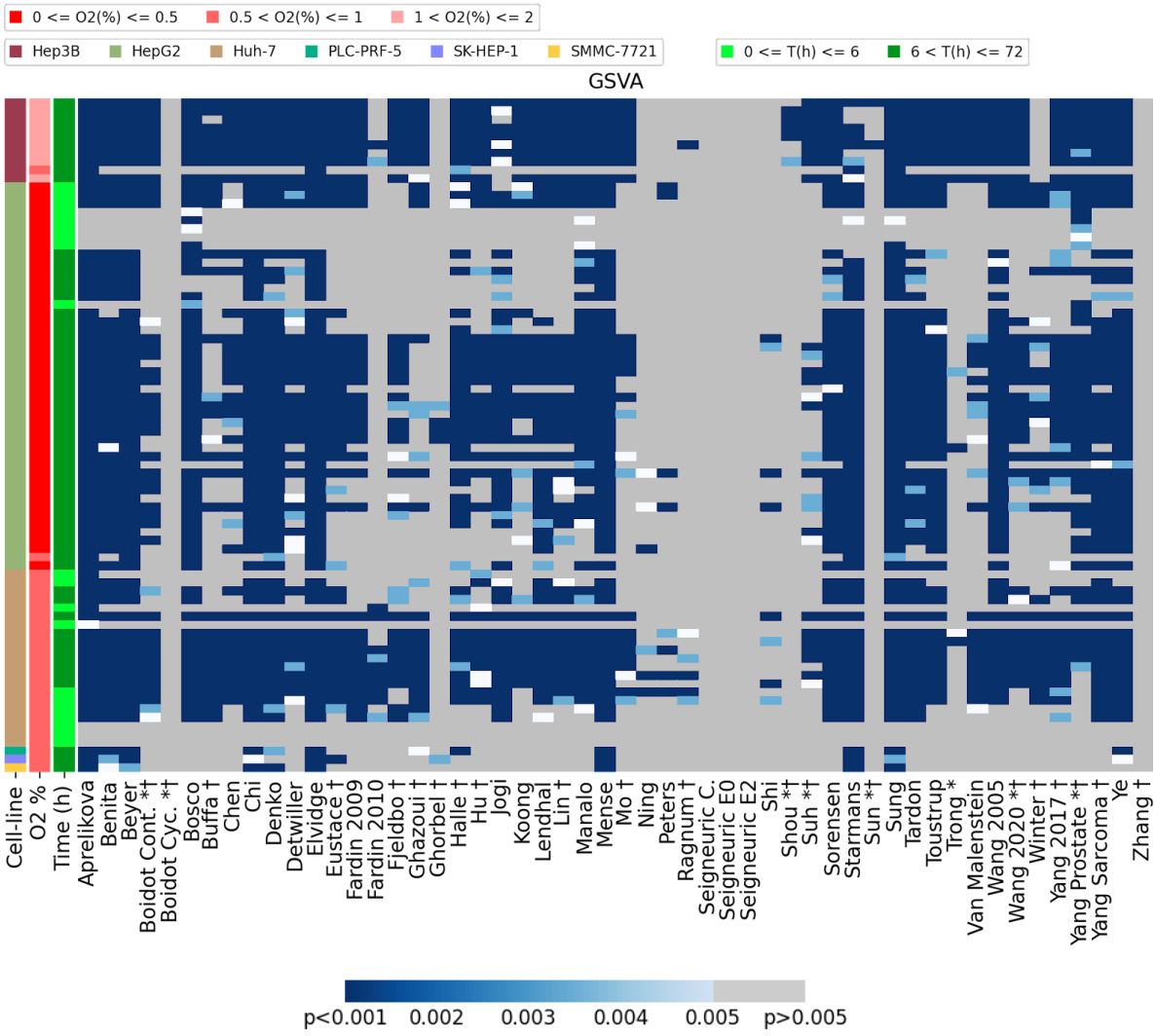
Cell line/samples	Number of Hypoxic samples	Number of Normoxic samples	Pairs
HepG2	16	10	46
Huh-7	12	4	21
Hep3B	4	4	10
PLC-PRF-5	1	1	1
SK-HEP-1	1	1	1
SMMC-7721	1	1	1
Total	**18**	**18**	**116**

Table 4.7: *Liver cancer cell lines in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using liver cancer cell lines.


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
Hep3B
HepG2
Huh-7
PLC-PRF-5
SK-HEP-1
SMMC-7721
0 <= T(h) <= 6
6 < T(h) <= 72
Median
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
Hep3B
HepG2
Huh-7
PLC-PRF-5
SK-HEP-1
SMMC-7721
0 <= T(h) <= 6
6 < T(h) <= 72
Mean
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
Hep3B
HepG2
Huh-7
PLC-PRF-5
SK-HEP-1
SMMC-7721
0 <= T(h) <= 6
6 < T(h) <= 72
GSVA
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

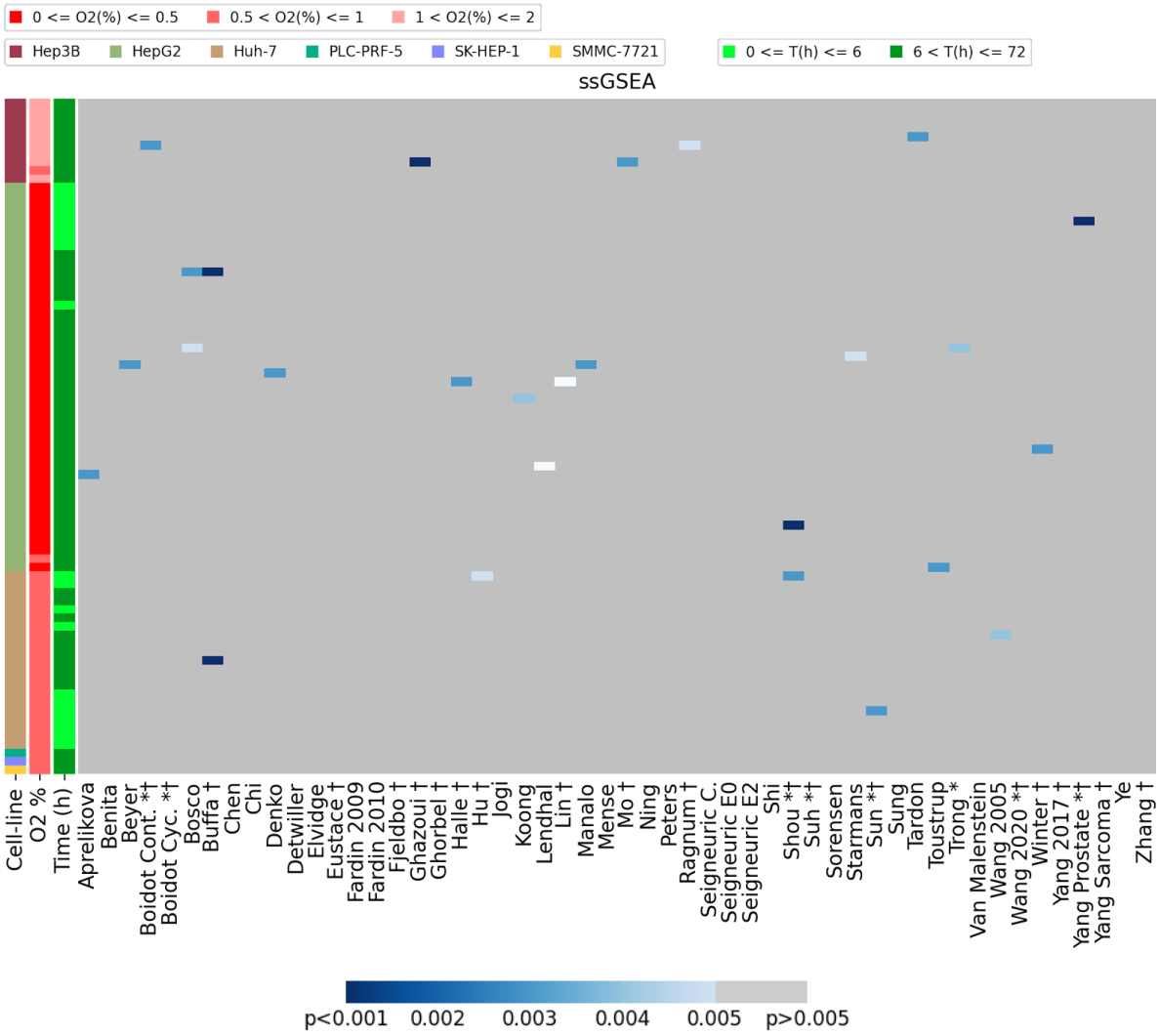


Figure 4.4: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in liver cancer cell lines*

Comparison of the scoring methods of the 53 hypoxia signatures across all liver cancer hypoxia experiments identified. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	95.00	81.25	88.75	1.25
Benita	92.50	80.00	81.25	0.00
Beyer	95.00	81.25	83.75	1.25
Boidot Cont. *†	68.75	62.50	55.00	1.25
Boidot Cyc. *†	2.50	0.00	0.00	0.00
Bosco	70.00	81.25	81.25	2.50
Buffa †	81.25	78.75	41.25	2.50
Chen	41.25	92.50	50.00	0.00
Chi	92.50	82.50	82.50	0.00
Denko	83.75	93.75	80.00	1.25
Detwiller	88.75	91.25	58.75	0.00
Elvidge	92.50	81.25	85.00	0.00
Eustace †	61.25	75.00	57.50	0.00
Fardin 2009	90.00	82.50	48.75	0.00
Fardin 2010	50.00	53.75	15.00	0.00
Fjeldbo †	53.75	58.75	46.25	0.00
Ghazoui †	38.75	67.50	40.00	1.25
Ghorbel †	1.25	1.25	6.25	0.00
Halle †	56.25	65.00	65.00	1.25
Hu †	70.00	83.75	51.25	1.25
Jogi	41.25	80.00	66.25	0.00
Koong	68.75	75.00	45.00	1.25
Lendhal	95.00	83.75	63.75	1.25
Lin †	93.75	91.25	56.25	1.25
Manalo	51.25	80.00	72.50	1.25
Mense	95.00	82.50	82.50	0.00
Mo †	37.50	48.75	33.75	1.25
Ning	15.00	8.75	7.50	0.00
Peters	8.75	0.00	10.00	0.00
Ragnum †	22.50	18.75	7.50	1.25
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	2.50	0.00	0.00	0.00
Seigneuric E2	2.50	0.00	0.00	0.00
Shi	2.50	33.75	10.00	0.00
Shou *†	3.75	2.50	6.25	2.50
Suh *†	52.50	56.25	35.00	0.00
Sorensen	95.00	83.75	75.00	0.00
Starmans	67.50	62.50	81.25	1.25
Sun *†	0.00	0.00	2.50	1.25
Sung	95.00	83.75	87.50	0.00
Tardon	95.00	93.75	73.75	1.25
Toustrup	90.00	95.00	67.50	1.25
Trong *	11.25	11.25	16.25	1.25
Van Malenstein	63.75	57.50	38.75	0.00
Wang 2005	92.50	83.75	75.00	1.25
Wang 2020 *†	48.75	80.00	48.75	0.00
Winter †	20.00	58.75	36.25	1.25
Yang 2017 †	40.00	66.25	57.50	0.00
Yang Prostate *†	61.25	81.25	66.25	1.25
Yang Sarcoma †	93.75	93.75	72.50	0.00
Ye	95.00	93.75	77.50	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	3.75	33.75	20.00	1.25

Table 4.8: *Percentage accuracy of determining hypoxic samples from normoxic samples in liver cancer cell lines for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

4.5 Performance of hypoxia signatures in cervical cancer cell lines

Cervical cancer cell lines with gene expression data in hypoxia and normoxia were identified from the Gene Expression Omnibus (see methods). Eight cell lines fell into this category: HeLa, SiHa, CaSki, C-33, C-41, HT-3, ME-180 and SW756 (*Table 4.9*). HeLa and SiHa made up the majority of the pairs that could be analysed (70/77). Both HeLa and SiHa cells have been reported to contain human papilloma virus sequences (according to the American Type Culture Collection), Hela containing HPV-18 and SiHa HPV-16.

I assessed the performance of the 53 signatures for identifying hypoxia in these eight cell lines in 8 different experiments (GSE72723, GSE36562, GSE55211, GSE3051, GSE42416, GSE147384, GSE141941, GSE33521). Median, mean and GSVA scores appeared the most effective overall, while ssGSEA performed worst (*Fig. 4.5*). The best performing signature was Aprelikova which, using the GSVA score, achieved 97.40% accuracy in identifying hypoxic samples (*Table 4.10*). The Aprelikova signature was developed from the MCF-7 breast cancer cell line at 0.5% oxygen at an 8h time point. However, it is notable that a large number of the published signatures accurately classified hypoxic samples in cervical cancer lines in over 90% of cases using a variety of scores (*Table 4.10*). In contrast, the worst performing signatures across all scores were Shou and Zhang with 0% accuracy across all scores. Seven signatures included cervical cancer lines in their development (Denko, Beyer, Benita, Lendhal, Toustrup, Halle, Fjbeldo); five of these reached 93.51% accuracy using mean or median (*Table 4.10*).

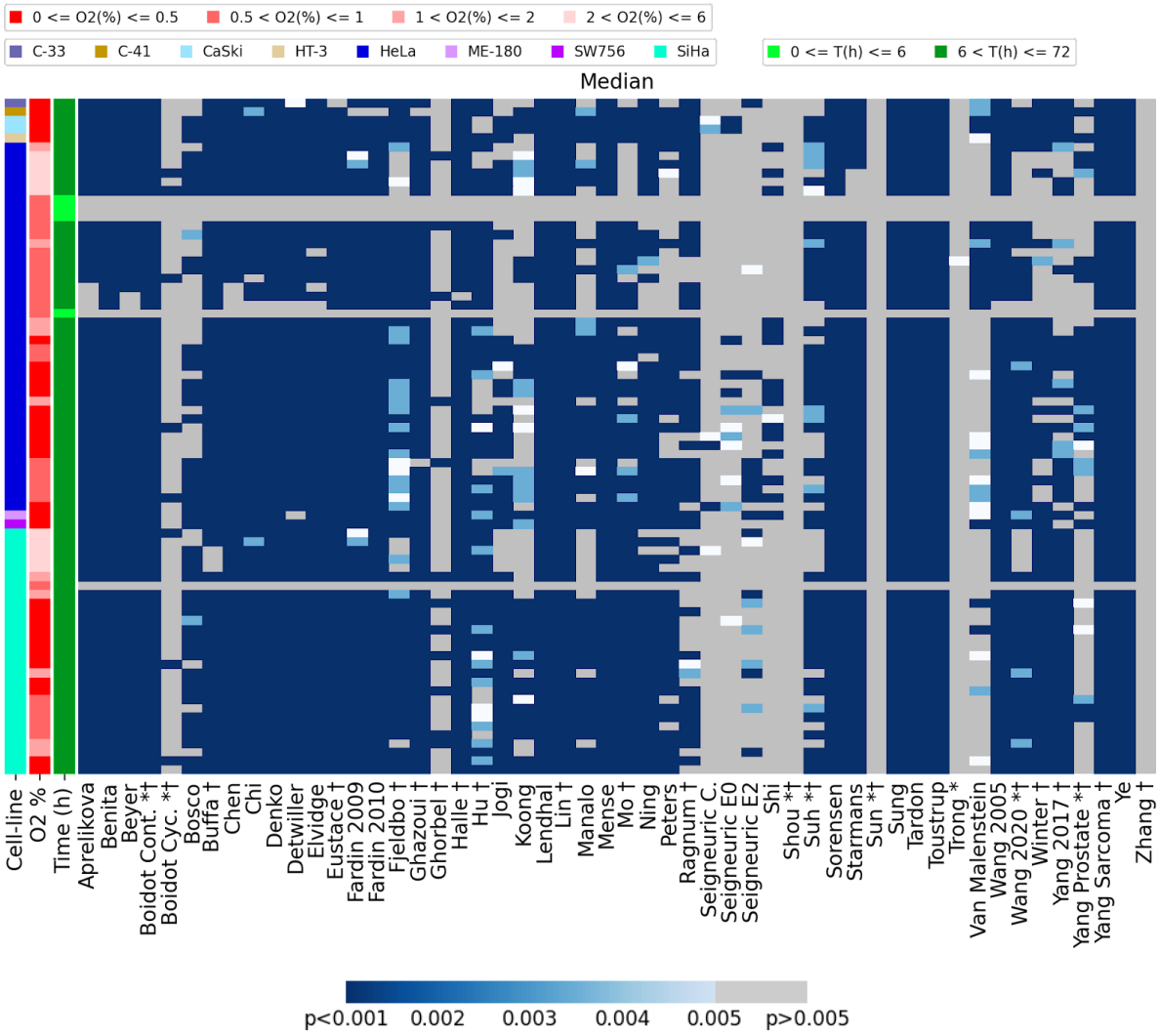
In this cancer specific analysis, measurements at 5% oxygen were also included. These are displayed in the heatmap as the only samples in the 2 < O2% >= 6 group. Although at the benchside 5% oxygen is generally not considered hypoxia, the goal was to investigate if hypoxia signatures could differentiate between this condition and ambient

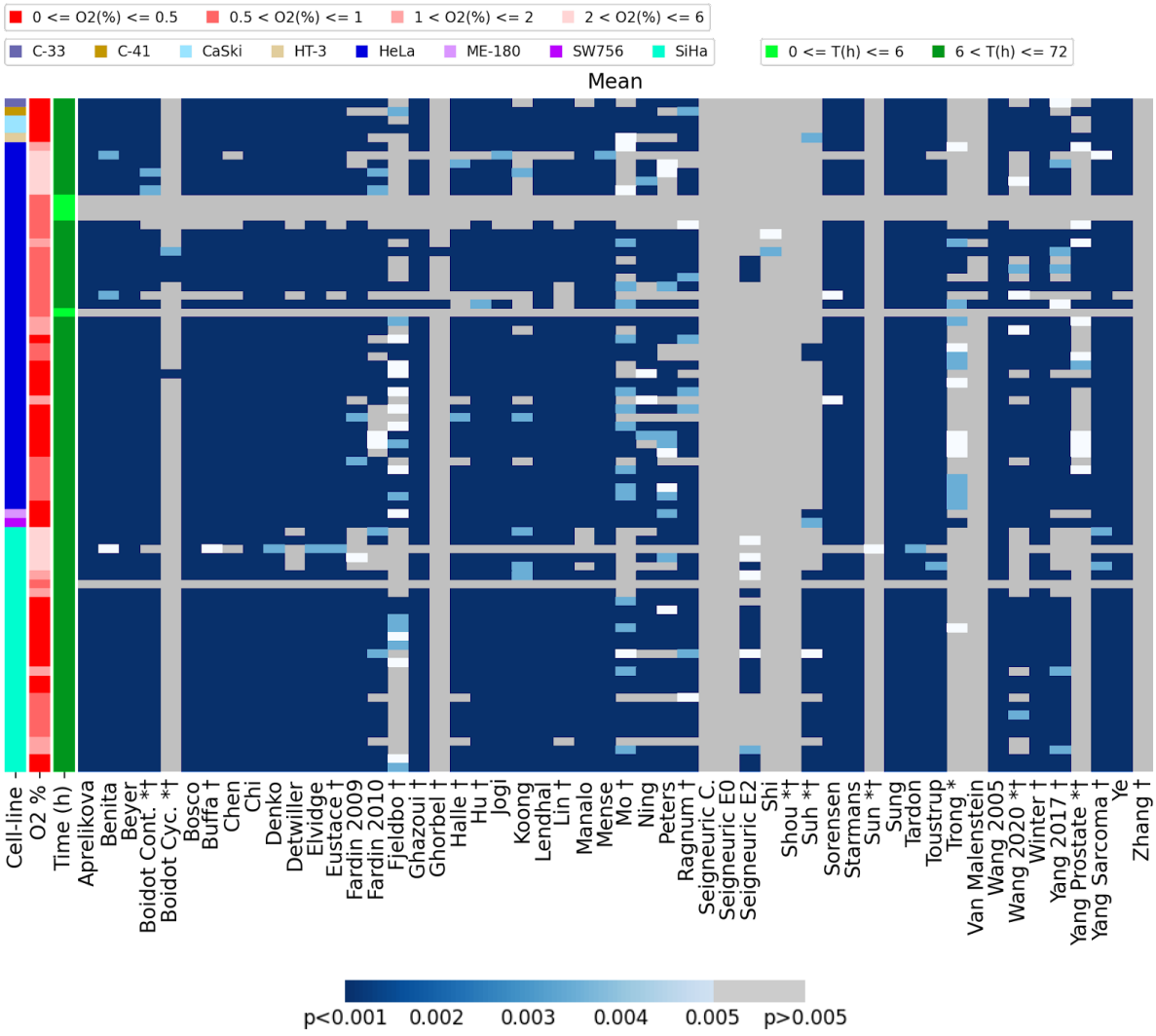
oxygen. Indeed, one can see that some hypoxia signatures, e.g. Aprelikova, can still differentiate between cells in this category.

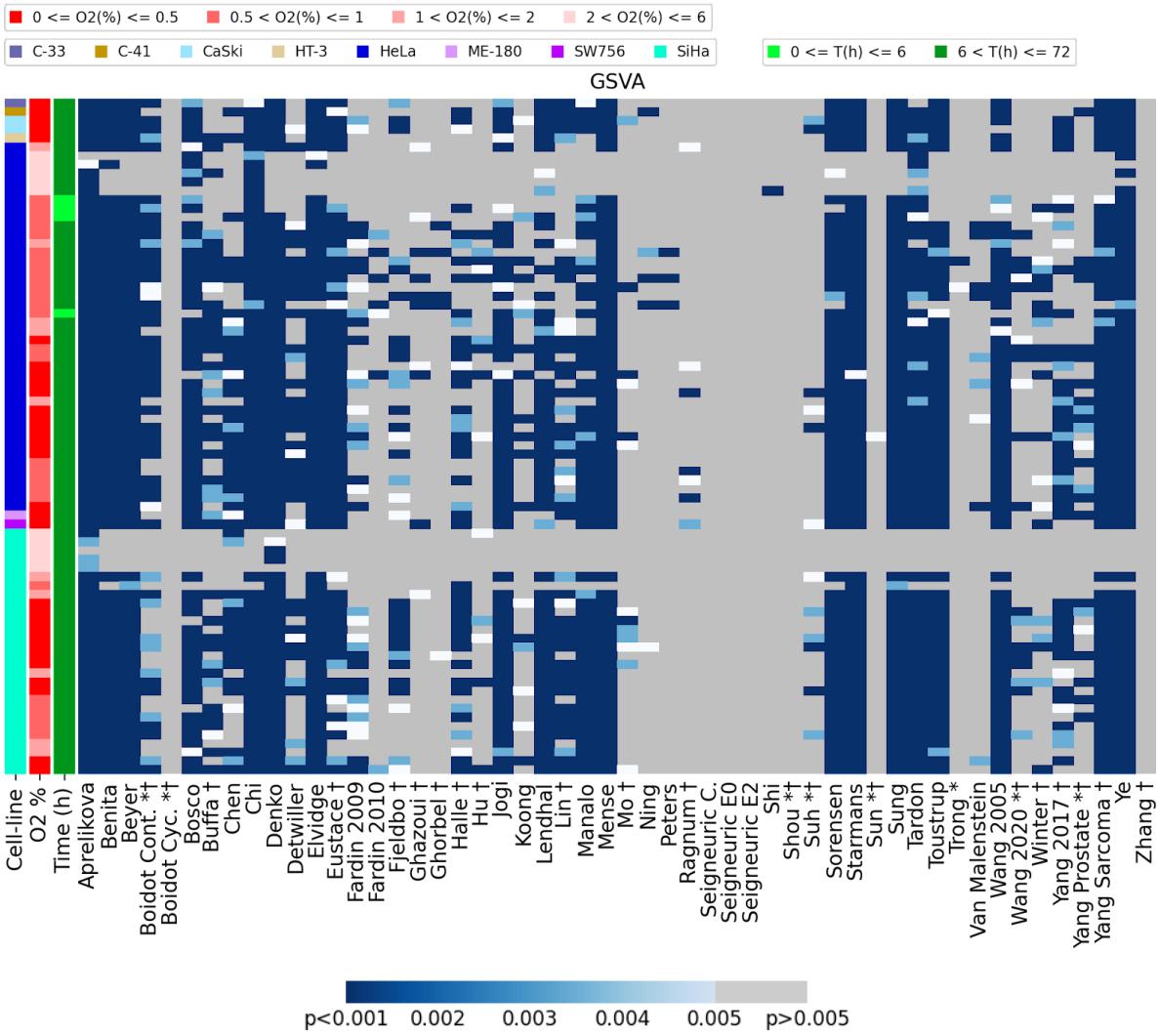
Cell line/samples	Number of Hypoxic samples	Number of Normoxic samples	Pairs
HeLa	13	14	42
SiHa	8	8	28
CaSki	2	2	2
C-33	1	1	1
C-41	1	1	1
HT-3	1	1	1
ME-180	1	1	1
SW756	1	1	1
Total	**28**	**29**	**77**

Table 4.9: *Cervical cancer cell lines in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using cervical cancer cell lines.


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
2 < O2(%) <= 6
C-33
C-41
CaSki
HT-3
HeLa
ME-180
SW756
SiHa
0 <= T(h) <= 6
6 < T(h) <= 72
Median
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
2 < O2(%) <= 6
C-33
C-41
CaSki
HT-3
HeLa
ME-180
SW756
SiHa
0 <= T(h) <= 6
6 < T(h) <= 72
Mean
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
2 < O2(%) <= 6
C-33
C-41
CaSki
HT-3
HeLa
ME-180
SW756
SiHa
0 <= T(h) <= 6
6 < T(h) <= 72
GSVA
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

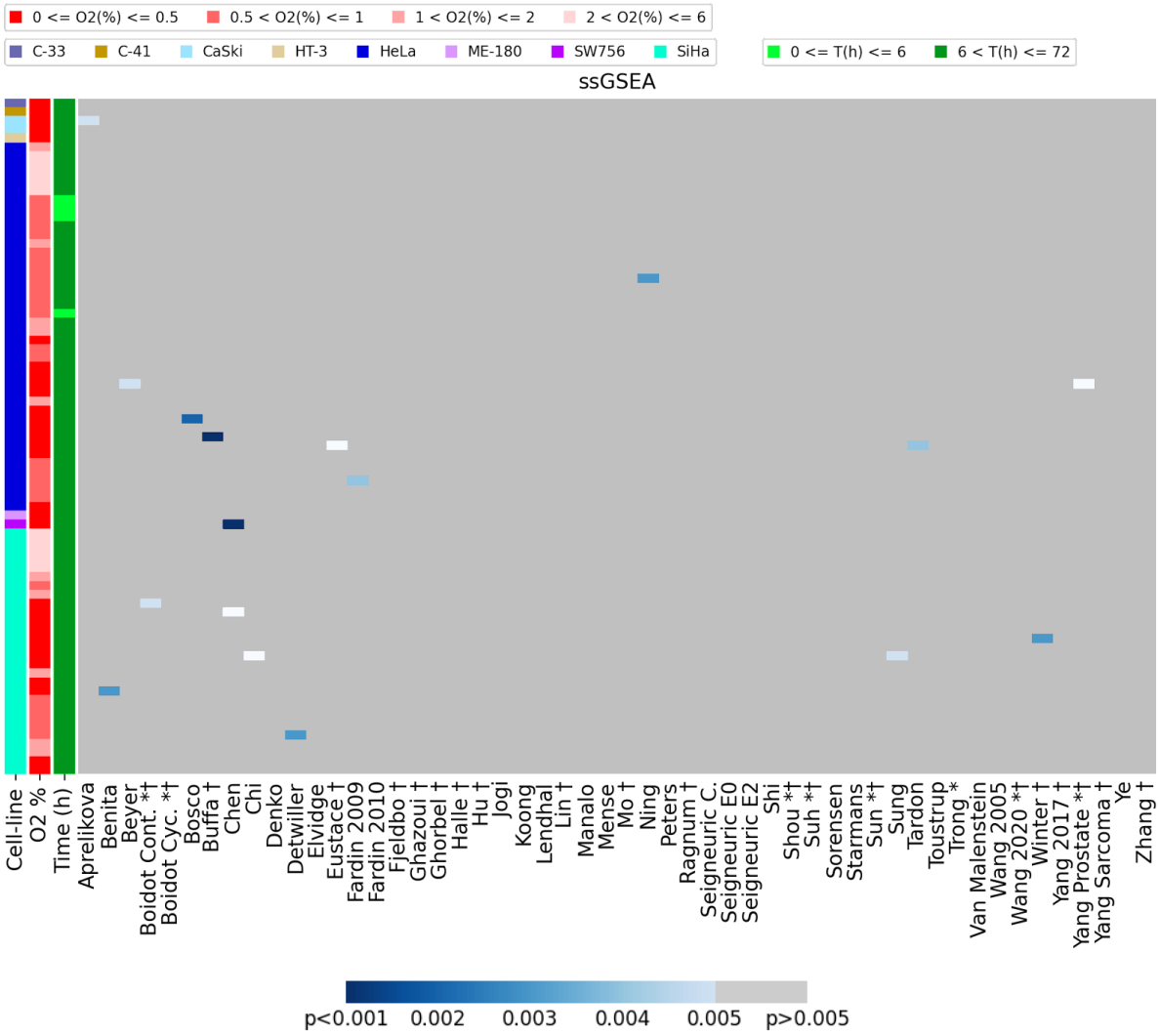


Figure 4.5: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in cervical cancer cell lines*

Comparison of the scoring methods of the 53 hypoxia signatures across all cervical cancer hypoxia experiments identified. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	89.61	93.51	97.40	1.30
Benita	93.51	93.51	87.01	1.30
Beyer	90.91	93.51	87.01	1.30
Boidot Cont. *†	93.51	89.61	64.94	1.30
Boidot Cyc. *†	10.39	6.49	0.00	0.00
Bosco	62.34	90.91	84.42	1.30
Buffa †	89.61	90.91	53.25	1.30
Chen	89.61	88.31	61.04	2.60
Chi	90.91	93.51	92.21	1.30
Denko	92.21	93.51	88.31	0.00
Detwiller	90.91	85.71	48.05	1.30
Elvidge	89.61	93.51	88.31	0.00
Eustace †	92.21	90.91	79.22	1.30
Fardin 2009	92.21	83.12	45.45	1.30
Fardin 2010	92.21	72.73	9.09	0.00
Fjeldbo †	83.12	45.45	40.26	0.00
Ghazoui †	90.91	88.31	12.99	0.00
Ghorbel †	15.58	2.60	6.49	0.00
Halle †	92.21	81.82	55.84	0.00
Hu †	76.62	89.61	18.18	0.00
Jogi	63.64	89.61	79.22	0.00
Koong	54.55	77.92	23.38	0.00
Lendhal	93.51	90.91	75.32	0.00
Lin †	93.51	84.42	57.14	0.00
Manalo	62.34	87.01	72.73	0.00
Mense	93.51	92.21	85.71	0.00
Mo †	77.92	62.34	16.88	0.00
Ning	83.12	76.62	6.49	1.30
Peters	64.94	71.43	3.90	0.00
Ragnum †	67.53	79.22	9.09	0.00
Seigneuric C.	7.79	0.00	0.00	0.00
Seigneuric E0	15.58	0.00	0.00	0.00
Seigneuric E2	24.68	27.27	0.00	0.00
Shi	20.78	2.60	1.30	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	67.53	32.47	15.58	0.00
Sorensen	93.51	90.91	84.42	0.00
Starmans	89.61	90.91	87.01	0.00
Sun *†	0.00	1.30	1.30	0.00
Sung	93.51	93.51	87.01	1.30
Tardon	93.51	93.51	87.01	1.30
Toustrup	93.51	90.91	80.52	0.00
Trong *	1.30	28.57	2.60	0.00
Van Malenstein	38.96	5.19	11.69	0.00
Wang 2005	92.21	93.51	81.82	0.00
Wang 2020 *†	76.62	64.94	24.68	0.00
Winter †	70.13	89.61	27.27	1.30
Yang 2017 †	68.83	75.32	57.14	0.00
Yang Prostate *†	40.26	32.47	29.87	1.30
Yang Sarcoma †	93.51	92.21	84.42	0.00
Ye	93.51	90.91	89.61	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	33.77	0.00	3.90	0.00

Table 4.10: *Percentage accuracy of determining hypoxic samples from normoxic samples in cervical cancer cell lines for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in T*able 4.2*.

4.6 Performance of hypoxia signatures in melanoma cancer cell lines

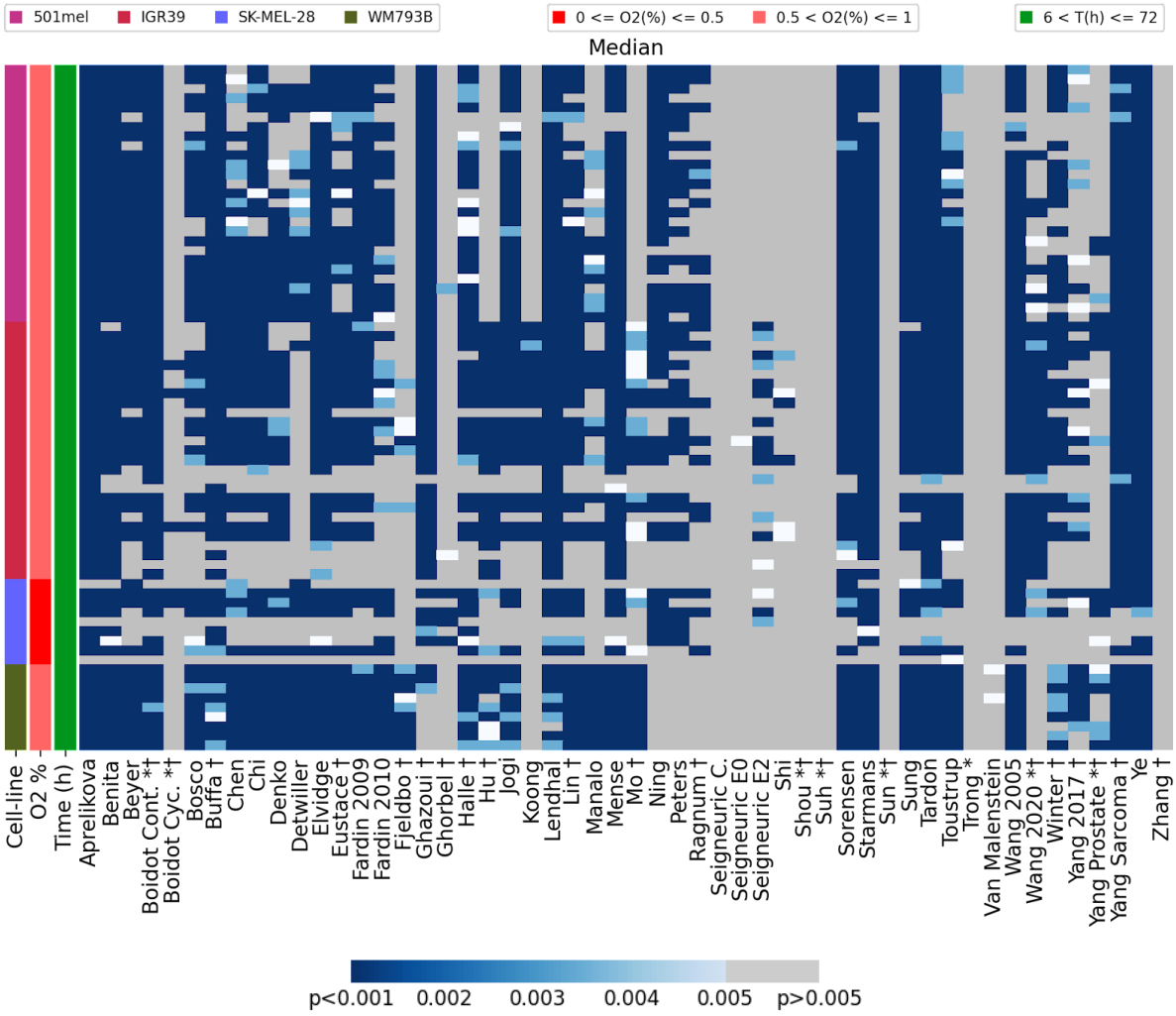
Melanoma cell lines with gene expression data in hypoxia and normoxia were identified from the Gene Expression Omnibus (as in Methods). Four such cell lines were identified: 501mel, IGR39, SK-MEL-28 and WM793B. The most common cell lines used in experiments were 501mel and IGR39 (27 pairs each, *Table 4.11*). I assessed the performance of the 53 signatures at defining hypoxia in these four cell lines in three different datasets (GSE53012, GSE85353, GSE95280). Overall, the median score performed the best (*Fig. 4.6*, top panel) and the highest performing signatures were the MCF-7 derived Aprelikova signature and Ye, which also was derived from 16 breast cancer cell lines (*Table 3.1*). Both could accurately differentiate hypoxic from normoxic samples with 95.83% accuracy (*Table 4.12*). Overall, ssGSEA was the worst performing score.

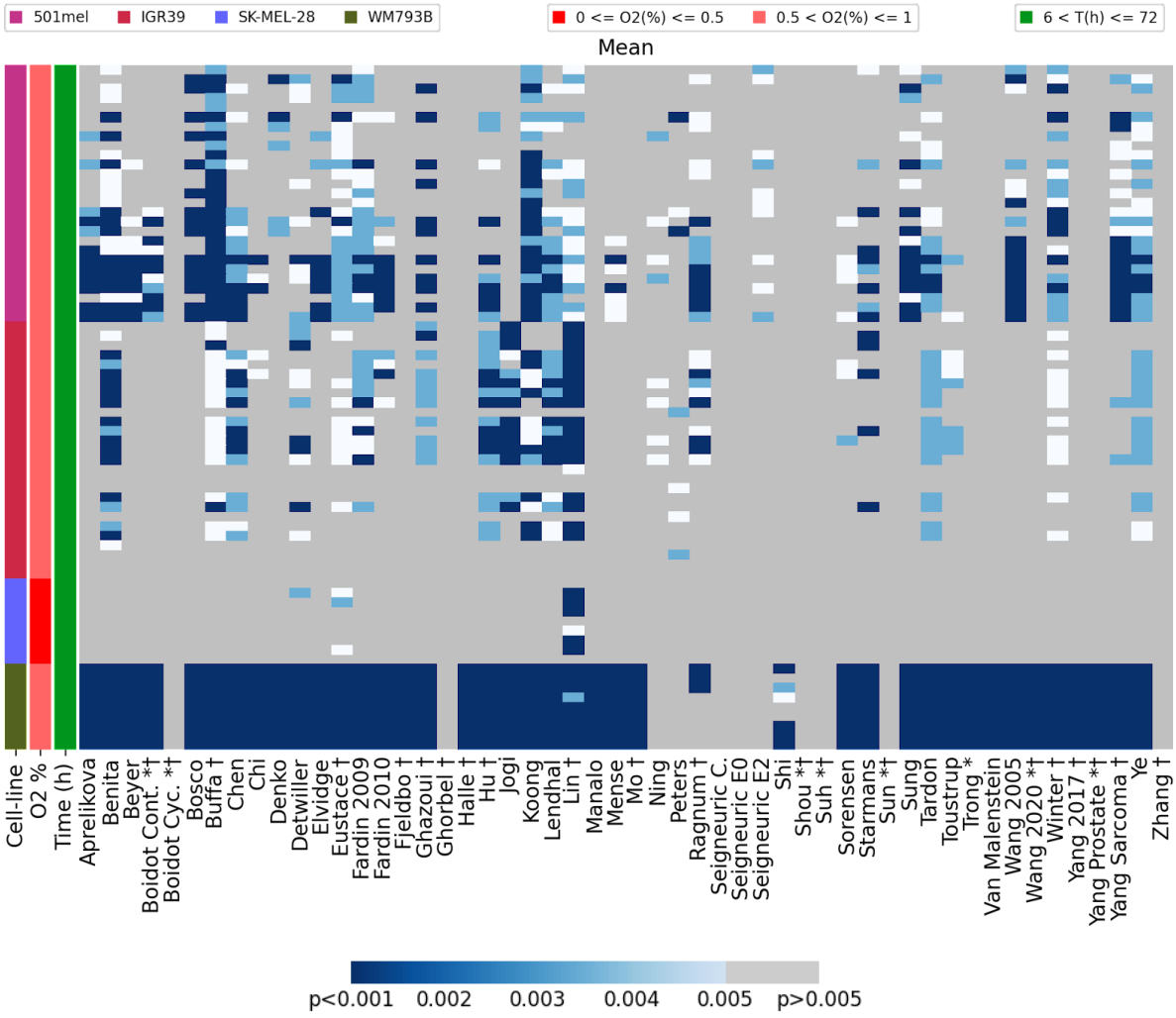
The only hypoxia signature actually derived from melanoma samples was that of Shou, who derived their 7-gene signature from the clinical samples found in The Cancer Genome Atlas (TCGA), coded as skin cutaneous melanoma (SKCM)[115]. However, this was a very poorly performing signature in identifying hypoxic melanoma cell line samples across all scores in this analysis, perhaps due to the existence of downregulated genes in the signature (*Fig. 4.6* and *Table 4.12*). The highest accuracy achieved by Shou was 1.39% using ssGSEA. This poor performance is surprising and could reflect a difference between the clinical and benchside environments, or highlight the risk of developing a signature with a relatively low number of genes. However, gene number does not seem to be the deciding factor since Fardin 2009, with only eight genes, was able to correctly classify hypoxic samples with 79.17% accuracy using the median score.

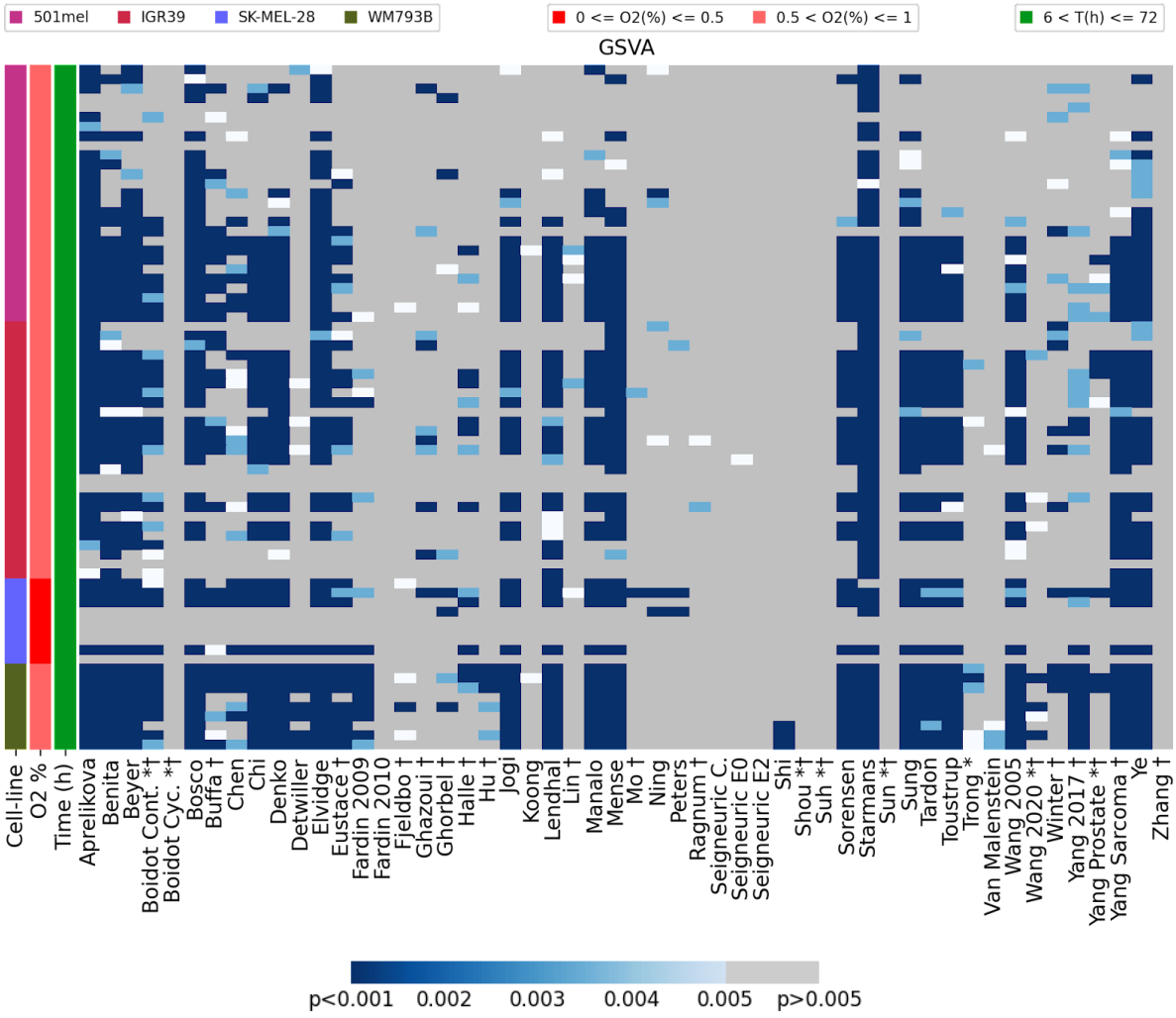
Cell line/samples	Number of Hypoxic samples	Number of Normoxic samples	Pairs
501mel	9	3	27
IGR39	9	3	27
SK-MEL-28	3	3	9
WM793B	3	3	9
Total	**24**	**12**	**72**

Table 4.11: *Melanoma cell lines in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using melanoma cell lines.


501mel
IGR39
SK-MEL-28
WM793B
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
6 < T(h) <= 72
Median
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


501mel
IGR39
SK-MEL-28
WM793B
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
6 < T(h) <= 72
Mean
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


501mel
IGR39
SK-MEL-28
WM793B
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
6 < T(h) <= 72
GSVA
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

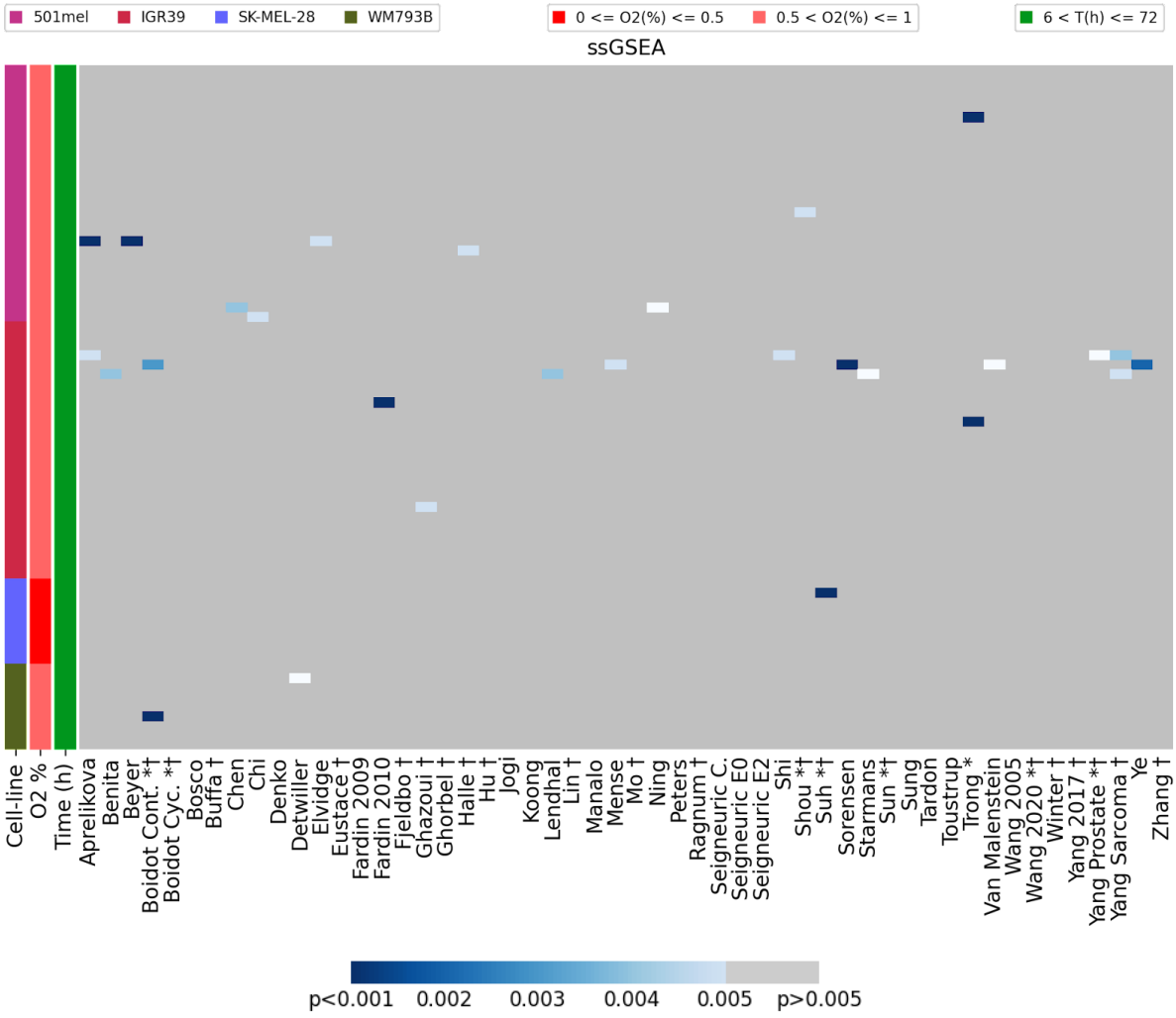


Figure 4.6: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in melanoma cell lines*

Comparison of the scoring methods of the 53 hypoxia signatures across all melanoma hypoxia experiments identified. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	95.83	29.17	83.33	2.78
Benita	91.67	68.06	69.44	1.39
Beyer	77.78	27.78	70.83	1.39
Boidot Cont. *†	91.67	27.78	52.78	2.78
Boidot Cyc. *†	6.94	0.00	0.00	0.00
Bosco	66.67	33.33	73.61	0.00
Buffa †	90.28	75.00	41.67	0.00
Chen	72.22	54.17	37.50	1.39
Chi	77.78	19.44	55.56	1.39
Denko	65.28	20.83	61.11	0.00
Detwiller	47.22	41.67	16.67	1.39
Elvidge	88.89	25.00	73.61	1.39
Eustace †	69.44	61.11	45.83	0.00
Fardin 2009	79.17	59.72	20.83	0.00
Fardin 2010	61.11	30.56	0.00	1.39
Fjeldbo †	20.83	12.50	6.94	0.00
Ghazoui †	84.72	41.67	13.89	1.39
Ghorbel †	8.33	0.00	11.11	0.00
Halle †	77.78	12.50	22.22	1.39
Hu †	45.83	50.00	6.94	0.00
Jogi	77.78	33.33	55.56	0.00
Koong	25.00	66.67	2.78	0.00
Lendhal	93.06	55.56	61.11	1.39
Lin †	75.00	73.61	6.94	0.00
Manalo	43.06	12.50	58.33	0.00
Mense	90.28	22.22	68.06	1.39
Mo †	41.67	12.50	2.78	0.00
Ning	63.89	11.11	9.72	1.39
Peters	51.39	9.72	4.17	0.00
Ragnum †	33.33	38.89	2.78	0.00
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	1.39	0.00	1.39	0.00
Seigneuric E2	23.61	13.89	0.00	0.00
Shi	8.33	8.33	4.17	1.39
Shou *†	0.00	0.00	0.00	1.39
Suh *†	0.00	0.00	0.00	1.39
Sorensen	90.28	23.61	54.17	1.39
Starmans	93.06	31.94	88.89	1.39
Sun *†	0.00	0.00	0.00	0.00
Sung	84.72	36.11	66.67	0.00
Tardon	94.44	54.17	51.39	0.00
Toustrup	80.56	25.00	52.78	0.00
Trong *	0.00	12.50	9.72	2.78
Van Malenstein	2.78	12.50	5.56	1.39
Wang 2005	88.89	36.11	58.33	0.00
Wang 2020 *†	50.00	12.50	9.72	0.00
Winter †	62.50	56.94	20.83	0.00
Yang 2017 †	41.67	12.50	36.11	0.00
Yang Prostate *†	26.39	12.50	15.28	1.39
Yang Sarcoma †	93.06	41.67	65.28	2.78
Ye	95.83	62.50	72.22	1.39
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	8.33	1.39	1.39

Table 4.12: *Percentage accuracy of determining hypoxic samples from normoxic samples in melanoma cell lines for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

4.7 Performance of hypoxia signatures in ovarian cancer cell lines

Ovarian cancer cell lines with gene expression data in hypoxia and normoxia were identified from the Gene Expression Omnibus (as in Methods). This yielded three cell lines: A2780, SKOV3ip.1 and SK-OV-3. The cell line most commonly used in experiments was A2780, an adenocarcinoma line isolated from a treatment-naive individual from Manchester, UK (*Table 4.13*). The performance of the 53 signatures at defining hypoxia was then investigated across the three cell lines (data from GEO series GSE53012, GSE66894, GSE52695). None of the published hypoxia signatures had been developed using ovarian cell lines or tissue.

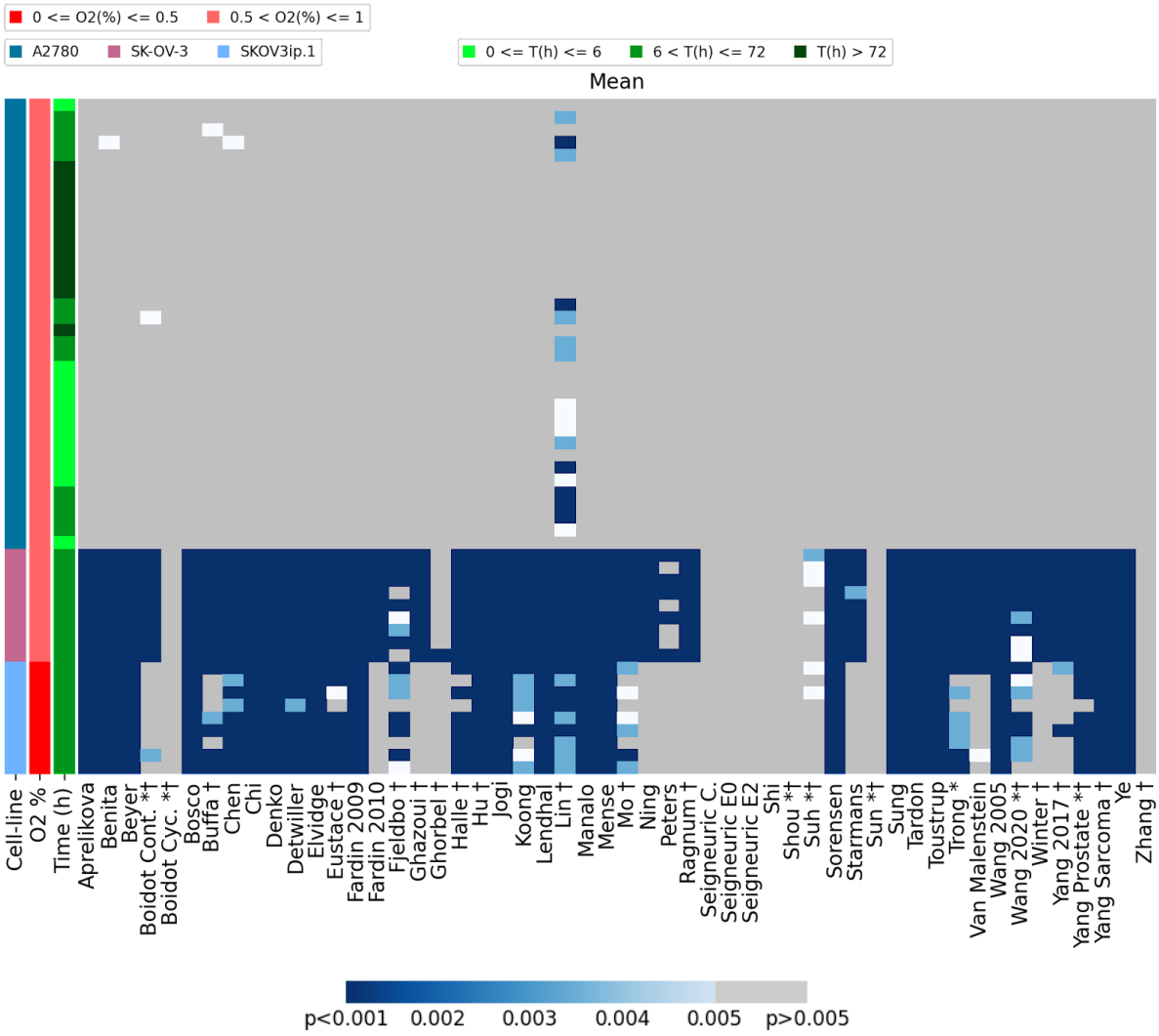
The median score was the overall best performing score (*Fig. 4.7*, top panel), with seven signatures all reaching 100% accuracy (Chen, Chi, Lendhal, Sung, Tardon, Yang Sarcoma and Ye, *Table 4.14*). This high level of accuracy was unexpected considering the origins of these signatures was away from ovarian cancer cell lines (*Table 3.1*). Six other signatures had >90% accuracy using the median score (*Table 4.14*). In addition, four signatures had >90% accuracy using GVSA (Aprelikova, Beyer, Elvidge and Sung). The worst performing signatures across all scores were Seigneuric common, Seigneuric early2, Shou, Zhang and Zou, with 0% accuracy using any of the four summary scores (*Table 4.14*). Again, ssGSEA appeared the worst performing overall score.

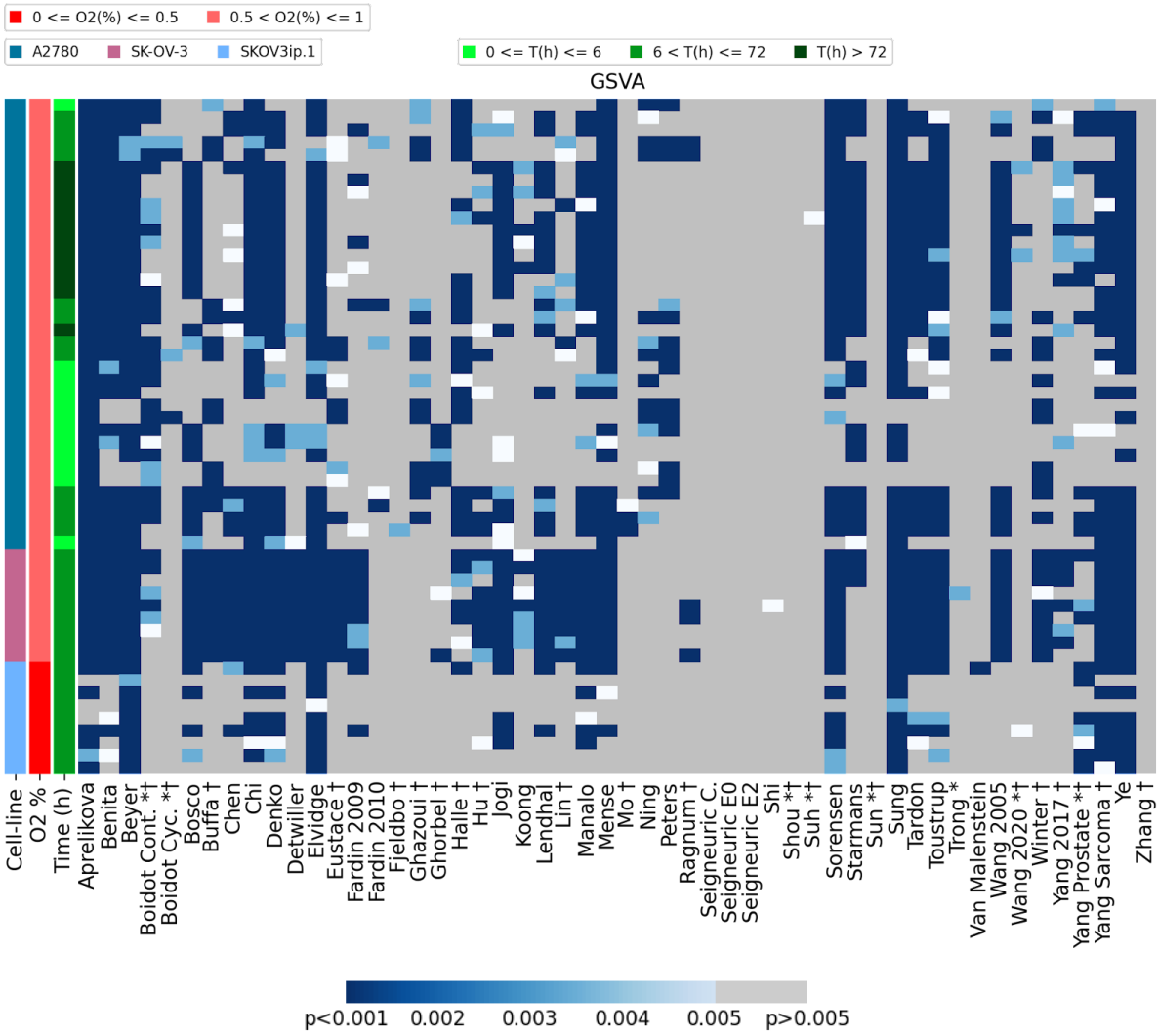
Cell line/samples	Number of Hypoxic samples	Number of Normoxic samples	Pairs
A2780	6	6	36
SKOV3ip.1	3	3	9
SK-OV-3	3	3	9
Total	**12**	**12**	**54**

Table 4.13: *Ovarian cell lines in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using ovarian cell lines.


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
A2780
SK-OV-3
SKOV3ip.1
0 <= T(h) <= 6
6 < T(h) <= 72
T(h) > 72
Median
Cell-line
S1 O2 %
S1 Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
A2780
SK-OV-3
SKOV3ip.1
0 <= T(h) <= 6
6 < T(h) <= 72
T(h) > 72
Mean
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
A2780
SK-OV-3
SKOV3ip.1
0 <= T(h) <= 6
6 < T(h) <= 72
T(h) > 72
GSVA
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

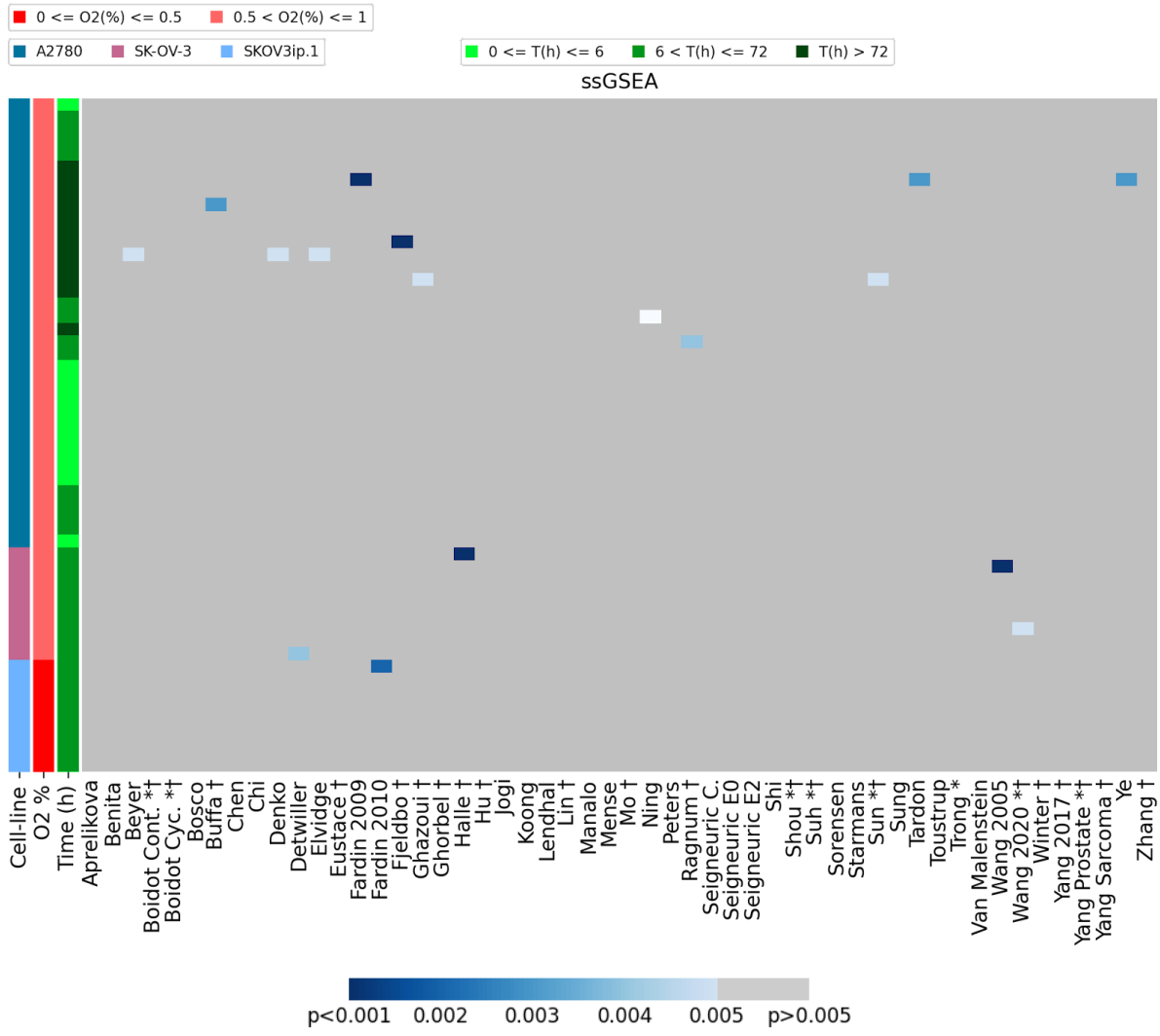


Figure 4.7: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in ovarian cancer cell lines*

Comparison of the scoring methods of the 53 hypoxia signatures across all ovarian cancer hypoxia experiments identified. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	96.30	33.33	94.44	0.00
Benita	87.04	35.19	81.48	0.00
Beyer	94.44	33.33	92.59	1.85
Boidot Cont. *†	75.93	20.37	61.11	0.00
Boidot Cyc. *†	20.37	0.00	7.41	0.00
Bosco	61.11	33.33	62.96	0.00
Buffa †	61.11	27.78	40.74	1.85
Chen	100.00	35.19	42.59	0.00
Chi	100.00	33.33	87.04	0.00
Denko	72.22	33.33	79.63	1.85
Detwiller	81.48	33.33	24.07	1.85
Elvidge	92.59	33.33	90.74	1.85
Eustace †	83.33	31.48	37.04	0.00
Fardin 2009	85.19	33.33	33.33	1.85
Fardin 2010	59.26	16.67	9.26	1.85
Fjeldbo †	33.33	25.93	1.85	1.85
Ghazoui †	40.74	16.67	27.78	1.85
Ghorbel †	7.41	1.85	12.96	0.00
Halle †	87.04	29.63	53.70	1.85
Hu †	75.93	33.33	35.19	0.00
Jogi	85.19	33.33	62.96	0.00
Koong	31.48	31.48	22.22	0.00
Lendhal	100.00	33.33	55.56	0.00
Lin †	70.37	64.81	29.63	0.00
Manalo	62.96	33.33	68.52	0.00
Mense	92.59	33.33	81.48	0.00
Mo †	50.00	27.78	5.56	0.00
Ning	57.41	16.67	25.93	1.85
Peters	33.33	9.26	25.93	0.00
Ragnum †	35.19	16.67	9.26	1.85
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	12.96	0.00	0.00	0.00
Seigneuric E2	0.00	0.00	0.00	0.00
Shi	22.22	0.00	1.85	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	1.85	11.11	1.85	0.00
Sorensen	98.15	33.33	81.48	0.00
Starmans	53.70	16.67	55.56	0.00
Sun *†	0.00	0.00	0.00	1.85
Sung	100.00	33.33	92.59	0.00
Tardon	100.00	33.33	64.81	1.85
Toustrup	96.30	33.33	72.22	0.00
Trong *	0.00	29.63	1.85	0.00
Van Malenstein	7.41	20.37	1.85	0.00
Wang 2005	85.19	33.33	55.56	1.85
Wang 2020 *†	11.11	29.63	7.41	1.85
Winter †	55.56	16.67	42.59	0.00
Yang 2017 †	9.26	20.37	33.33	0.00
Yang Prostate *†	22.22	31.48	40.74	0.00
Yang Sarcoma †	100.00	33.33	75.93	0.00
Ye	100.00	33.33	83.33	1.85
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	0.00	0.00	0.00

Table 4.14: *Percentage accuracy of determining hypoxic samples from normoxic samples in ovarian cancer cell lines for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

4.8 Performance of hypoxia signatures in cancer cell lines originating from the central nervous system

Cancer cell lines originating from the central nervous system (CNS) with gene expression data in hypoxia and normoxia were identified from the Gene Expression Omnibus (as in Methods). This yielded eight cell lines: U87 (glioma), LN229 (glioma), NCH421k (glioma), NCH601 (glioma), NCH644 (glioma), NCH660h (glioma), U373 (astrocytoma) and DAOY (medulloblastoma). The most common cell line used was U87 and the analysis was biased towards glioma because of the hypoxia experimental data available in the public domain (*Table 4.15*).

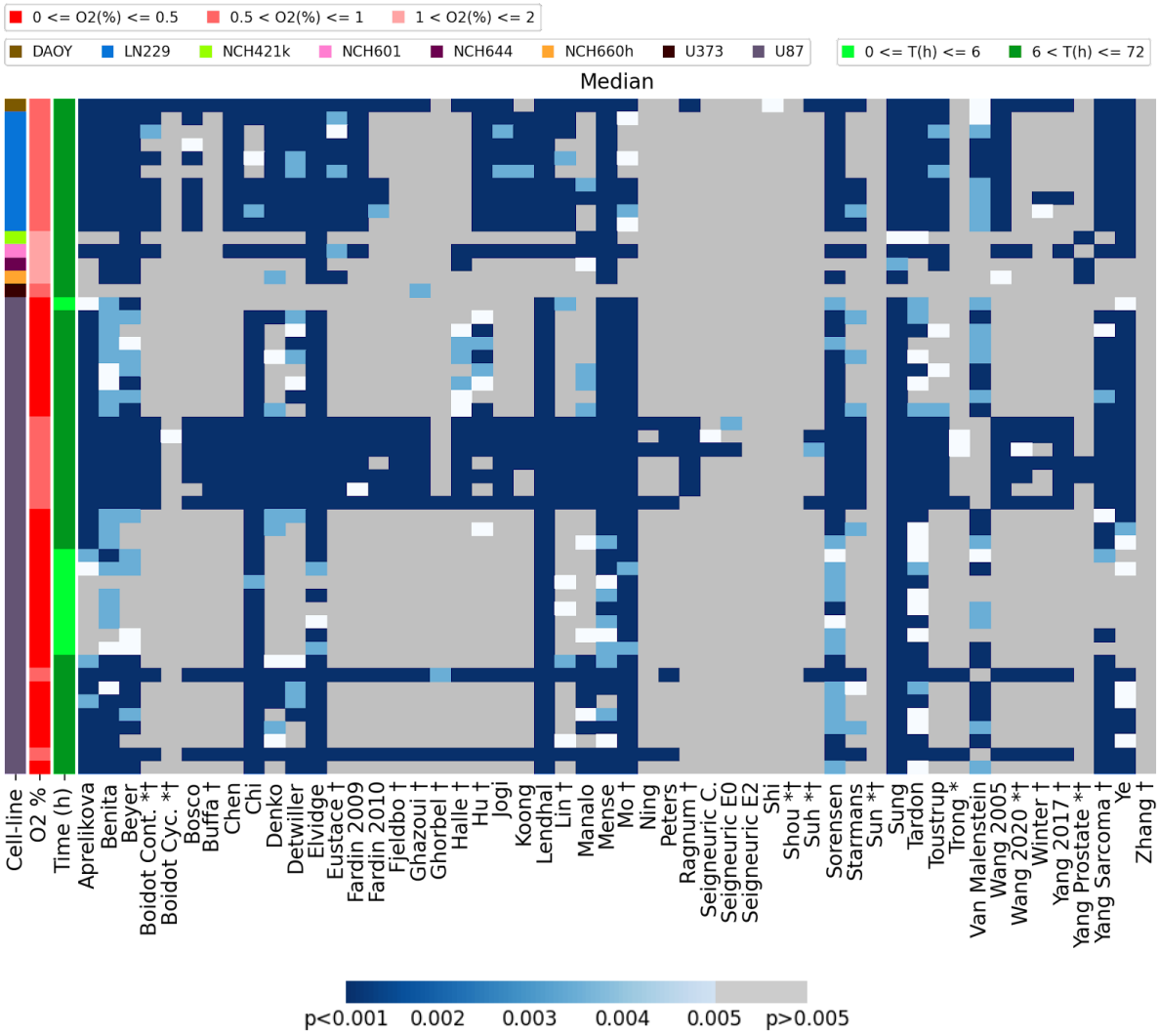
I assessed the performance of the 53 signatures at defining hypoxia in these eight CNS cell lines from six different datasets (GSE18494, GSE118683, GSE45301, GSE113353, GSE42416, GSE27523). The mean outperforms the median across the 53 signatures (*Fig. 4.8*). Impressively, several signatures reached 98.04% hypoxia identification accuracy: eight using the mean score (Benita, Beyer, Chi, Denko, Fjeldbo, Mense, Sung, Tardon and Yang Sarcoma) and one using the median (Sung) (*Table 4.16*). The Sung 90-gene signature, determined from nasopharynx carcinoma cell lines at 0.1% oxygen at a 16 hour time point, achieved 98.04% accuracy with both mean and median scores. The accuracy does not reach 100% in these signatures as the only pair derived from U373 cells (glioblastoma) is incorrectly classified (*Fig. 4.8*, third panel).

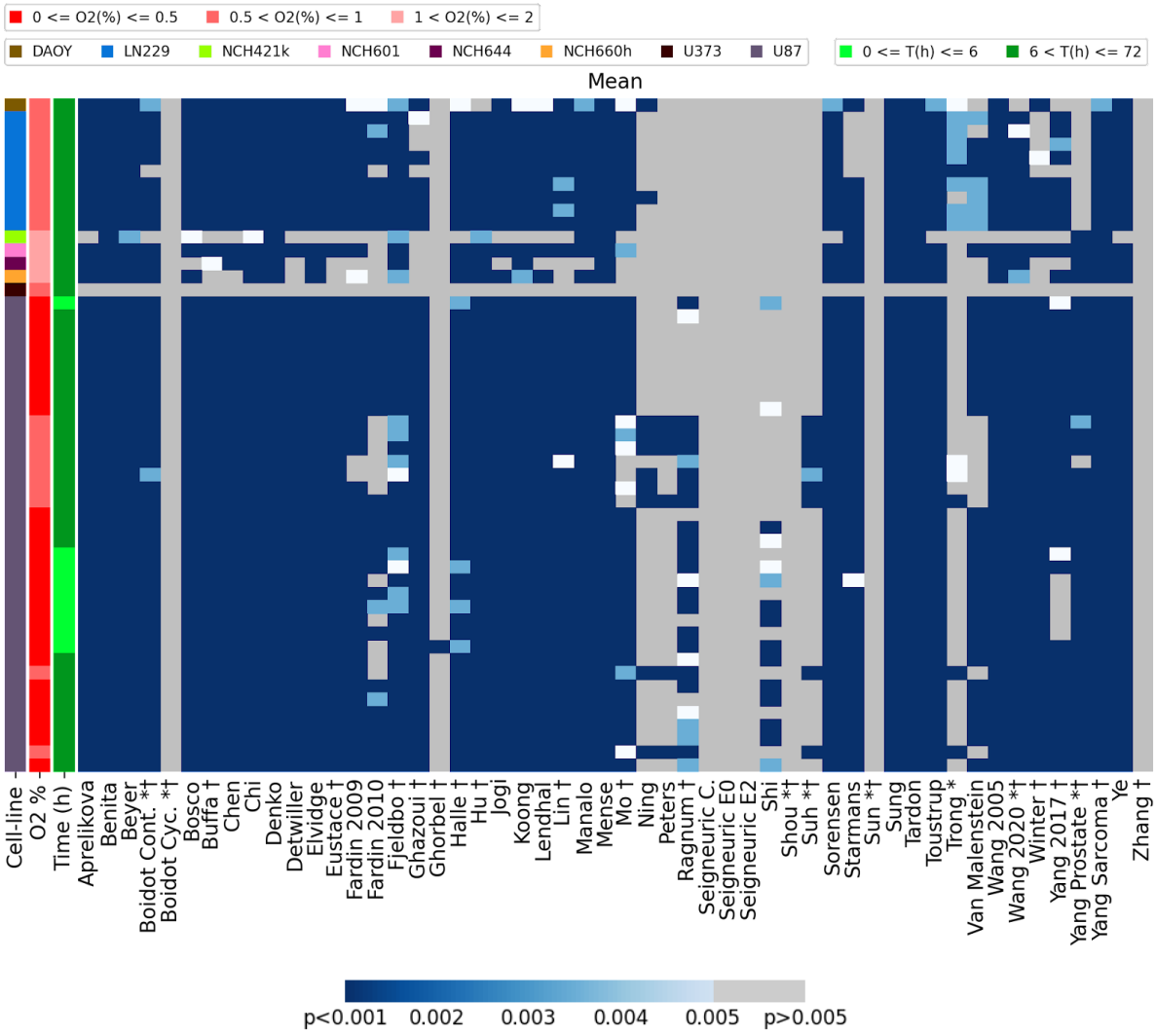
The worst performing signature was the three gene signature from Zhang. Zhang was derived from liver cancer cell lines (HepG2 and Huh-7) and had 0% accuracy in identifying hypoxia across all scores (*Table 4.16*). As commonly observed above, ssGSEA was the worst at discrimination.

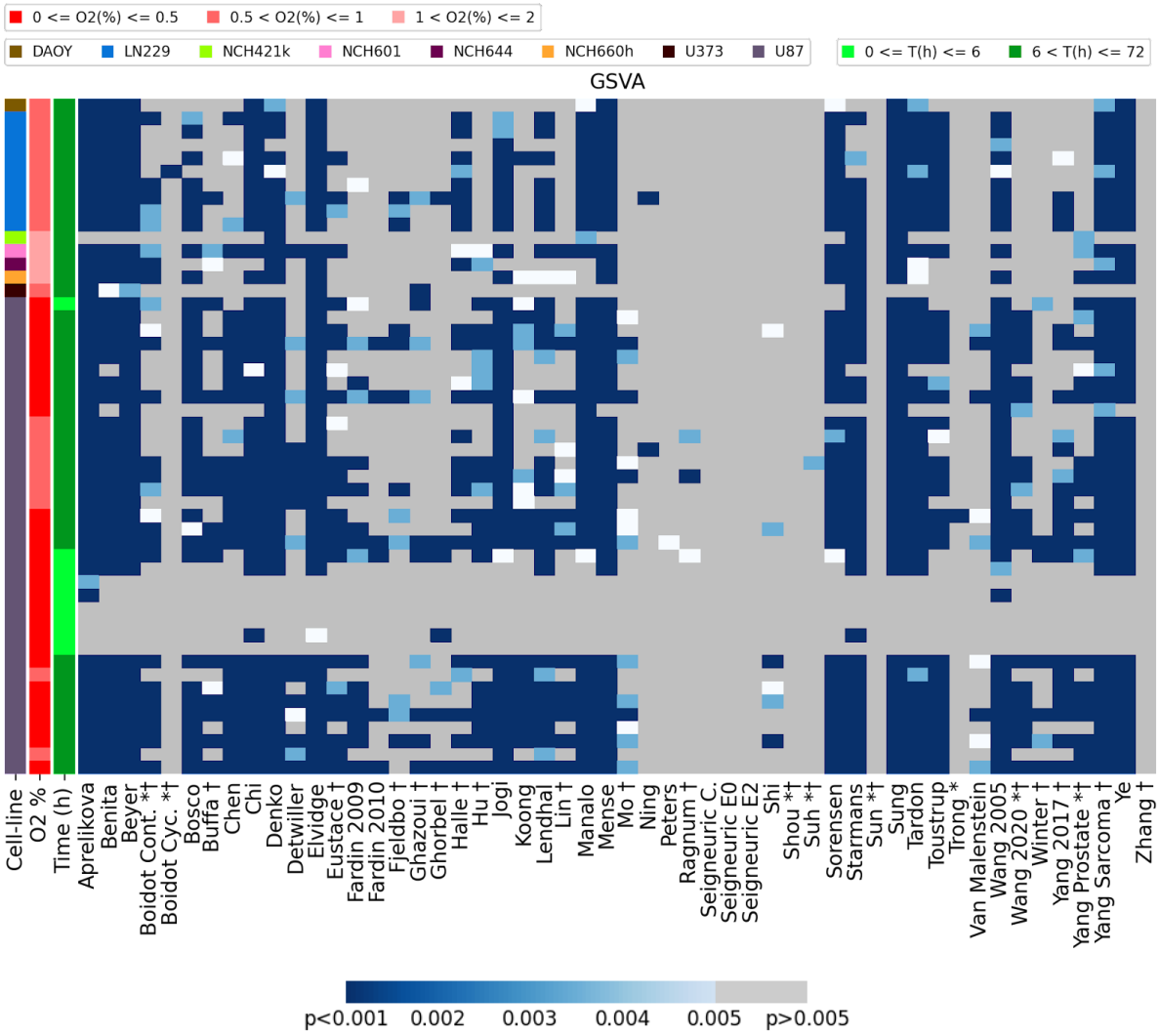
Cell line/samples	Number of Hypoxic samples	Number of Normoxic samples	Pairs
U87	12	6	36
LN229	3	3	9
NCH421k	1	1	1
NCH601	1	1	1
NCH644	1	1	1
NCH660h	1	1	1
U373	1	1	1
DAOY	1	1	1
Total	**21**	**15**	**51**

Table 4.15: *Cancer cell lines originating from the central nervous system used in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using cancer cell lines originating from the central nervous system.


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
DAOY
LN229
NCH421k
NCH601
NCH644
NCH660h
U373
U87
0 <= T(h) <= 6
6 < T(h) <= 72
Median
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
DAOY
LN229
NCH421k
NCH601
NCH644
NCH660h
U373
U87
0 <= T(h) <= 6
6 < T(h) <= 72
Mean
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
DAOY
LN229
NCH421k
NCH601
NCH644
NCH660h
U373
U87
0 <= T(h) <= 6
6 < T(h) <= 72
GSVA
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

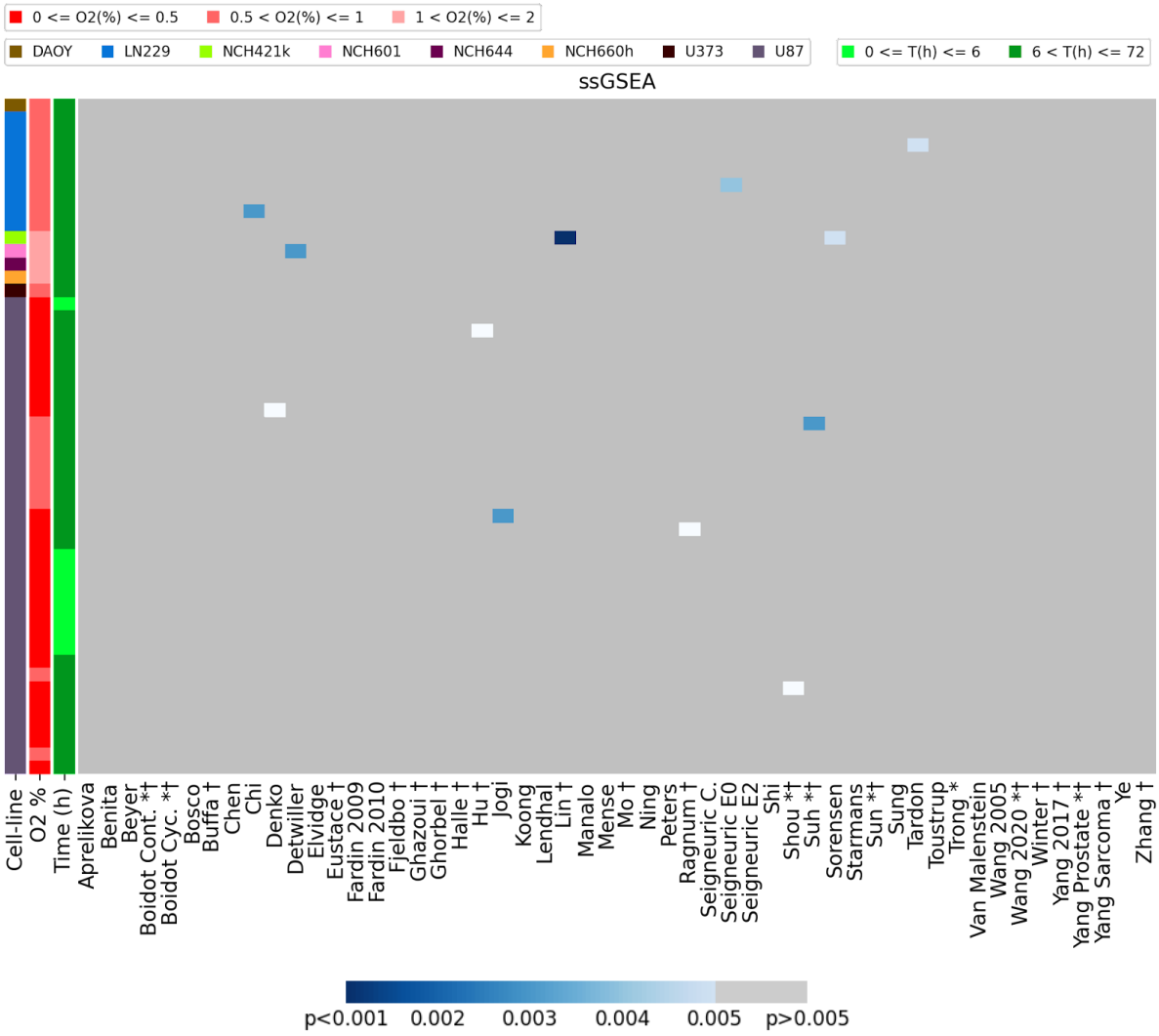


Figure 4.8: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in cancer cell lines originating from the central nervous system*

Comparison of the scoring methods of the 53 hypoxia signatures across all hypoxia experiments identified which included cancer cell lines originating from the central nervous system. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	80.39	96.08	90.20	0.00
Benita	92.16	98.04	82.35	0.00
Beyer	82.35	98.04	86.27	0.00
Boidot Cont. *†	35.29	94.12	50.98	0.00
Boidot Cyc. *†	3.92	0.00	1.96	0.00
Bosco	31.37	96.08	68.63	0.00
Buffa †	19.61	94.12	35.29	0.00
Chen	39.22	94.12	54.90	0.00
Chi	84.31	98.04	82.35	1.96
Denko	60.78	98.04	84.31	1.96
Detwiller	62.75	92.16	33.33	1.96
Elvidge	92.16	96.08	88.24	0.00
Eustace †	41.18	92.16	47.06	0.00
Fardin 2009	39.22	90.20	33.33	0.00
Fardin 2010	23.53	66.67	11.76	0.00
Fjeldbo †	19.61	98.04	23.53	0.00
Ghazoui †	21.57	84.31	21.57	0.00
Ghorbel †	5.88	1.96	13.73	0.00
Halle †	35.29	96.08	49.02	0.00
Hu †	50.98	96.08	45.10	1.96
Jogi	39.22	94.12	78.43	1.96
Koong	33.33	96.08	41.18	0.00
Lendhal	92.16	94.12	64.71	0.00
Lin †	45.10	92.16	33.33	1.96
Manalo	50.98	96.08	80.39	0.00
Mense	96.08	98.04	82.35	0.00
Mo †	86.27	86.27	29.41	0.00
Ning	7.84	19.61	3.92	0.00
Peters	11.76	11.76	1.96	0.00
Ragnum †	13.73	50.98	5.88	1.96
Seigneuric C.	3.92	0.00	0.00	0.00
Seigneuric E0	3.92	0.00	0.00	1.96
Seigneuric E2	0.00	0.00	0.00	0.00
Shi	1.96	29.41	11.76	0.00
Shou *†	0.00	0.00	0.00	1.96
Suh *†	9.80	17.65	1.96	1.96
Sorensen	94.12	96.08	78.43	1.96
Starmans	41.18	88.24	78.43	0.00
Sun *†	0.00	0.00	0.00	0.00
Sung	98.04	98.04	86.27	0.00
Tardon	90.20	98.04	80.39	1.96
Toustrup	47.06	96.08	64.71	0.00
Trong *	9.80	27.45	1.96	0.00
Van Malenstein	66.67	70.59	17.65	0.00
Wang 2005	41.18	96.08	82.35	0.00
Wang 2020 *†	19.61	94.12	39.22	0.00
Winter †	17.65	84.31	13.73	0.00
Yang 2017 †	23.53	80.39	45.10	0.00
Yang Prostate *†	9.80	76.47	43.14	0.00
Yang Sarcoma †	78.43	98.04	84.31	0.00
Ye	78.43	96.08	82.35	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	54.90	11.76	0.00

Table 4.16: *Percentage accuracy of determining hypoxic samples from normoxic samples in cancer cell lines originating from the central nervous system cancer for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

4.9 Performance of hypoxia signatures in prostate cancer cell lines

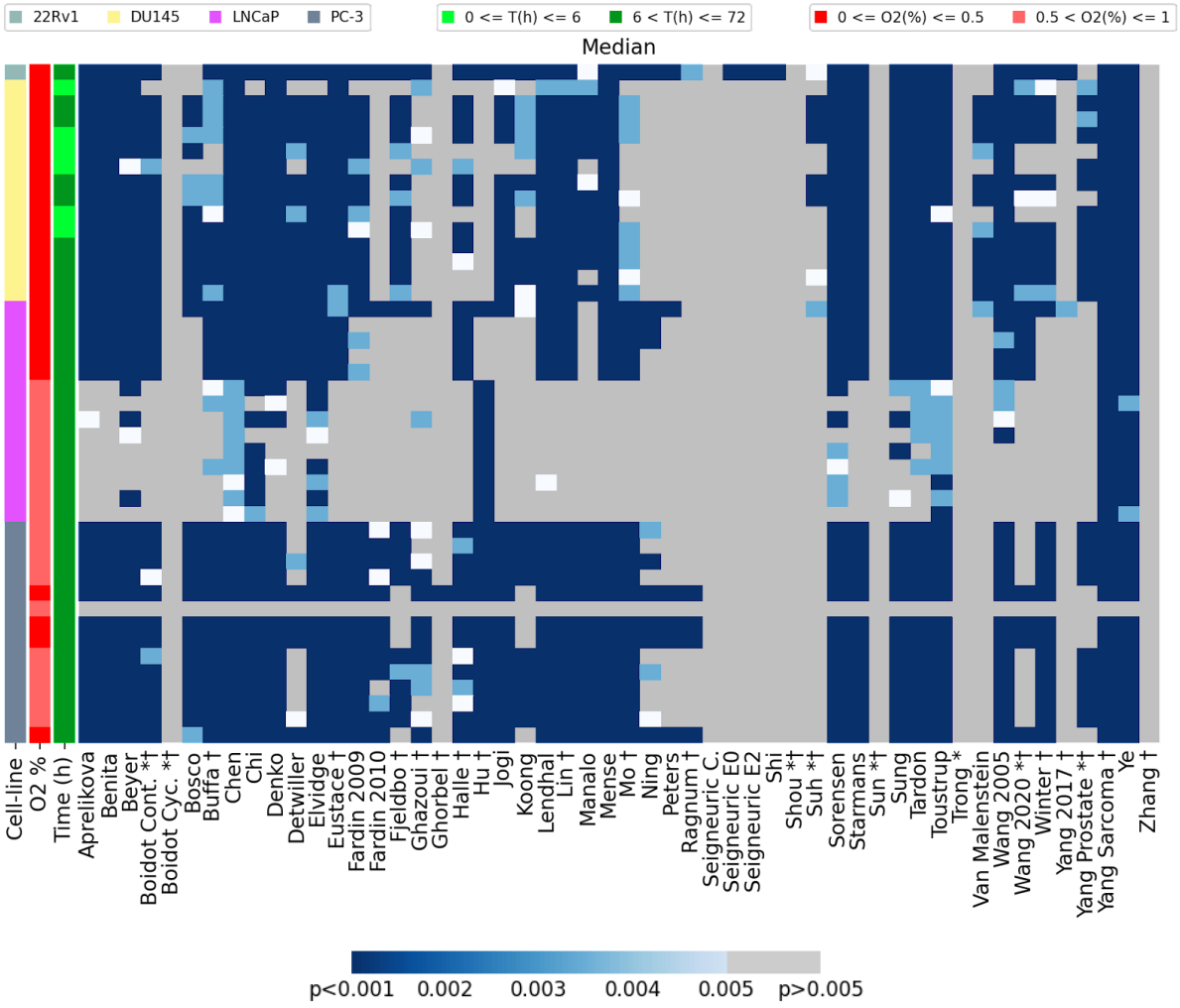
Differential gene expression profiles in hypoxic prostate cancer cell lines were identified from the Gene Expression Omnibus (as in Methods). These included the following four cell lines: (i) PC-3 (prostate carcinoma, derived from bone metastasis) (ii) DU145 (prostate carcinoma, derived from brain metastases) (iii) LNCaP (prostate carcinoma, derived from the left supraclavicular lymph node metastasis) (iv) 22Rv1 (prostate carcinoma, derived from a xenograft that was serially propagated in mice after castration-induced regression). DU145, PC-3, LNCaPI comprise approximately a third each of the pairwise combinations in the analysis, with 22Rv1 adding only one pair (*Table 4.17*). The performance of the 53 signatures in defining hypoxia in these four cell lines was then assessed.

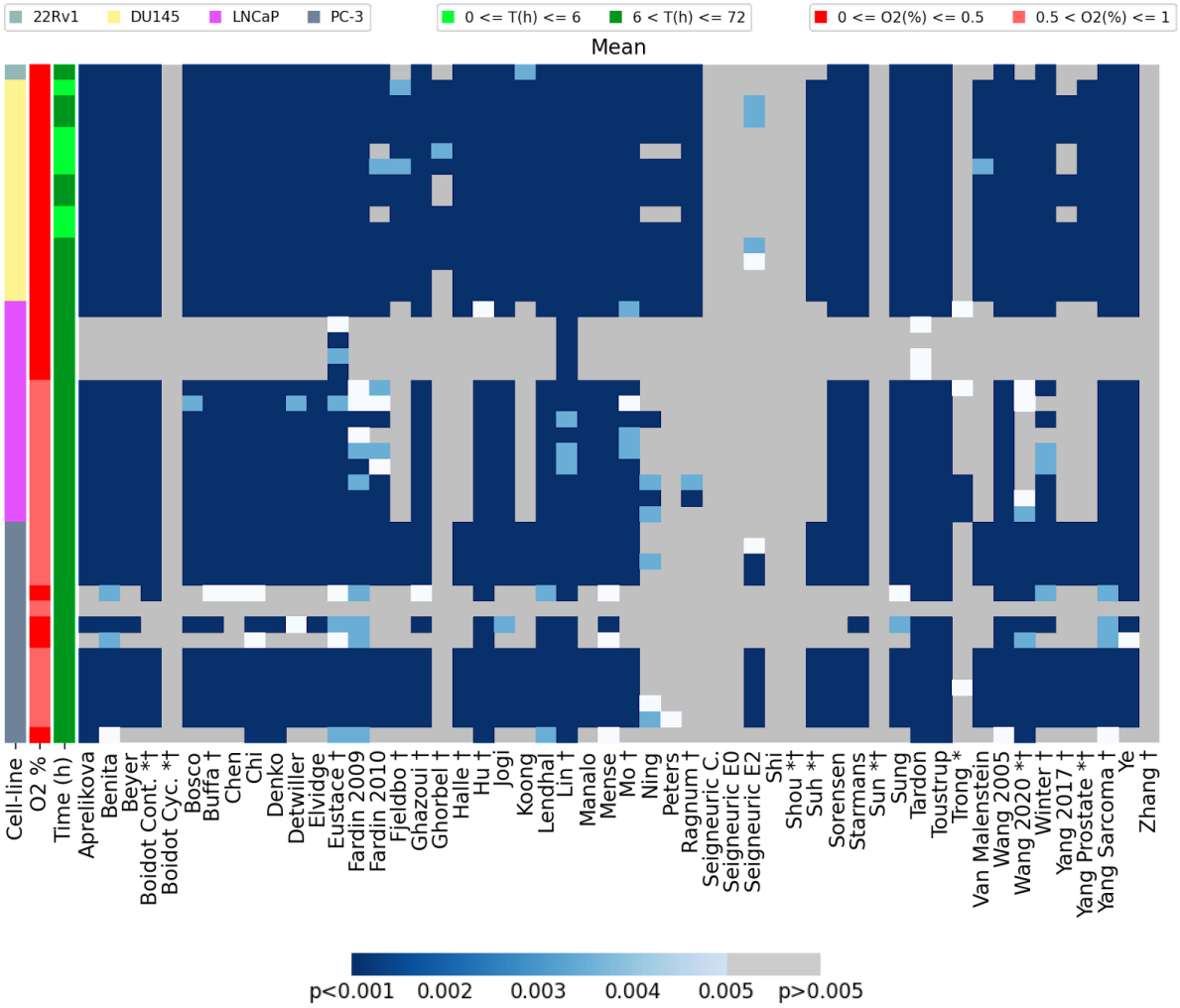
Three scores (median, mean and GSVA) overall performed well in identifying hypoxia, although, as previously observed, ssGSEA performed poorly irrespective of signature tested (*Fig. 4.9*). The highest accuracy was achieved using both median and mean scores in any one of six signatures (97.67% accuracy with Chen, Toustrup, Yang Sarcoma and Ye on median; Lin, Eustace on mean, highlighted red in *Table 4.18*). Interestingly, none of these six signatures were developed from prostate cancer cell lines. The two signatures derived solely from prostate cancer cell lines (Ragnum and Yang Prostate) only reached a maximum of 53.49% accuracy (Yang Prostate: Mean). The worst performing signatures had 0% accuracy on all summary scores (Seigneuric common, Sun and Zhang).

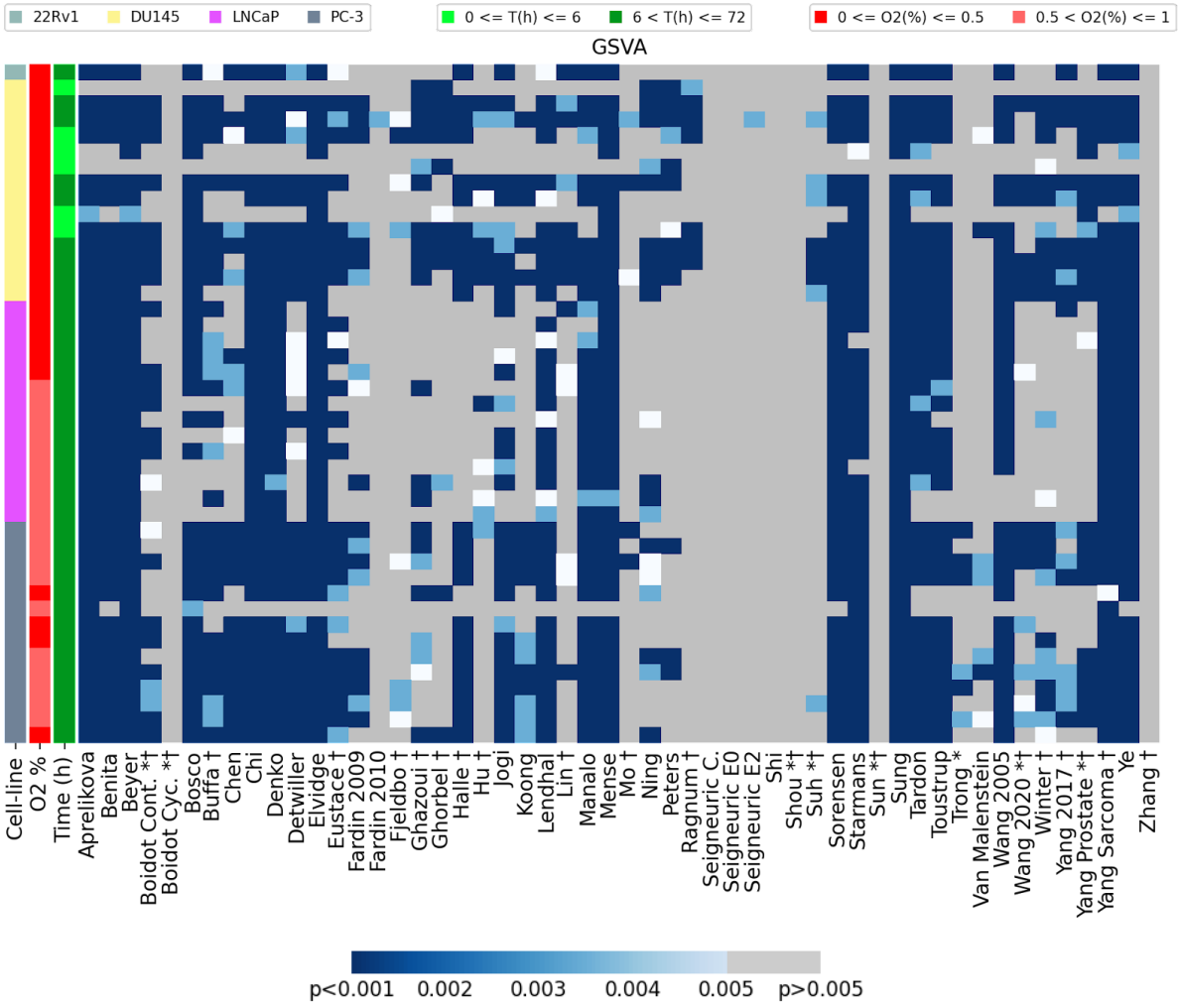
Cell line	Number of Hypoxic samples	Number of Normoxic samples	Pairs
DU145	14	2	14
PC-3	16	6	14
LNCaP	6	6	14
22Rv1	1	1	1
Total	**27**	**16**	**43**

Table 4.17: *Prostate cancer cell lines in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using prostate cancer cancer cell lines.


22Rv1
DU145
LNCaP
PC-3
0 <= T(h) <= 6
6 < T(h) <= 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
Median
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


22Rv1
DU145
LNCaP
PC-3
0 <= T(h) <= 6
6 < T(h) <= 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
Mean
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


22Rv1
DU145
LNCaP
PC-3
0 <= T(h) <= 6
6 < T(h) <= 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
GSVA
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

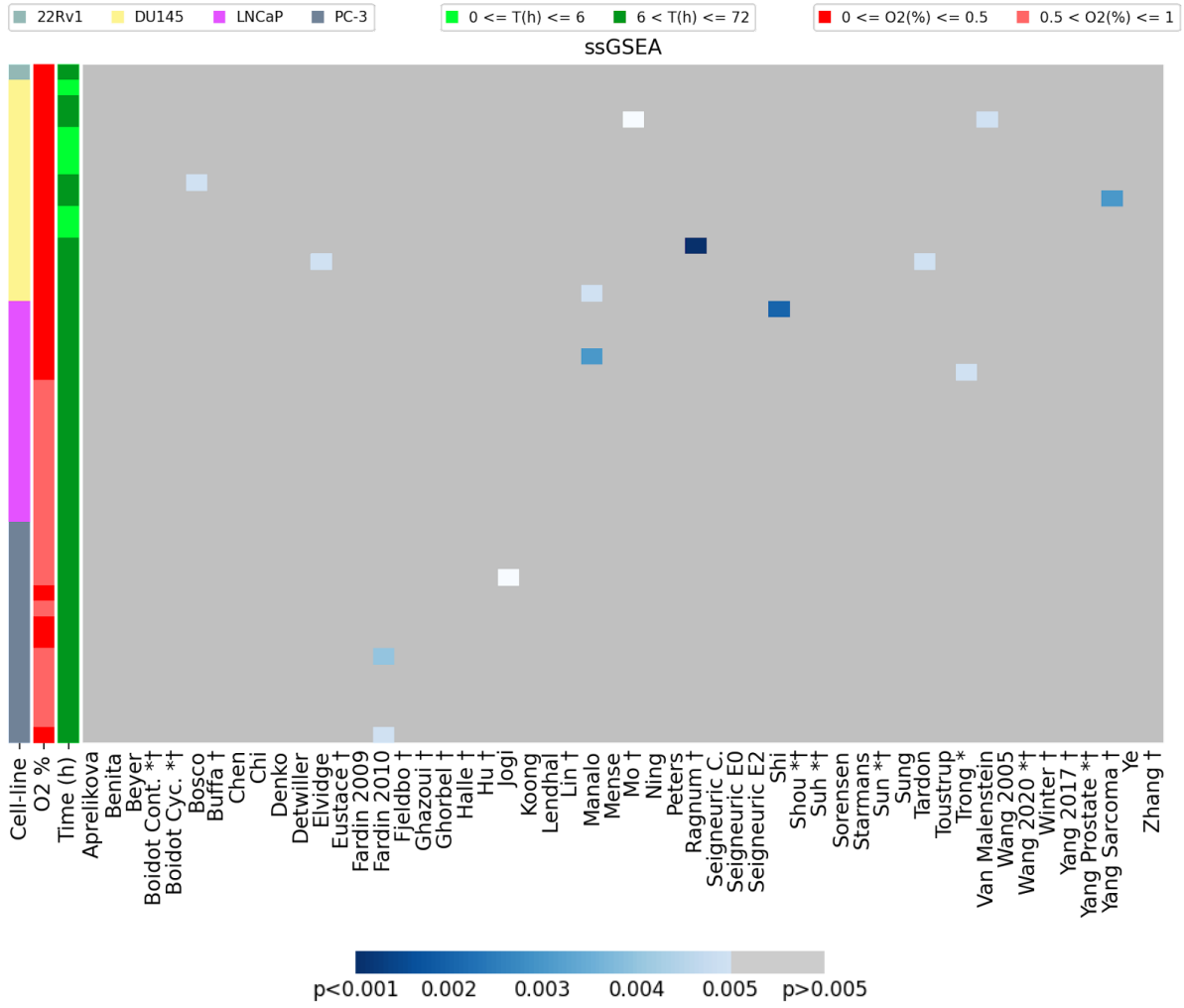


Figure 4.9: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in prostate cancer*

Comparison of the scoring methods of the 53 hypoxia signatures across all prostate cancer hypoxia experiments identified. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	79.07	83.72	93.02	0.00
Benita	76.74	88.37	88.37	0.00
Beyer	86.05	81.40	95.35	0.00
Boidot Cont. *†	74.42	81.40	55.81	0.00
Boidot Cyc. *†	0.00	0.00	0.00	0.00
Bosco	60.47	81.40	81.40	2.33
Buffa †	79.07	83.72	67.44	0.00
Chen	97.67	81.40	53.49	0.00
Chi	90.70	88.37	88.37	0.00
Denko	86.05	83.72	88.37	0.00
Detwiller	55.81	81.40	69.77	0.00
Elvidge	93.02	81.40	93.02	2.33
Eustace †	76.74	97.67	62.79	0.00
Fardin 2009	69.77	88.37	44.19	0.00
Fardin 2010	32.56	72.09	2.33	4.65
Fjeldbo †	53.49	53.49	18.60	0.00
Ghazoui †	41.86	83.72	48.84	0.00
Ghorbel †	2.33	23.26	27.91	0.00
Halle †	65.12	58.14	53.49	0.00
Hu †	55.81	88.37	30.23	0.00
Jogi	62.79	81.40	72.09	2.33
Koong	48.84	55.81	37.21	0.00
Lendhal	79.07	88.37	86.05	0.00
Lin †	76.74	97.67	32.56	0.00
Manalo	60.47	79.07	83.72	4.65
Mense	76.74	88.37	93.02	0.00
Mo †	65.12	79.07	11.63	2.33
Ning	27.91	48.84	46.51	0.00
Peters	13.95	34.88	30.23	0.00
Ragnum †	11.63	41.86	16.28	2.33
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	2.33	0.00	0.00	0.00
Seigneuric E2	2.33	27.91	2.33	0.00
Shi	2.33	0.00	0.00	2.33
Shou *†	0.00	0.00	0.00	0.00
Suh *†	18.60	53.49	18.60	0.00
Sorensen	90.70	79.07	88.37	0.00
Starmans	76.74	81.40	90.70	0.00
Sun *†	0.00	0.00	0.00	0.00
Sung	86.05	83.72	95.35	0.00
Tardon	88.37	95.35	83.72	2.33
Toustrup	97.67	88.37	79.07	0.00
Trong *	0.00	13.95	13.95	2.33
Van Malenstein	27.91	53.49	16.28	2.33
Wang 2005	86.05	88.37	81.40	0.00
Wang 2020 *†	48.84	74.42	32.56	0.00
Winter †	60.47	79.07	55.81	0.00
Yang 2017 †	4.65	44.19	39.53	0.00
Yang Prostate *†	48.84	53.49	51.16	0.00
Yang Sarcoma †	97.67	88.37	90.70	2.33
Ye	97.67	86.05	93.02	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	18.60	0.00	2.33

Table 4.18: *Percentage accuracy of determining hypoxic samples from normoxic samples in prostate cancer for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

4.10 Performance of hypoxia signatures in pancreatic cancer cell lines

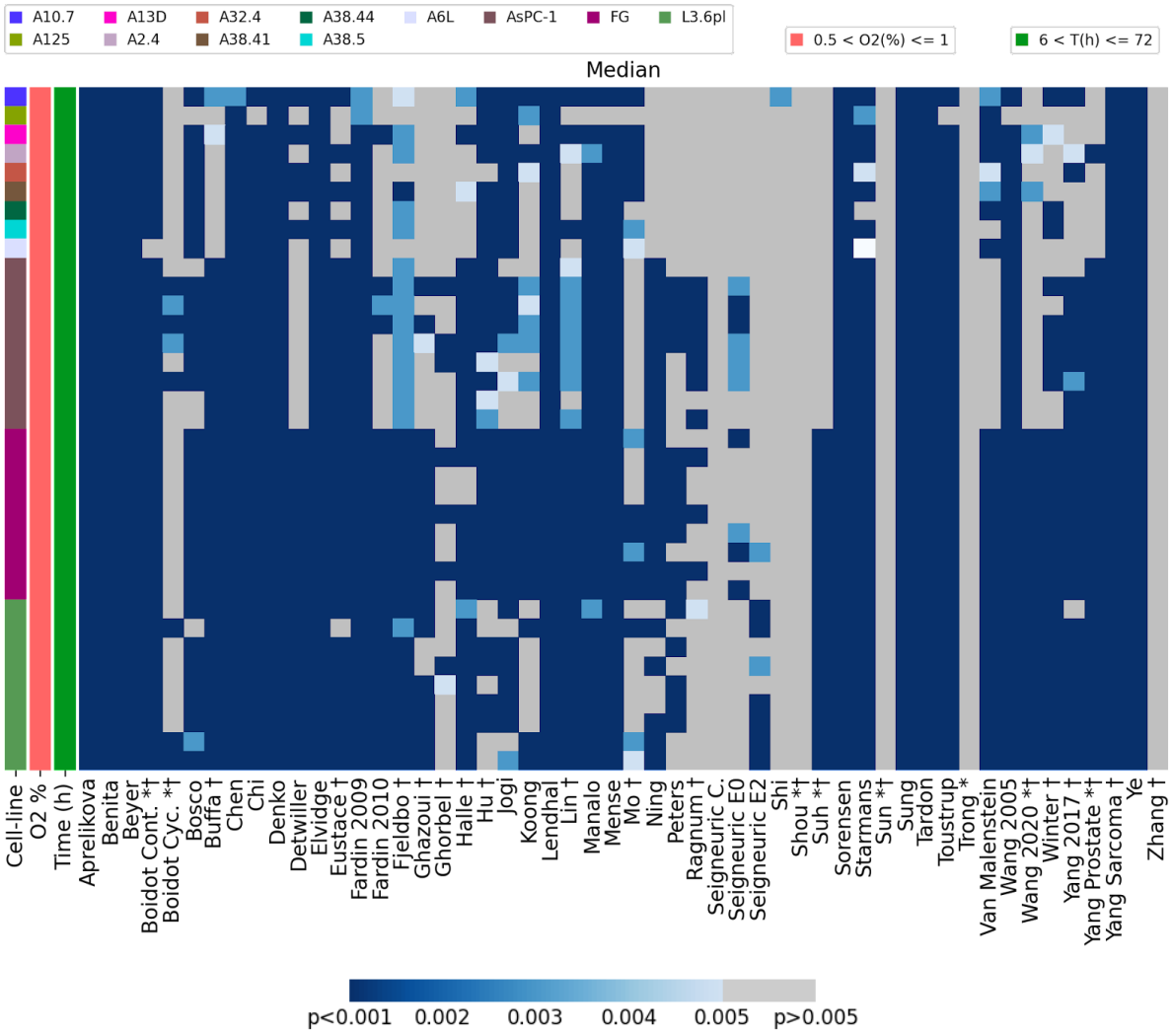
Pancreatic cancer cell lines with gene expression data in hypoxia and normoxia were identified from the Gene Expression Omnibus (as in Methods). This revealed 12 cell lines. Three cell lines, AsPC-1 (ductal adenocarcinoma), FG (adenosquamous carcinoma) and L3.6pl (adenosquamous carcinoma), contributed nine pairwise combinations to the analysis, whereas the other nine lines contributed only one pair each (*Table 4.19*). The performance of the 53 signatures at defining hypoxia in these 3 cell lines was assessed (GSE9350, GSE139673, GSE67549).

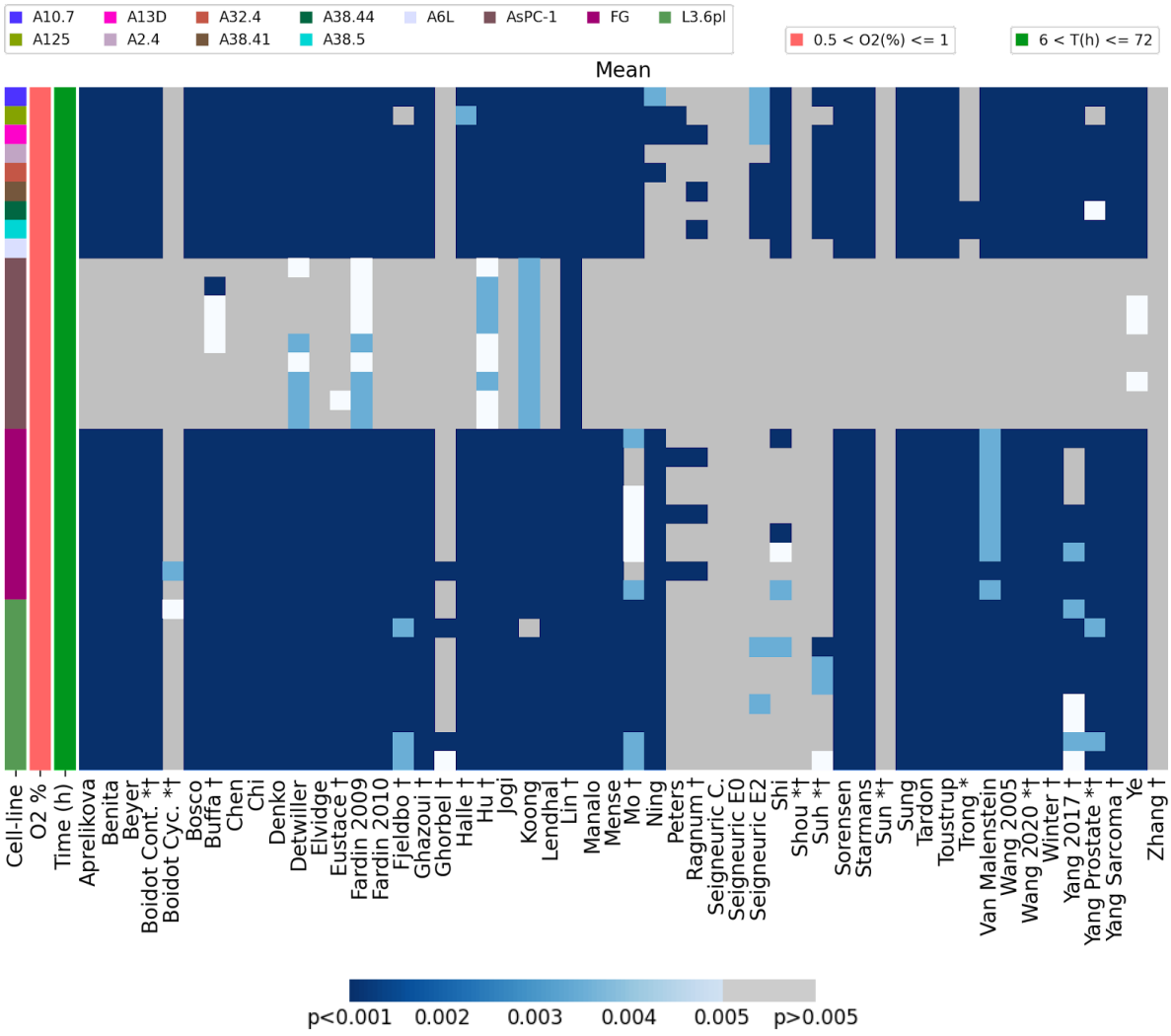
Three scores (median, mean and GSVA) performed well overall (*Fig. 4.10* and *Table 4.20*) and a number of signatures had 100% hypoxia identification accuracy using these scores: median - 13 signatures, mean - 3 signatures and GSVA - 10 signatures (*Table 4.20*). The worst performing signatures were Seigneuric common, Shou, Sun and Zhang with 0% accuracy on all scores. As commonly observed above, the ssGSEA score was the worst performing score.

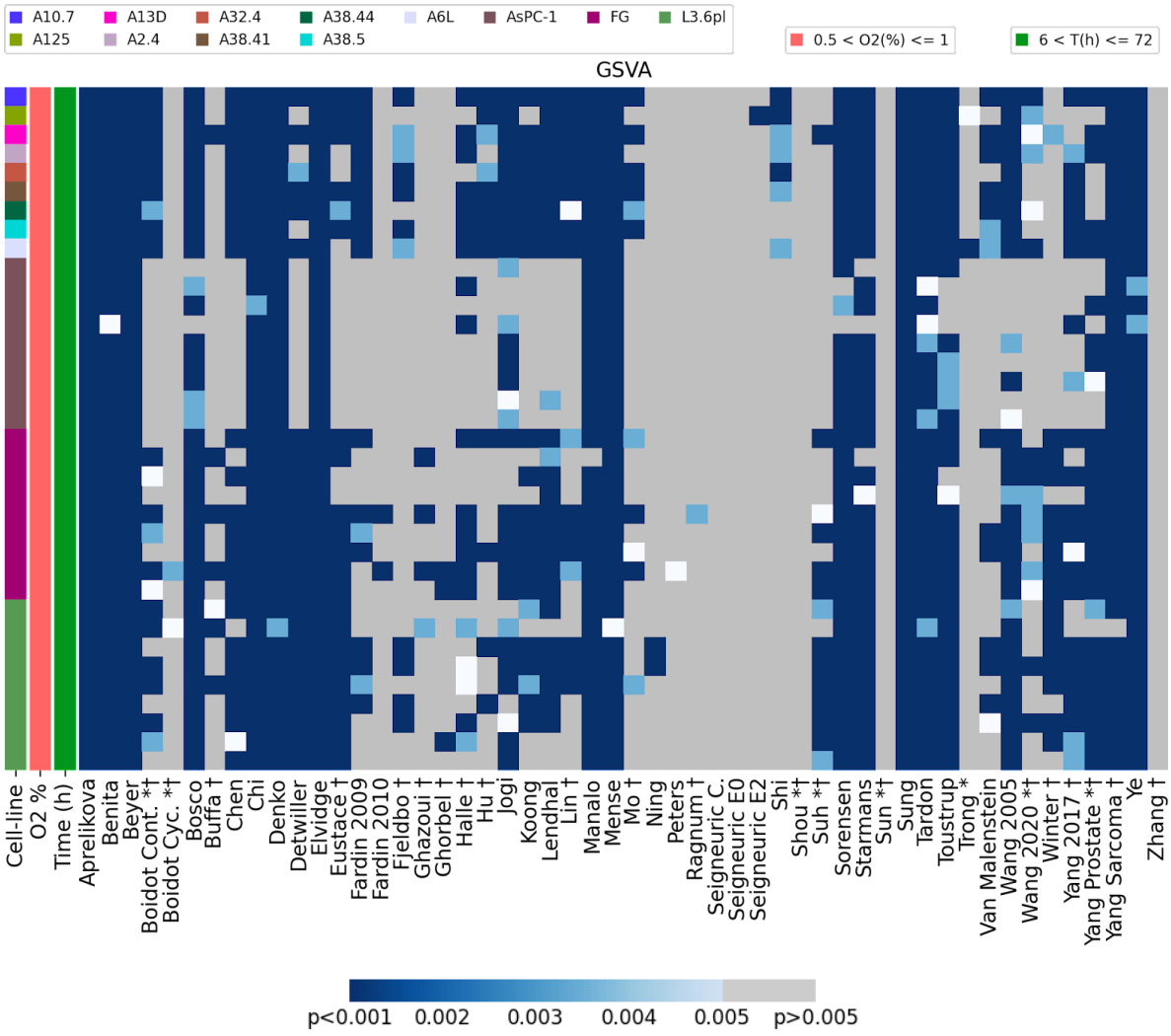
Cell line/samples	Number of Hypoxic samples	Number of Normoxic samples	Pairs
AsPC-1	3	3	9
FG	3	3	9
L3.6pl	3	3	9
A10.7	1	1	1
A125	1	1	1
A13D	1	1	1
A2.4	1	1	1
A32.4	1	1	1
A38.41	1	1	1
A38.44	1	1	1
A38.5	1	1	1
A6L	1	1	1
Total	**18**	**18**	**36**

Table 4.19: *Cancer cell lines originating from the central nervous system used in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using cancer cell lines originating from the central nervous system.


A10.7
A125
A13D
A2.4
A32.4
A38.41
A38.44
A38.5
A6L
AsPC-1
FG
L3.6pl
0.5 < O2(%) <= 1
6 < T(h) <= 72
Median
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


A10.7
A125
A13D
A2.4
A32.4
A38.41
A38.44
A38.5
A6L
AsPC-1
FG
L3.6pl
0.5 < O2(%) <= 1
6 < T(h) <= 72
Mean
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


A10.7
A125
A13D
A2.4
A32.4
A38.41
A38.44
A38.5
A6L
AsPC-1
FG
L3.6pl
0.5 < O2(%) <= 1
6 < T(h) <= 72
GSVA
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

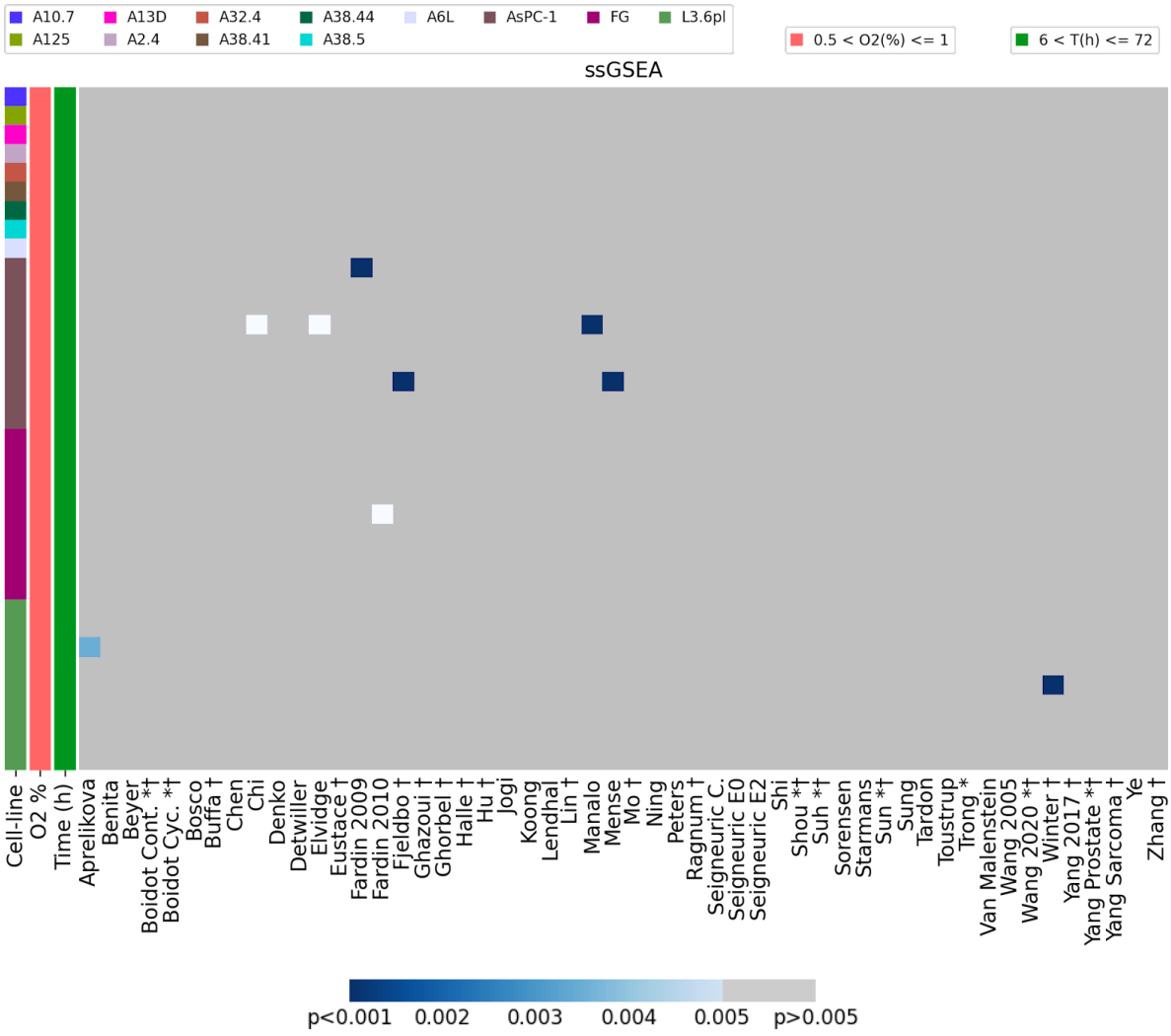


Figure 4.10: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in pancreatic cancer*

Comparison of the scoring methods of the 53 hypoxia signatures across all pancreatic cancer hypoxia experiments identified. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	100.00	75.00	100.00	2.78
Benita	100.00	75.00	100.00	0.00
Beyer	100.00	75.00	100.00	0.00
Boidot Cont. *†	97.22	75.00	58.33	0.00
Boidot Cyc. *†	22.22	5.56	5.56	0.00
Bosco	86.11	75.00	94.44	0.00
Buffa †	80.56	86.11	19.44	0.00
Chen	100.00	75.00	66.67	0.00
Chi	97.22	75.00	100.00	2.78
Denko	100.00	75.00	100.00	0.00
Detwiller	63.89	91.67	66.67	0.00
Elvidge	100.00	75.00	100.00	2.78
Eustace †	83.33	77.78	61.11	0.00
Fardin 2009	100.00	100.00	47.22	2.78
Fardin 2010	61.11	75.00	5.56	2.78
Fjeldbo †	91.67	72.22	30.56	2.78
Ghazoui †	52.78	75.00	11.11	0.00
Ghorbel †	22.22	11.11	8.33	0.00
Halle †	77.78	75.00	55.56	0.00
Hu †	83.33	100.00	33.33	0.00
Jogi	83.33	75.00	80.56	0.00
Koong	58.33	97.22	58.33	0.00
Lendhal	100.00	75.00	69.44	0.00
Lin †	83.33	100.00	47.22	0.00
Manalo	97.22	75.00	97.22	2.78
Mense	97.22	75.00	100.00	2.78
Mo †	44.44	66.67	30.56	0.00
Ning	63.89	61.11	5.56	0.00
Peters	44.44	13.89	2.78	0.00
Ragnum †	30.56	16.67	2.78	0.00
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	27.78	0.00	0.00	0.00
Seigneuric E2	22.22	25.00	2.78	0.00
Shi	2.78	38.89	19.44	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	50.00	30.56	38.89	0.00
Sorensen	100.00	75.00	94.44	0.00
Starmans	97.22	75.00	91.67	0.00
Sun *†	0.00	0.00	0.00	0.00
Sung	100.00	75.00	100.00	0.00
Tardon	100.00	75.00	100.00	0.00
Toustrup	97.22	75.00	88.89	0.00
Trong *	0.00	55.56	5.56	0.00
Van Malenstein	72.22	75.00	38.89	0.00
Wang 2005	97.22	75.00	83.33	0.00
Wang 2020 *†	58.33	75.00	41.67	0.00
Winter †	75.00	75.00	44.44	2.78
Yang 2017 †	77.78	66.67	63.89	0.00
Yang Prostate *†	77.78	72.22	72.22	0.00
Yang Sarcoma †	100.00	75.00	97.22	0.00
Ye	100.00	83.33	100.00	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	41.67	5.56	0.00

Table 4.20: *Percentage accuracy of determining hypoxic samples from normoxic samples in pancreatic cancer for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

4.11 Performance of hypoxia signatures in immortalised non-cancer cell lines

Immortalised non-cancer derived cell lines form a key part of biological research in areas including hypoxia. Although not cancer per se, these cells should not be viewed as “normal,” as they have been modified to divide indefinitely and can express unique gene expression patterns not found in any cell type *in vivo*. Also, after periods of continuous growth, cell characteristics can change and care must be taken to periodically evaluate their characteristics to ensure they still are representative of the original parent cell. The aim of this analysis is to assess the performance of hypoxia signatures within these cells, particularly as clinical biopsies which are often used to assess hypoxic status using gene expression signatures can also contain non-cancer tissue.

Non-cancer cell lines with gene expression data in hypoxia and normoxia were identified from the Gene Expression Omnibus (as in Methods). This yielded 10 cell lines originating from mammary epithelial cells, oesophageal epithelial cells, gastric myofibroblasts, dendritic cells and astrocytes (*Table 4.21*). Samples originating from the mammary epithelial cells made up 25 of the 77 pairwise combinations, there were 16 immortalised esophageal epithelial cell lines and the rest of the samples were evenly split between the other cell types.

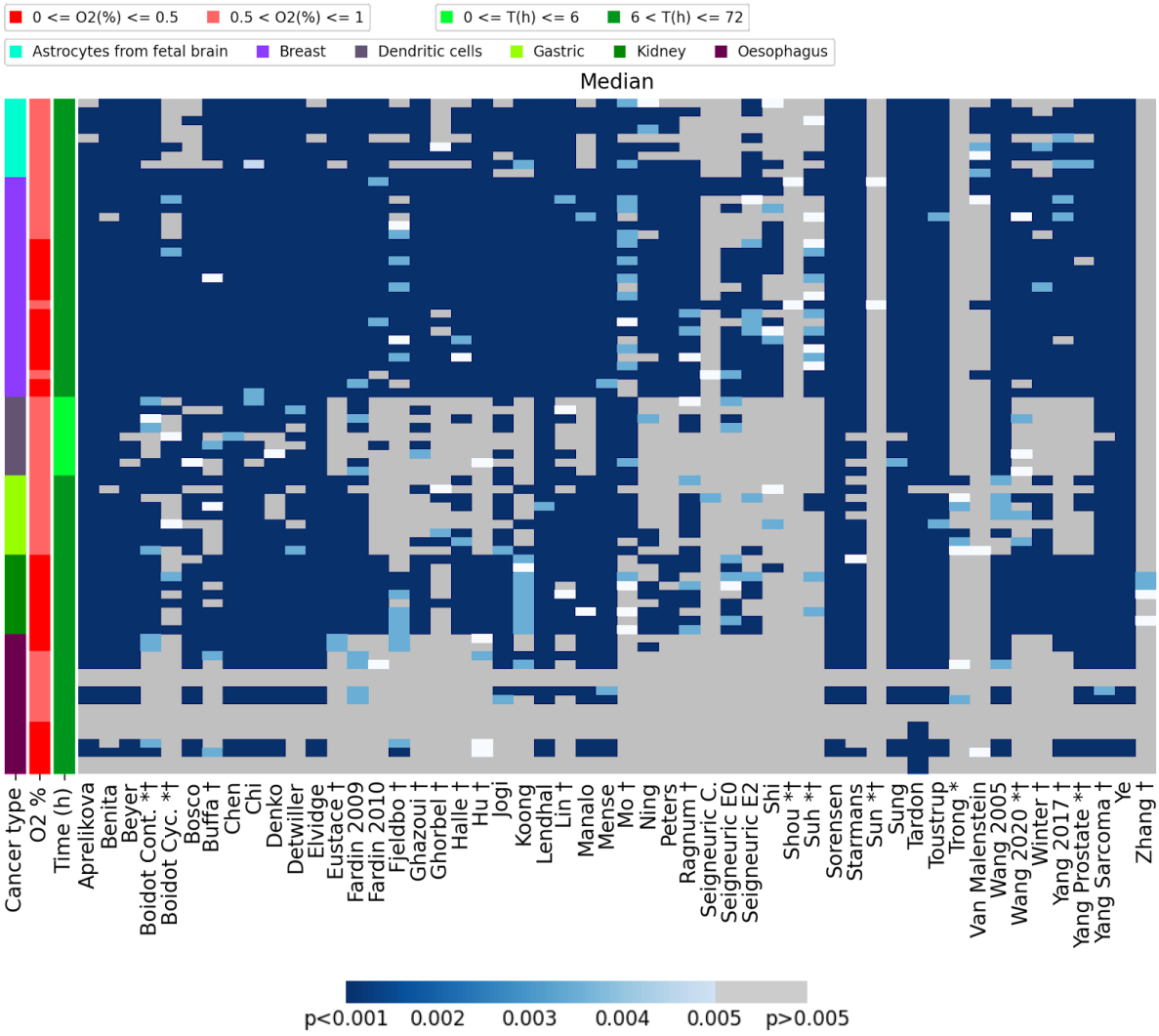
I then assessed the performance of the 53 signatures at defining hypoxia in these 10 cell lines. Overall, GVSA and median both perform well (*Fig. 4.11*). The best performing signatures in the immortalised non-cancer cells were the MCF-7 derived Aprekoliva and multi-cell line derived Beyer signatures using GVSA (98.70% accuracy, *Table 4.22*). Also, Elvidge (MCF-7 derived) and Mense (derived from peripheral blood monocytes from normal individuals) using GSVA achieved 97.40% accuracy at identifying hypoxic samples. Again, the worst performing signature across all scores was Shou, achieving a maximum of only 2.94% accuracy. As with the other analyses, ssGSEA was the worst performing score (*Fig. 4.11*).

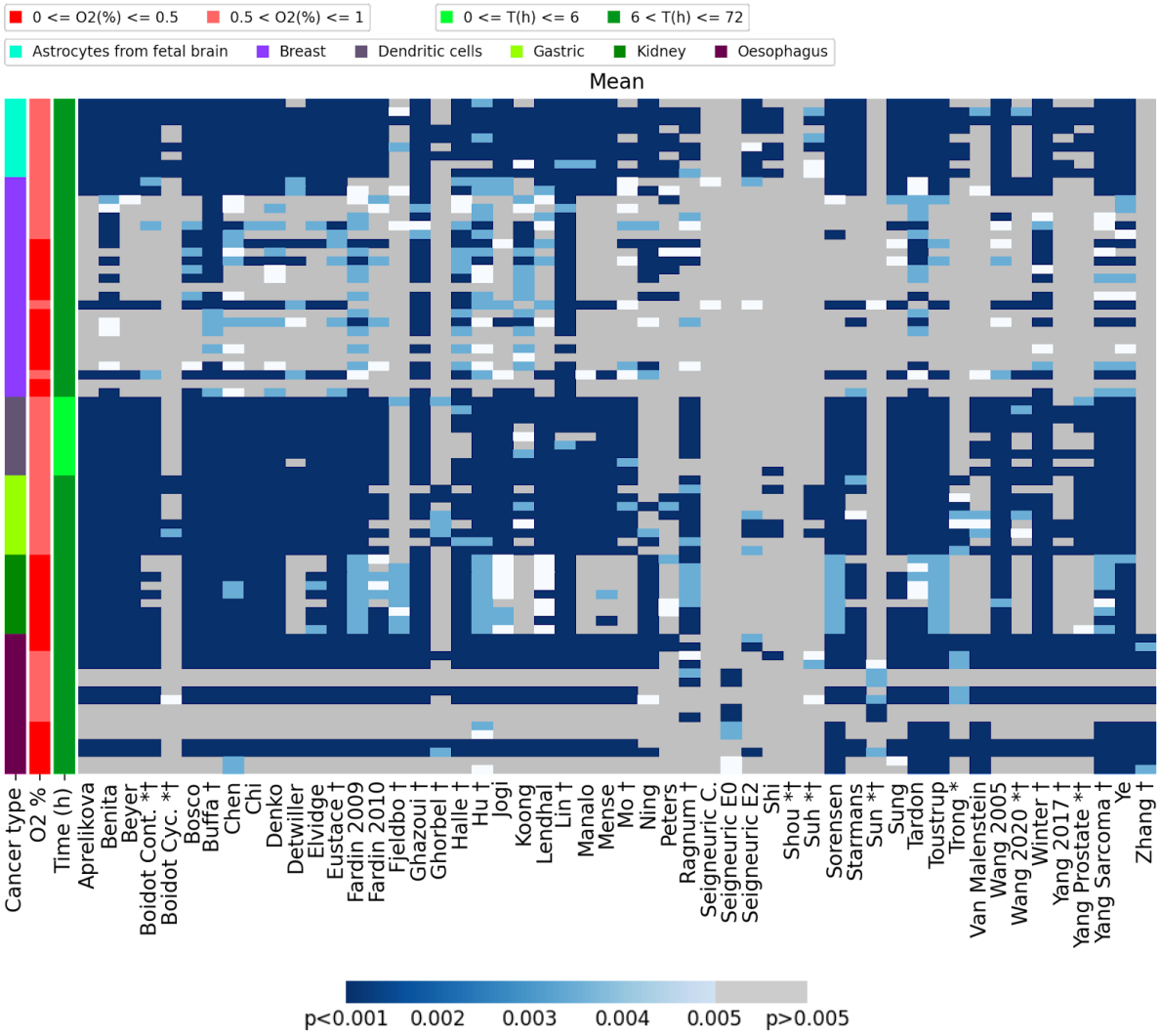
It is desirable that hypoxia signatures that perform well in the cancer samples of one tissue type also perform well on the non-cancer lines of the sample tissue/organ type. Reassuringly, this does seem to be the case and lays a solid foundation for exploring normal and cancer biopsies from the TCGA (Chapter 6). The most direct comparison here compares breast cancer cell lines with immortalised mammary cells where the median score for Sorensen (the best signatures in breast cancer lines) performed close to 100% accuracy in defining hypoxia in the non-cancer lines (*Appendix 2, Supplementary Table S2*).

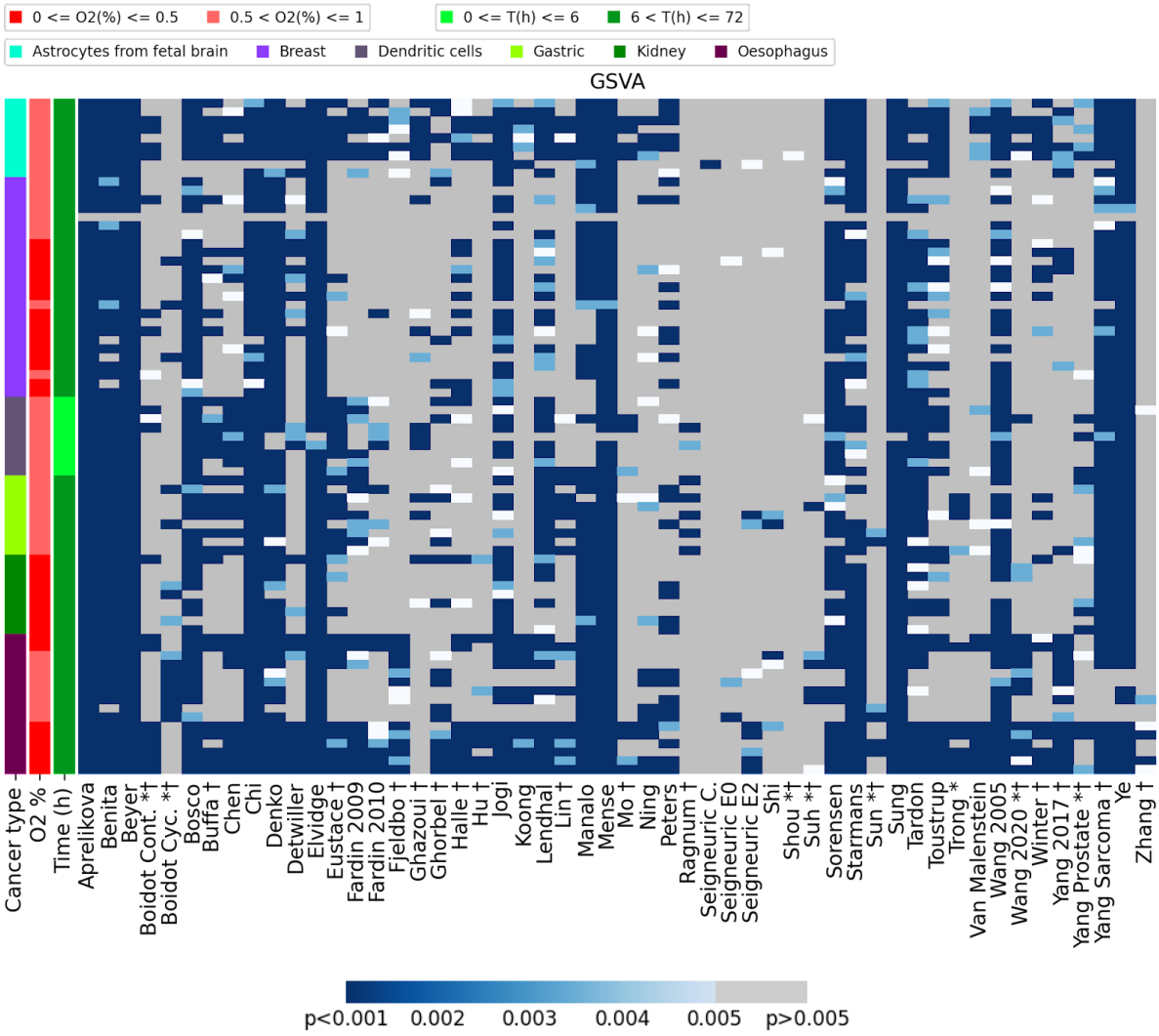
Cell line/samples	Number of Hypoxic samples	Number of Normoxic samples	Pairs
MCF10A (immortalised mammary epithelial cells)	7	7	21
EPC2 (immortalised oesophageal epithelial cells)	4	4	16
Immortalised gastric myofibroblasts	3	3	9
Immortalised astrocytes (fetal brain)	3	3	9
Immortalised dendritic cells	3	3	9
HKC8 (immortalised human renal proximal cell line)	3	3	9
HBL100 (immortalised mammary epithelial cells)	1	1	1
HME2 (immortalised mammary epithelial cells)	1	1	1
MCF12A (immortalised mammary epithelial cells)	1	1	1
hTERT-HME (immortalised mammary epithelial cells)	1	1	1
Total	**27**	**27**	**77**

Table 4.21: *Immortalised non-cancer cell lines used in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using immortalised non-cancer cell lines.


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
0 <= T(h) <= 6
6 < T(h) <= 72
Astrocytes from fetal brain
Breast
Dendritic cells
Gastric
Kidney
Oesophagus
Median
Cancer type
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
0 <= T(h) <= 6
6 < T(h) <= 72
Astrocytes from fetal brain
Breast
Dendritic cells
Gastric
Kidney
Oesophagus
Mean
Cancer type
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
0 <= T(h) <= 6
6 < T(h) <= 72
Astrocytes from fetal brain
Breast
Dendritic cells
Gastric
Kidney
Oesophagus
GSVA
Cancer type
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

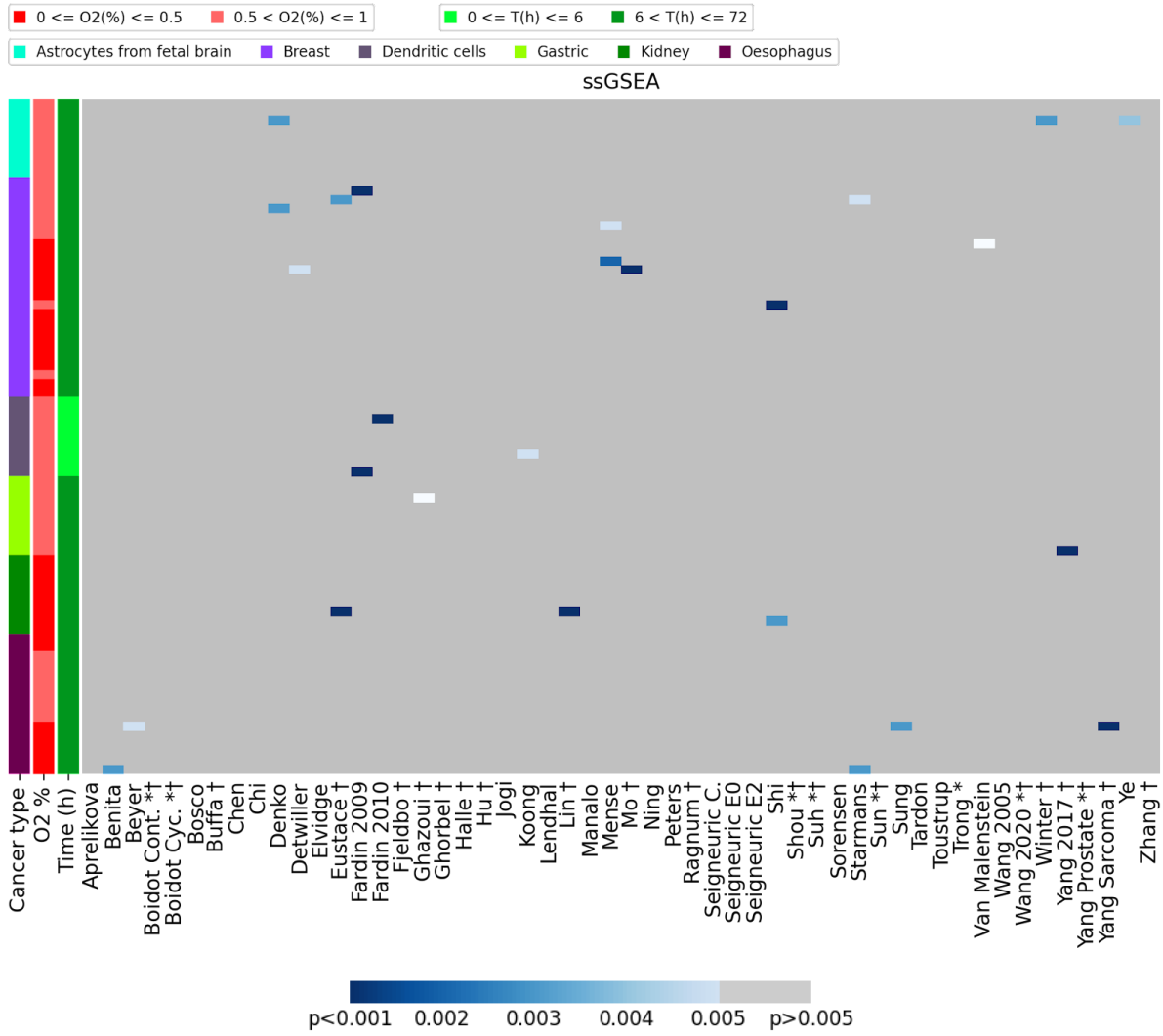


Figure 4.11: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in immortalised non-cancer cell lines*

Comparison of the scoring methods of the 53 hypoxia signatures across immortalised non-cancer cell lines. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	87.01	62.34	98.70	0.00
Benita	85.71	79.22	93.51	1.30
Beyer	87.01	63.64	98.70	1.30
Boidot Cont. *†	77.92	58.44	25.97	0.00
Boidot Cyc. *†	50.65	16.88	20.78	0.00
Bosco	75.32	75.32	89.61	0.00
Buffa †	68.83	84.42	41.56	0.00
Chen	89.61	80.52	45.45	0.00
Chi	87.01	67.53	96.10	0.00
Denko	83.12	75.32	92.21	2.60
Detwiller	81.82	51.95	38.96	1.30
Elvidge	87.01	61.04	97.40	0.00
Eustace †	68.83	70.13	48.05	2.60
Fardin 2009	79.22	79.22	38.96	2.60
Fardin 2010	55.84	58.44	28.57	1.30
Fjeldbo †	49.35	28.57	25.97	0.00
Ghazoui †	62.34	80.52	22.08	1.30
Ghorbel †	41.56	18.18	32.47	0.00
Halle †	61.04	75.32	38.96	0.00
Hu †	62.34	81.82	15.58	0.00
Jogi	75.32	64.94	81.82	0.00
Koong	62.34	62.34	23.38	1.30
Lendhal	89.61	70.13	72.73	0.00
Lin †	72.73	85.71	28.57	1.30
Manalo	71.43	49.35	80.52	0.00
Mense	89.61	53.25	97.40	2.60
Mo †	61.04	50.65	15.58	1.30
Ning	51.95	49.35	25.97	0.00
Peters	50.65	27.27	41.56	0.00
Ragnum †	58.44	57.14	9.09	0.00
Seigneuric C.	6.49	2.60	1.30	0.00
Seigneuric E0	37.66	10.39	3.90	0.00
Seigneuric E2	44.16	22.08	11.69	0.00
Shi	31.17	14.29	7.79	2.60
Shou *†	2.94	0.00	1.47	0.00
Suh *†	33.77	18.18	9.09	0.00
Sorensen	89.61	70.13	84.42	0.00
Starmans	81.82	59.74	84.42	2.60
Sun *†	2.60	10.39	7.79	0.00
Sung	88.31	64.94	96.10	1.30
Tardon	93.51	85.71	79.22	0.00
Toustrup	88.31	71.43	54.55	0.00
Trong *	10.39	28.57	12.99	0.00
Van Malenstein	18.18	35.06	22.08	1.30
Wang 2005	84.42	68.83	85.71	0.00
Wang 2020 *†	59.74	33.77	25.97	0.00
Winter †	57.14	77.92	29.87	1.30
Yang 2017 †	58.44	27.27	33.77	1.30
Yang Prostate *†	67.53	37.66	25.97	0.00
Yang Sarcoma †	88.31	80.52	88.31	1.30
Ye	89.61	83.12	93.51	1.30
Zhang †	5.19	15.58	7.79	0.00
Zou *†	38.96	27.27	7.79	0.00

Table 4.22: *Percentage accuracy of determining hypoxic samples from normoxic samples immortalised non-cancer cell lines for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

4.12 Overall performance of hypoxia signatures

I then investigated the performance of all four summary scores and 53 published signatures across the 90 cancer cell line/tissue experiments identified from the Gene Expression Omnibus. This pan-cancer analysis includes nine cancer types and ~1000 pairwise comparisons of hypoxic and normoxic samples. Three extra tumour types (Burkitt's lymphoma, Ewing's sarcoma and fibrosarcoma) and a cancer-associated gastric myofibroblast experiment were also included in this global approach (which each had a low number of total samples, so were not discussed in individual sections). *Fig. 4.12* shows a summary of this comprehensive analysis.

However, it is noteworthy that VHL loss-of-function mutations are known to activate HIF-1a in the absence of hypoxia, thus making non-hypoxic cells take an HIF-1a activated "hypoxic phenotype". Such mutations are common in renal cancer. Therefore, for the sake of identifying hypoxia itself rather than VHL mutations, all cell lines/tissues without VHL mutations in the analysis presented here were analysed. The performance of signatures within VHL mutated cell lines are discussed in detail in Chapter 5.3. Finally, the pan-cancer approach was revisited with different considerations in an even more comprehensive analysis in Chapter 5.8.

The best performing signature/score combination was the Aprelikova signature using GVSA score. This signature achieved an accuracy across non-VHL mutated cancer types of 88.01% (highlighted red in *Table 4.24*). When adding in immortalised non-cancer cell lines and a cancer-associated cell experiment (an extra 74 paired samples), Aprelikova using the GVSA score remained the best performing with an accuracy of 88.76% (*Appendix 2, Supplementary Table S2*). Impressively, the Aprelikova signature using the GVSA score achieved >80% accuracy on 11 of the 14 cancer classes tested and of these 9 reached >90% accuracy (breast, prostate, cervical, ovarian, pancreas, Burkitt Lymphoma, Ewing's sarcoma, CNS and bone, *Table 4.25*).

Generally, poor performance in lung cancer was a common theme across the signatures and scores (*Table 4.25* and *Fig. 4.12*). The Aprelikova signature using the GVSA score does relatively well compared to the rest of the signatures, with 68.10% accuracy. However, improved accuracy can be achieved in identifying hypoxia in lung cell lines/tissue using Starmans or Mense and the mean score (86.21%, see Chapter 4.2). The 16-gene Sun signature and 7-gene Shou signature were generally ineffective at identifying hypoxia, demonstrating the importance of the choice of hypoxia gene expression signature used in the laboratory (or in the clinic) in identifying the state of hypoxia.

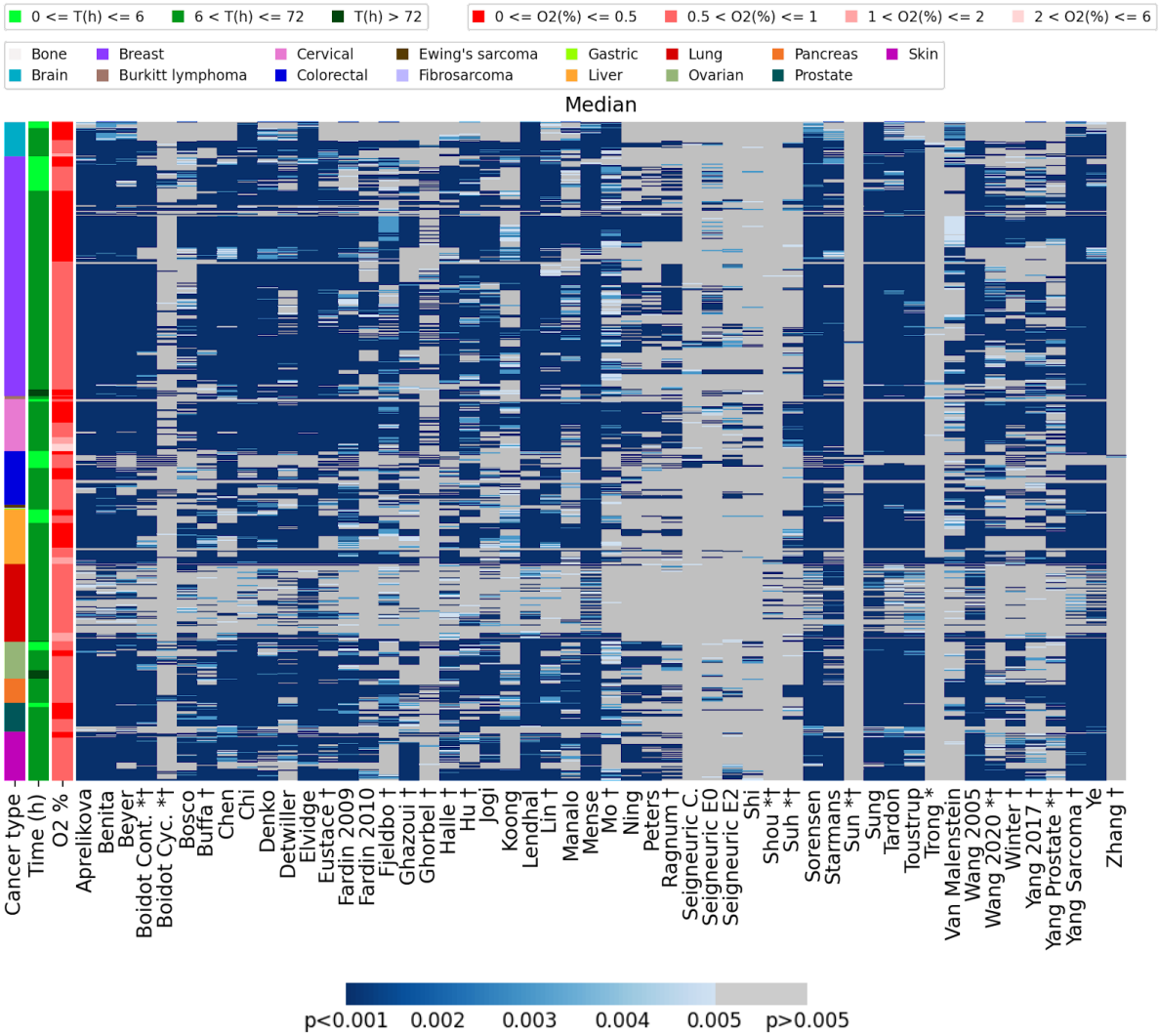
To summarise, this work highlights that Aprelikova signature using the GSVA score gives the best indication of hypoxia amongst a variety of cell types. However, in lung cancer particularly, specific signatures might be preferable (e.g. Starmans using the mean score). Or perhaps the lung cancer results from mainly *ex vivo* samples suggest one might see different behaviour in human tumours compared to cell lines. It is worth noting, all tissue specific analyses yielded higher accuracy than the global approach (e.g. the Lendhal signature using the median score yielded 100% accuracy across 54 paired samples in ovarian lines whereas Aprelikova using GSVA reached 94.44%).

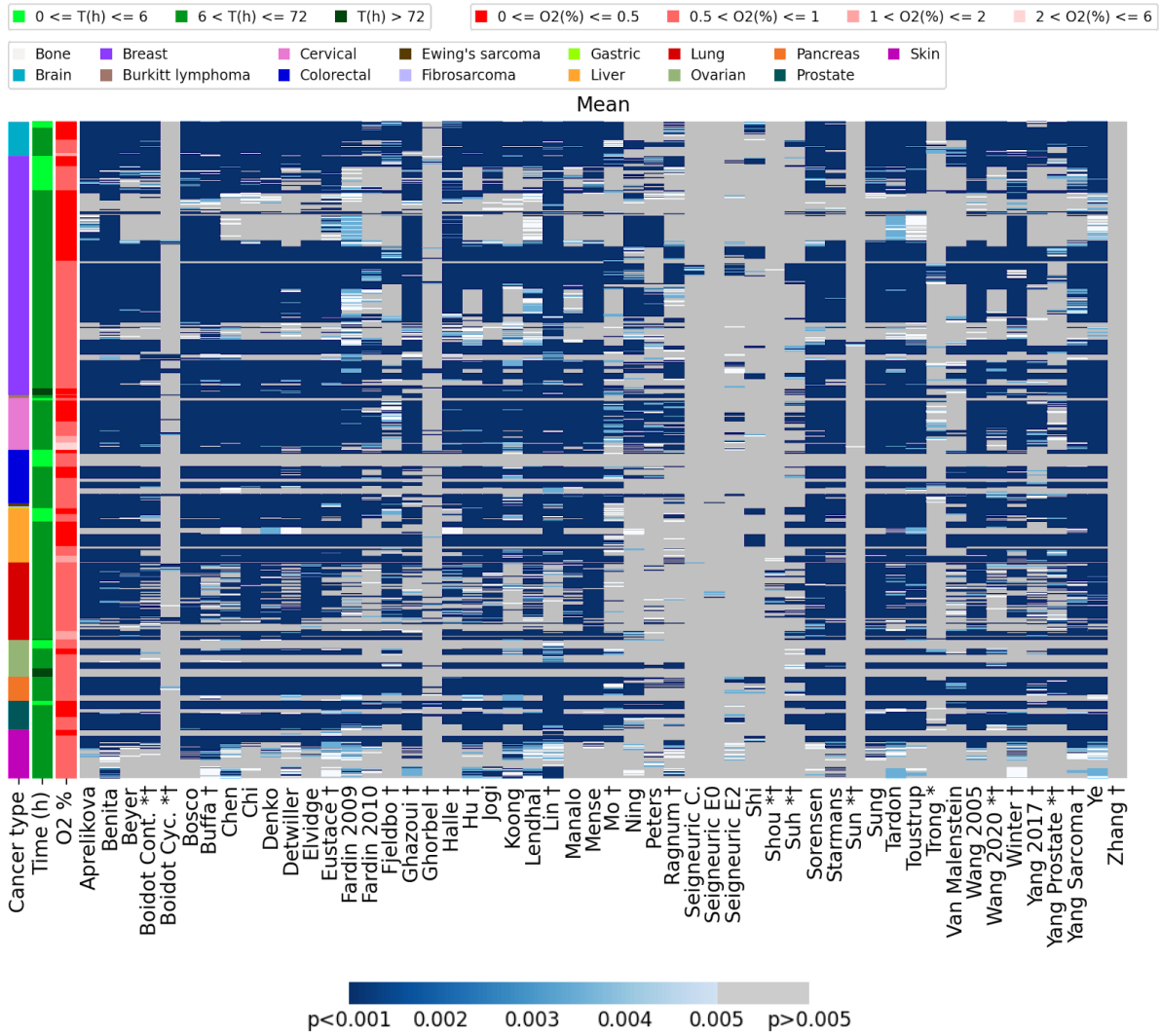
However, it should be noted that the majority of this work comes from cancer cell lines, not clinical samples, so how transferable these results are to the clinic remains to be elucidated. For this reason, the study will move on to assess the performance of hypoxia signatures looking at matched normal-cancer pair samples in clinical datasets using the TCGA (Chapter 6). A similar methodology to that exploited in the evaluation of the hypoxic state of cancer cell lines will be applied. However, the next section investigates how hypoxia signatures perform when genetic perturbations are introduced. It will also explore the accuracy of hypoxia gene expression signatures under various experimental constraints that aim to mimic other tumour microenvironmental conditions.

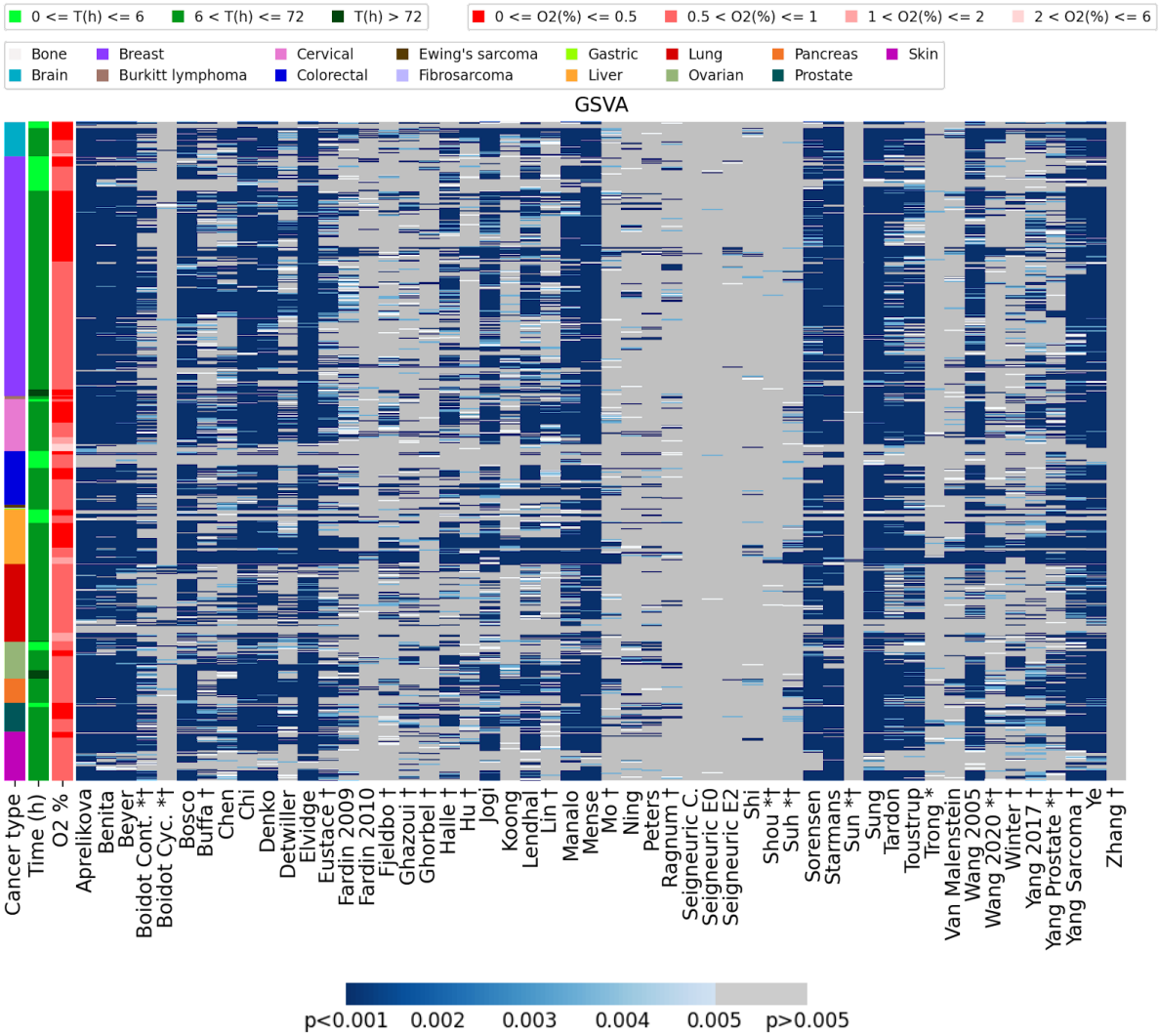
Cell line/samples	Number of Hypoxic samples	Number of Normoxic samples	Pairs
MCF-7	79	48	243
NSCLC	10	10	100
HCT116	19	16	57
MDA-MB-231	20	14	56
HepG2	16	10	46
HeLa	13	14	42
A2780	6	6	36
U87	12	6	36
501mel	9	3	27
IGR39	9	3	27
SiHa	8	8	28
Huh-7	12	4	21
HT29	15	3	15
A549	6	6	14
DU145	14	2	14
LNCaP	6	6	14
PC-3	6	6	14
T-47D	5	5	11
Hep3B	4	4	10
Others	106	97	165
Total	**370**	**271**	**976**

Table 4.23: *Cell lines used in hypoxia experiments in GEO and included in the pan-cancer hypoxia signature analysis*

Details of the hypoxia versus normoxia experiments in GEO using all cancer and cancer-associated cell lines.


0 <= T(h) <= 6
6 < T(h) <= 72
T(h) > 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
2 < O2(%) <= 6
Bone
Brain
Breast
Burkitt lymphoma
Cervical
Colorectal
Ewing's sarcoma
Fibrosarcoma
Gastric
Liver
Lung
Ovarian
Pancreas
Prostate
Skin
Median
Cancer type
Time (h)
O2 %
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fieldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= T(h) <= 6
6 < T(h) <= 72
T(h) > 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
2 < O2(%) <= 6
Bone
Brain
Breast
Burkitt lymphoma
Cervical
Colorectal
Ewing's sarcoma
Fibrosarcoma
Gastric
Liver
Lung
Ovarian
Pancreas
Prostate
Skin
Mean
Cancer type
Time (h)
O2 %
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fieldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


0 <= T(h) <= 6
6 < T(h) <= 72
T(h) > 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
1 < O2(%) <= 2
2 < O2(%) <= 6
Bone
Brain
Breast
Burkitt lymphoma
Cervical
Colorectal
Ewing's sarcoma
Fibrosarcoma
Gastric
Liver
Lung
Ovarian
Pancreas
Prostate
Skin
GSVA
Cancer type
Time (h)
O2 %
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fieldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

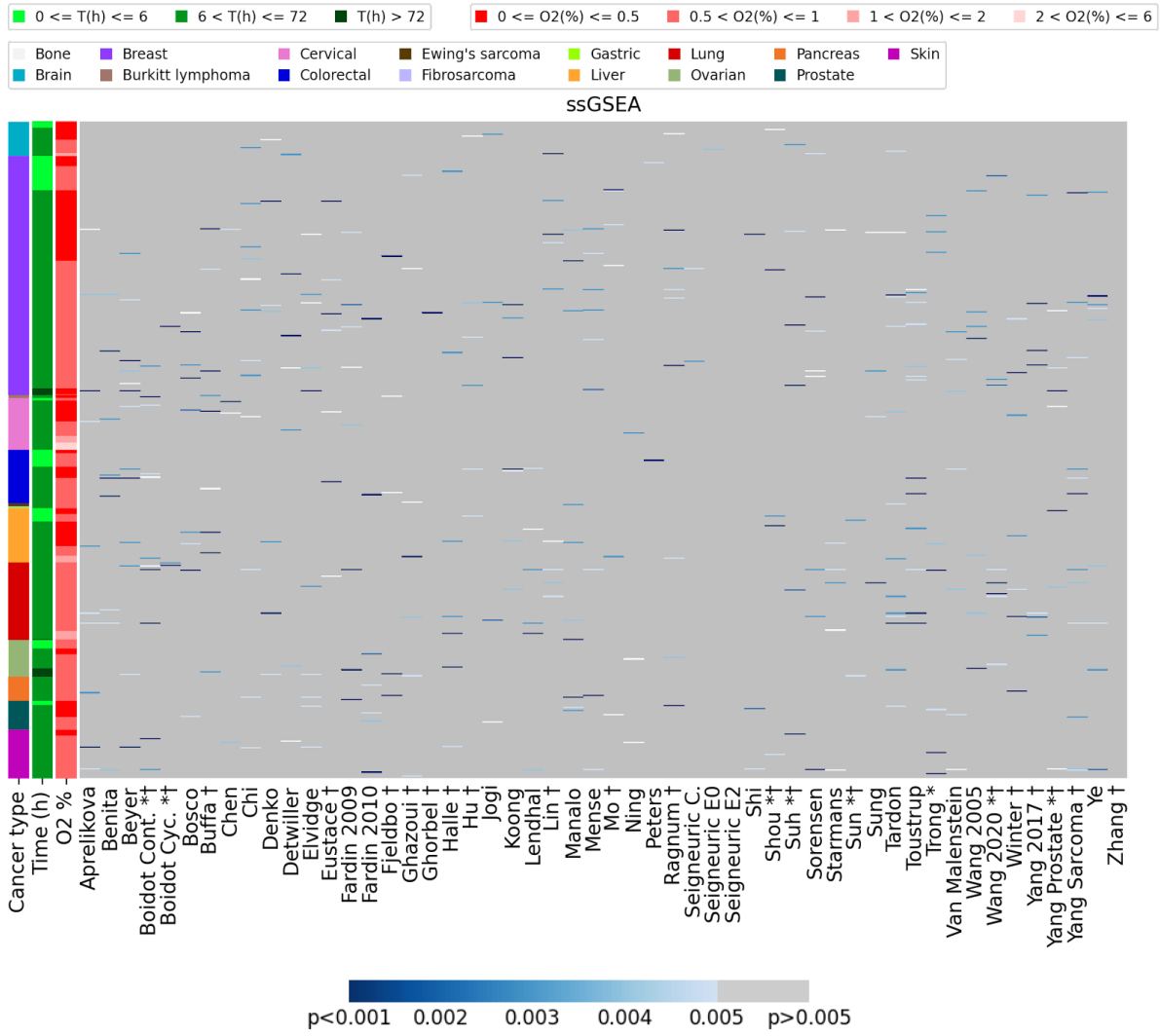


Figure 4.12: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in hypoxia experiments on cancer cell lines/tissue identified in the Gene Expression Omnibus (GEO)*

Comparison of the scoring methods of the 53 hypoxia signatures in hypoxia experiments on cancer cell lines/tissue. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	82.99	72.85	88.01	1.02
Benita	85.45	80.02	80.53	0.92
Beyer	80.94	67.32	82.17	1.33
Boidot Cont. *†	69.67	62.60	46.93	1.13
Boidot Cyc. *†	9.94	4.10	2.87	0.31
Bosco	51.84	74.59	73.57	0.92
Buffa †	66.09	75.61	39.65	0.92
Chen	73.57	71.31	44.77	0.41
Chi	79.20	70.08	79.92	1.23
Denko	74.49	68.44	74.08	0.72
Detwiller	63.11	62.81	37.81	0.72
Elvidge	85.14	68.55	84.43	1.02
Eustace †	70.49	72.34	50.00	0.72
Fardin 2009	69.06	69.67	32.07	0.82
Fardin 2010	46.72	44.98	6.45	0.72
Fjeldbo †	53.79	45.39	20.49	0.51
Ghazoui †	48.16	64.86	20.80	0.72
Ghorbel †	13.63	5.74	9.43	0.10
Halle †	65.06	66.50	45.39	0.72
Hu †	64.55	60.86	20.80	0.61
Jogi	61.78	66.91	62.60	0.41
Koong	39.14	52.66	22.85	0.72
Lendhal	83.91	70.59	59.94	0.61
Lin †	73.16	78.48	28.48	0.92
Manalo	48.05	57.58	70.18	1.02
Mense	86.78	70.49	83.81	0.92
Mo †	51.95	40.16	13.42	0.61
Ning	36.78	37.30	9.63	0.31
Peters	26.84	26.02	8.30	0.20
Ragnum †	33.81	33.81	7.48	0.92
Seigneuric C.	4.82	1.54	0.10	0.20
Seigneuric E0	11.68	0.51	0.61	0.10
Seigneuric E2	11.07	10.76	1.54	0.00
Shi	6.35	10.66	3.79	0.31
Shou *†	1.95	2.36	1.13	0.51
Suh *†	18.34	21.82	7.58	0.61
Sorensen	87.40	69.98	77.25	0.82
Starmans	76.84	64.55	81.76	0.61
Sun *†	1.02	0.61	0.31	0.20
Sung	86.89	71.62	85.35	0.41
Tardon	85.55	76.43	69.67	1.43
Toustrup	83.40	73.87	65.68	1.33
Trong *	2.56	15.57	3.89	0.82
Van Malenstein	41.39	36.27	13.22	0.72
Wang 2005	78.79	69.57	68.24	0.72
Wang 2020 *†	46.82	48.98	18.85	0.92
Winter †	40.78	65.37	22.64	0.72
Yang 2017 †	44.16	40.88	38.32	0.72
Yang Prostate *†	35.96	40.27	34.63	0.72
Yang Sarcoma †	85.04	71.93	76.02	1.23
Ye	85.14	78.59	82.07	0.92
Zhang †	0.20	0.00	0.00	0.00
Zou *†	11.68	16.39	4.71	0.82

Table 4.24: *Percentage accuracy of determining hypoxic samples from normoxic samples in experiments involving cancer cell lines/samples for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

a) Median

	Breast	Prostate	Colorectal	Lung	Cervical	Ovarian	Liver	Brain	Skin	Pancreas	Burkitt lymphoma	Ewing's sarcoma	Bone	Fibrosarcoma
Aprelikova	95.21	79.07	63.75	30.17	89.61	96.30	**95.00**	80.39	**95.83**	**100.00**	**100.00**	75.00	0.00	**0.00**
Benita	96.90	76.74	75.00	37.93	**93.51**	87.04	92.50	92.16	91.67	**100.00**	**100.00**	**100.00**	**100.00**	**0.00**
Beyer	92.96	86.05	47.50	37.93	90.91	94.44	**95.00**	82.35	77.78	**100.00**	**100.00**	**100.00**	**100.00**	**0.00**
Boidot Cont. *†	83.38	74.42	48.75	18.10	**93.51**	75.93	68.75	35.29	91.67	97.22	**100.00**	0.00	0.00	**0.00**
Boidot Cyc. *†	13.80	0.00	13.75	0.86	10.39	20.37	2.50	3.92	6.94	22.22	0.00	0.00	0.00	**0.00**
Bosco	54.93	60.47	18.75	31.03	62.34	61.11	70.00	31.37	66.67	86.11	0.00	0.00	**100.00**	**0.00**
Buffa †	83.94	79.07	31.25	11.21	89.61	61.11	81.25	19.61	90.28	80.56	**100.00**	0.00	0.00	**0.00**
Chen	91.83	**97.67**	**82.50**	12.07	89.61	**100.00**	41.25	39.22	72.22	**100.00**	50.00	**100.00**	0.00	**0.00**
Chi	93.80	90.70	36.25	25.86	90.91	**100.00**	92.50	84.31	77.78	97.22	**100.00**	**100.00**	0.00	**0.00**
Denko	88.45	86.05	50.00	30.17	92.21	72.22	83.75	60.78	65.28	**100.00**	75.00	**100.00**	**100.00**	**0.00**
Detwiller	71.27	55.81	38.75	23.28	90.91	81.48	88.75	62.75	47.22	63.89	75.00	25.00	**100.00**	**0.00**
Elvidge	96.34	93.02	55.00	46.55	89.61	92.59	92.50	92.16	88.89	**100.00**	**100.00**	**100.00**	**100.00**	**0.00**
Eustace †	87.04	76.74	52.50	28.45	92.21	83.33	61.25	41.18	69.44	83.33	**100.00**	0.00	0.00	**0.00**
Fardin 2009	75.21	69.77	61.25	12.93	92.21	85.19	90.00	39.22	79.17	**100.00**	**100.00**	**100.00**	**100.00**	**0.00**
Fardin 2010	57.18	32.56	12.50	3.45	92.21	59.26	50.00	23.53	61.11	61.11	**100.00**	0.00	0.00	**0.00**
Fjeldbo †	68.45	53.49	55.00	22.41	83.12	33.33	53.75	19.61	20.83	91.67	**100.00**	25.00	**100.00**	**0.00**
Ghazoui †	56.06	41.86	36.25	5.17	90.91	40.74	38.75	21.57	84.72	52.78	**100.00**	0.00	0.00	**0.00**
Ghorbel †	23.38	2.33	17.50	0.86	15.58	7.41	1.25	5.88	8.33	22.22	0.00	0.00	0.00	**0.00**
Halle †	79.72	65.12	30.00	25.00	92.21	87.04	56.25	35.29	77.78	77.78	**100.00**	0.00	0.00	**0.00**
Hu †	79.72	55.81	68.75	14.66	76.62	75.93	70.00	50.98	45.83	83.33	**100.00**	25.00	**100.00**	**0.00**
Jogi	78.87	62.79	30.00	27.59	63.64	85.19	41.25	39.22	77.78	83.33	**100.00**	50.00	0.00	**0.00**
Koong	45.92	48.84	17.50	11.21	54.55	31.48	68.75	33.33	25.00	58.33	0.00	25.00	0.00	**0.00**
Lendhal	96.90	79.07	55.00	32.76	**93.51**	**100.00**	**95.00**	92.16	93.06	**100.00**	**100.00**	0.00	**100.00**	**0.00**
Lin †	88.45	76.74	66.25	15.52	**93.51**	70.37	93.75	45.10	75.00	83.33	**100.00**	0.00	0.00	**0.00**
Manalo	53.52	60.47	16.25	16.38	62.34	62.96	51.25	50.98	43.06	97.22	50.00	75.00	**100.00**	**0.00**
Mense	92.11	76.74	80.00	56.90	**93.51**	92.59	**95.00**	96.08	90.28	97.22	**100.00**	**100.00**	0.00	**0.00**
Mo †	65.35	65.12	33.75	11.21	77.92	50.00	37.50	86.27	41.67	44.44	0.00	0.00	0.00	**0.00**
Ning	40.85	27.91	23.75	2.59	83.12	57.41	15.00	7.84	63.89	63.89	0.00	0.00	0.00	**0.00**
Peters	29.58	13.95	18.75	1.72	64.94	33.33	8.75	11.76	51.39	44.44	0.00	0.00	0.00	**0.00**
Ragnum †	50.14	11.63	12.50	4.31	67.53	35.19	22.50	13.73	33.33	30.56	25.00	0.00	0.00	**0.00**
Seigneuric C.	5.92	0.00	21.25	0.86	7.79	0.00	0.00	3.92	0.00	0.00	0.00	0.00	0.00	**0.00**
Seigneuric E0	21.41	2.33	3.75	0.00	15.58	12.96	2.50	3.92	1.39	27.78	0.00	0.00	0.00	**0.00**
Seigneuric E2	11.55	2.33	22.50	1.72	24.68	0.00	2.50	0.00	23.61	22.22	0.00	0.00	0.00	**0.00**
Shi	5.07	2.33	5.00	0.86	20.78	22.22	2.50	1.96	8.33	2.78	0.00	0.00	0.00	**0.00**
Shou *†	0.28	0.00	0.00	12.93	0.00	0.00	3.75	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
Suh *†	11.27	18.60	3.75	8.62	67.53	1.85	52.50	9.80	0.00	50.00	0.00	0.00	0.00	**0.00**
Sorensen	**97.75**	90.70	75.00	40.52	**93.51**	98.15	**95.00**	94.12	90.28	**100.00**	**100.00**	**100.00**	0.00	**0.00**
Starmans	83.38	76.74	62.50	**74.14**	89.61	53.70	67.50	41.18	93.06	97.22	**100.00**	75.00	**100.00**	**0.00**
Sun *†	0.56	0.00	10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
Sung	94.37	86.05	71.25	51.72	**93.51**	**100.00**	**95.00**	**98.04**	84.72	**100.00**	**100.00**	**100.00**	0.00	**0.00**
Tardon	93.80	88.37	77.50	33.62	**93.51**	**100.00**	**95.00**	90.20	94.44	**100.00**	**100.00**	**100.00**	**100.00**	**0.00**
Toustrup	97.18	**97.67**	71.25	39.66	**93.51**	96.30	90.00	47.06	80.56	97.22	**100.00**	**100.00**	**100.00**	**0.00**
Trong *	1.69	0.00	0.00	0.00	1.30	0.00	11.25	9.80	0.00	0.00	**100.00**	0.00	0.00	**0.00**
Van Malenstein	50.14	27.91	48.75	18.97	38.96	7.41	63.75	66.67	2.78	72.22	**100.00**	0.00	**100.00**	**0.00**
Wang 2005	86.76	86.05	60.00	50.00	92.21	85.19	92.50	41.18	88.89	97.22	**100.00**	75.00	0.00	**0.00**
Wang 2020 *†	60.85	48.84	47.50	5.17	76.62	11.11	48.75	19.61	50.00	58.33	50.00	75.00	0.00	**0.00**
Winter †	44.79	60.47	22.50	10.34	70.13	55.56	20.00	17.65	62.50	75.00	50.00	0.00	0.00	**0.00**
Yang 2017 †	61.69	4.65	41.25	11.21	68.83	9.26	40.00	23.53	41.67	77.78	25.00	0.00	**100.00**	**0.00**
Yang Prostate *†	39.44	48.84	33.75	14.66	40.26	22.22	61.25	9.80	26.39	77.78	0.00	0.00	0.00	**0.00**
Yang Sarcoma †	97.46	**97.67**	73.75	24.14	**93.51**	**100.00**	93.75	78.43	93.06	**100.00**	**100.00**	**100.00**	**100.00**	**0.00**
Ye	91.55	**97.67**	71.25	43.10	**93.51**	**100.00**	**95.00**	78.43	**95.83**	**100.00**	**100.00**	**100.00**	0.00	**0.00**
Zhang †	0.00	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
Zou *†	16.06	0.00	35.00	0.00	33.77	0.00	3.75	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**

b) Mean

	Breast	Prostate	Colorectal	Lung	Cervical	Ovarian	Liver	Brain	Skin	Pancreas	Burkitt lymphoma	Ewing's sarcoma	Bone	Fibrosarcoma
Aprelikova	78.03	83.72	56.25	77.59	**93.51**	33.33	81.25	96.08	29.17	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Benita	87.32	88.37	58.75	81.03	**93.51**	35.19	80.00	**98.04**	68.06	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Beyer	65.35	81.40	56.25	70.69	**93.51**	33.33	81.25	**98.04**	27.78	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Boidot Cont. *†	65.07	81.40	50.00	60.34	89.61	20.37	62.50	94.12	27.78	75.00	**100.00**	75.00	**100.00**	**0.00**
Boidot Cyc. *†	5.35	0.00	2.50	10.34	6.49	0.00	0.00	0.00	0.00	5.56	0.00	0.00	0.00	**0.00**
Bosco	85.92	81.40	56.25	68.10	90.91	33.33	81.25	96.08	33.33	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Buffa †	88.45	83.72	56.25	44.83	90.91	27.78	78.75	94.12	**75.00**	86.11	**100.00**	75.00	**100.00**	**0.00**
Chen	76.34	81.40	63.75	45.69	88.31	35.19	92.50	94.12	54.17	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Chi	70.42	88.37	55.00	81.03	**93.51**	33.33	82.50	**98.04**	19.44	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Denko	69.86	83.72	56.25	61.21	**93.51**	33.33	93.75	**98.04**	20.83	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Detwiller	58.31	81.40	50.00	45.69	85.71	33.33	91.25	92.16	41.67	91.67	**100.00**	**100.00**	**100.00**	**0.00**
Elvidge	68.45	81.40	58.75	72.41	**93.51**	33.33	81.25	96.08	25.00	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Eustace †	80.56	**97.67**	56.25	49.14	90.91	31.48	75.00	92.16	61.11	77.78	**100.00**	75.00	**100.00**	**0.00**
Fardin 2009	77.75	88.37	58.75	30.17	83.12	33.33	82.50	90.20	59.72	**100.00**	**100.00**	**100.00**	**100.00**	**0.00**
Fardin 2010	46.20	72.09	32.50	16.38	72.73	16.67	53.75	66.67	30.56	75.00	**100.00**	75.00	0.00	**0.00**
Fjeldbo †	34.65	53.49	52.50	57.76	45.45	25.93	58.75	**98.04**	12.50	72.22	0.00	**100.00**	**100.00**	**0.00**
Ghazoui †	81.69	83.72	32.50	37.93	88.31	16.67	67.50	84.31	41.67	75.00	**100.00**	0.00	0.00	**0.00**
Ghorbel †	4.79	23.26	10.00	9.48	2.60	1.85	1.25	1.96	0.00	11.11	0.00	**25.00**	0.00	**0.00**
Halle †	85.07	58.14	53.75	46.55	81.82	29.63	65.00	96.08	12.50	75.00	**100.00**	50.00	**100.00**	**0.00**
Hu †	48.45	88.37	58.75	43.97	89.61	33.33	83.75	96.08	50.00	**100.00**	**100.00**	**100.00**	**100.00**	**0.00**
Jogi	67.61	81.40	53.75	63.79	89.61	33.33	80.00	94.12	33.33	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Koong	43.10	55.81	45.00	19.83	77.92	31.48	75.00	96.08	66.67	97.22	**100.00**	75.00	0.00	**0.00**
Lendhal	71.27	88.37	58.75	60.34	90.91	33.33	83.75	94.12	55.56	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Lin †	**94.08**	**97.67**	**66.25**	18.10	84.42	**64.81**	91.25	92.16	73.61	**100.00**	**100.00**	50.00	**100.00**	**0.00**
Manalo	47.61	79.07	55.00	60.34	87.01	33.33	80.00	96.08	12.50	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Mense	69.01	88.37	57.50	**86.21**	92.21	33.33	82.50	**98.04**	22.22	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Mo †	34.93	79.07	46.25	14.66	62.34	27.78	48.75	**86.27**	12.50	66.67	0.00	25.00	0.00	**0.00**
Ning	56.62	48.84	17.50	7.76	76.62	16.67	8.75	19.61	11.11	61.11	**100.00**	0.00	0.00	**0.00**
Peters	36.90	34.88	16.25	11.21	71.43	9.26	0.00	11.76	9.72	13.89	**100.00**	0.00	0.00	**0.00**
Ragnum †	41.13	41.86	16.25	3.45	79.22	16.67	18.75	50.98	38.89	16.67	**100.00**	0.00	0.00	**0.00**
Seigneuric C.	**3.94**	0.00	0.00	0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
Seigneuric E0	0.00	0.00	1.25	**3.45**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
Seigneuric E2	14.37	**27.91**	2.50	0.00	27.27	0.00	0.00	0.00	13.89	25.00	0.00	0.00	0.00	**0.00**
Shi	11.27	0.00	0.00	0.00	2.60	0.00	33.75	29.41	8.33	**38.89**	0.00	0.00	0.00	**0.00**
Shou *†	0.28	0.00	0.00	**17.24**	0.00	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
Suh *†	15.77	53.49	7.50	25.86	32.47	11.11	**56.25**	17.65	0.00	30.56	0.00	0.00	0.00	**0.00**
Sorensen	71.55	79.07	58.75	76.72	90.91	33.33	83.75	96.08	23.61	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Starmans	61.41	81.40	52.50	**86.21**	90.91	16.67	62.50	88.24	31.94	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Sun *†	**1.41**	0.00	0.00	0.00	1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
Sung	70.70	83.72	58.75	81.03	**93.51**	33.33	83.75	**98.04**	36.11	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Tardon	80.28	95.35	58.75	69.83	**93.51**	33.33	93.75	**98.04**	54.17	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Toustrup	77.18	88.37	62.50	77.59	90.91	33.33	**95.00**	96.08	25.00	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Trong *	10.42	13.95	17.50	0.86	28.57	29.63	11.25	27.45	12.50	55.56	**100.00**	0.00	0.00	**0.00**
Van Malenstein	33.24	53.49	52.50	30.17	5.19	20.37	57.50	70.59	12.50	75.00	0.00	0.00	**100.00**	**0.00**
Wang 2005	72.11	88.37	58.75	58.62	**93.51**	33.33	83.75	96.08	36.11	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Wang 2020 *†	43.38	74.42	52.50	25.86	64.94	29.63	80.00	94.12	12.50	75.00	0.00	75.00	**100.00**	**0.00**
Winter †	77.18	79.07	42.50	46.55	89.61	16.67	58.75	84.31	56.94	75.00	**100.00**	0.00	0.00	**0.00**
Yang 2017 †	34.93	44.19	21.25	31.03	75.32	20.37	66.25	80.39	12.50	66.67	0.00	**100.00**	**100.00**	**0.00**
Yang Prostate *†	24.51	53.49	56.25	43.97	32.47	31.48	81.25	76.47	12.50	72.22	0.00	75.00	**100.00**	**0.00**
Yang Sarcoma †	70.42	88.37	58.75	73.28	92.21	33.33	93.75	**98.04**	41.67	75.00	**100.00**	**100.00**	**100.00**	**0.00**
Ye	81.97	86.05	58.75	81.03	90.91	33.33	93.75	96.08	62.50	83.33	**100.00**	**100.00**	**100.00**	**0.00**
Zhang †	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Zou *†	19.72	18.60	0.00	5.17	0.00	0.00	33.75	54.90	8.33	41.67	0.00	0.00	0.00	**0.00**

c) GSVA

	Breast	Prostate	Colorectal	Lung	Cervical	Ovarian	Liver	Brain	Skin	Pancreas	Burkitt lymphoma	Ewing's sarcoma	Bone	Fibrosarcoma
Aprelikova	94.08	93.02	70.00	68.10	97.40	94.44	88.75	90.20	83.33	100.00	100.00	100.00	100.00	0.00
Benita	86.76	88.37	60.00	66.38	87.01	81.48	81.25	82.35	69.44	100.00	100.00	100.00	100.00	0.00
Beyer	89.86	95.35	68.75	52.59	87.01	92.59	83.75	86.27	70.83	100.00	100.00	100.00	100.00	0.00
Boidot Cont. *†	46.20	55.81	32.50	23.28	64.94	61.11	55.00	50.98	52.78	58.33	50.00	50.00	0.00	0.00
Boidot Cyc. *†	1.41	0.00	0.00	13.79	0.00	7.41	0.00	1.96	0.00	5.56	0.00	0.00	0.00	0.00
Bosco	77.46	81.40	66.25	50.86	84.42	62.96	81.25	68.63	73.61	94.44	100.00	75.00	100.00	0.00
Buffa †	43.38	67.44	27.50	21.55	53.25	40.74	41.25	35.29	41.67	19.44	100.00	0.00	0.00	100.00
Chen	45.63	53.49	38.75	21.55	61.04	42.59	50.00	54.90	37.50	66.67	25.00	100.00	0.00	0.00
Chi	88.73	88.37	52.50	62.07	92.21	87.04	82.50	82.35	55.56	100.00	100.00	100.00	100.00	0.00
Denko	80.28	88.37	66.25	33.62	88.31	79.63	80.00	84.31	61.11	100.00	100.00	75.00	100.00	0.00
Detwiller	39.15	69.77	20.00	21.55	48.05	24.07	58.75	33.33	16.67	66.67	100.00	75.00	0.00	0.00
Elvidge	93.24	93.02	70.00	57.76	88.31	90.74	85.00	88.24	73.61	100.00	100.00	100.00	100.00	0.00
Eustace †	49.30	62.79	40.00	33.62	79.22	37.04	57.50	47.06	45.83	61.11	100.00	50.00	0.00	100.00
Fardin 2009	29.01	44.19	36.25	13.79	45.45	33.33	48.75	33.33	20.83	47.22	50.00	25.00	0.00	0.00
Fardin 2010	7.32	2.33	1.25	2.59	9.09	9.26	15.00	11.76	0.00	5.56	0.00	0.00	0.00	0.00
Fjeldbo †	13.52	18.60	18.75	25.86	40.26	1.85	46.25	23.53	6.94	30.56	25.00	0.00	0.00	0.00
Ghazoui †	15.49	48.84	18.75	23.28	12.99	27.78	40.00	21.57	13.89	11.11	50.00	0.00	0.00	0.00
Ghorbel †	7.89	27.91	11.25	5.17	6.49	12.96	6.25	13.73	11.11	8.33	0.00	25.00	0.00	0.00
Halle †	51.55	53.49	33.75	18.10	55.84	53.70	65.00	49.02	22.22	55.56	50.00	25.00	0.00	0.00
Hu †	13.52	30.23	17.50	7.76	18.18	35.19	51.25	45.10	6.94	33.33	100.00	25.00	0.00	0.00
Jogi	65.92	72.09	42.50	40.52	79.22	62.96	66.25	78.43	55.56	80.56	50.00	75.00	100.00	0.00
Koong	21.97	37.21	17.50	3.45	23.38	22.22	45.00	41.18	2.78	58.33	0.00	0.00	0.00	0.00
Lendhal	63.38	86.05	48.75	31.90	75.32	55.56	63.75	64.71	61.11	69.44	50.00	25.00	100.00	0.00
Lin †	24.23	32.56	28.75	6.90	57.14	29.63	56.25	33.33	6.94	47.22	50.00	25.00	0.00	0.00
Manalo	76.06	83.72	56.25	47.41	72.73	68.52	72.50	80.39	58.33	97.22	75.00	100.00	100.00	0.00
Mense	91.55	93.02	67.50	73.28	85.71	81.48	82.50	82.35	68.06	100.00	100.00	100.00	100.00	0.00
Mo †	9.01	11.63	18.75	6.90	16.88	5.56	33.75	29.41	2.78	30.56	0.00	0.00	0.00	0.00
Ning	9.01	46.51	6.25	0.00	6.49	25.93	7.50	3.92	9.72	5.56	0.00	0.00	0.00	0.00
Peters	5.92	30.23	10.00	6.90	3.90	25.93	10.00	1.96	4.17	2.78	0.00	0.00	0.00	0.00
Ragnum †	8.73	16.28	6.25	3.45	9.09	9.26	7.50	5.88	2.78	2.78	50.00	0.00	0.00	0.00
Seigneuric C.	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Seigneuric E0	0.85	0.00	1.25	0.86	0.00	0.00	0.00	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Seigneuric E2	3.66	2.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.78	0.00	0.00	0.00	0.00
Shi	2.54	0.00	2.50	0.00	1.30	1.85	10.00	11.76	4.17	19.44	0.00	0.00	0.00	0.00
Shou *†	0.85	0.00	0.00	2.59	0.00	0.00	6.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Suh *†	1.41	18.60	3.75	1.72	15.58	1.85	35.00	1.96	0.00	38.89	0.00	0.00	0.00	0.00
Sorensen	84.79	88.37	60.00	63.79	84.42	81.48	75.00	78.43	54.17	94.44	100.00	100.00	100.00	0.00
Starmans	84.79	90.70	68.75	80.17	87.01	55.56	81.25	78.43	88.89	91.67	100.00	100.00	100.00	0.00
Sun *†	0.00	0.00	0.00	0.00	1.30	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sung	90.99	95.35	82.50	66.38	87.01	92.59	87.50	86.27	66.67	100.00	100.00	100.00	100.00	0.00
Tardon	75.77	83.72	61.25	35.34	87.01	64.81	73.75	80.39	51.39	100.00	75.00	100.00	100.00	0.00
Toustrup	66.20	79.07	52.50	52.59	80.52	72.22	67.50	64.71	52.78	88.89	100.00	100.00	100.00	0.00
Trong *	1.41	13.95	0.00	0.00	2.60	1.85	16.25	1.96	9.72	5.56	0.00	25.00	0.00	0.00
Van Malenstein	7.89	16.28	20.00	7.76	11.69	1.85	38.75	17.65	5.56	38.89	25.00	0.00	0.00	0.00
Wang 2005	73.52	81.40	62.50	37.93	81.82	55.56	75.00	82.35	58.33	83.33	50.00	100.00	100.00	0.00
Wang 2020 *†	11.27	32.56	15.00	8.62	24.68	7.41	48.75	39.22	9.72	41.67	0.00	50.00	0.00	0.00
Winter †	13.80	55.81	18.75	16.38	27.27	42.59	36.25	13.73	20.83	44.44	50.00	0.00	0.00	0.00
Yang 2017 †	41.97	39.53	13.75	11.21	57.14	33.33	57.50	45.10	36.11	63.89	25.00	50.00	0.00	0.00
Yang Prostate *†	24.51	51.16	61.25	18.10	29.87	40.74	66.25	43.14	15.28	72.22	0.00	0.00	0.00	0.00
Yang Sarcoma †	81.97	90.70	60.00	55.17	84.42	75.93	72.50	84.31	65.28	97.22	100.00	100.00	100.00	0.00
Ye	90.70	93.02	65.00	59.48	89.61	83.33	77.50	82.35	72.22	100.00	100.00	100.00	100.00	100.00
Zhang †	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Zou *†	3.10	0.00	6.25	0.86	3.90	0.00	20.00	11.76	1.39	5.56	25.00	0.00	0.00	0.00

d) ssGSEA

	Breast	Prostate	Colorectal	Lung	Cervical	Ovarian	Liver	Brain	Skin	Pancreas	Burkitt lymphoma	Ewing's sarcoma	Bone	Fibrosarcoma
Aprelikova	0.85	0.00	0.00	1.72	1.30	0.00	1.25	0.00	2.78	2.78	0.00	0.00	0.00	0.00
Benita	0.56	0.00	3.75	1.72	1.30	0.00	0.00	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Beyer	1.69	0.00	2.50	0.86	1.30	1.85	1.25	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Boidot Cont. *†	0.28	0.00	2.50	2.59	1.30	0.00	1.25	0.00	2.78	0.00	25.00	0.00	0.00	0.00
Boidot Cyc. *†	0.28	0.00	0.00	1.72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bosco	1.13	2.33	0.00	0.86	1.30	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Buffa †	0.85	0.00	1.25	0.00	1.30	1.85	2.50	0.00	0.00	0.00	25.00	0.00	0.00	0.00
Chen	0.28	0.00	0.00	0.00	2.60	0.00	0.00	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Chi	1.97	0.00	0.00	0.86	1.30	0.00	0.00	1.96	1.39	2.78	0.00	0.00	0.00	0.00
Denko	0.85	0.00	0.00	0.86	0.00	1.85	1.25	1.96	0.00	0.00	0.00	0.00	0.00	0.00
Detwiller	0.85	0.00	0.00	0.00	1.30	1.85	0.00	1.96	1.39	0.00	0.00	0.00	0.00	0.00
Elvidge	1.41	2.33	0.00	0.86	0.00	1.85	0.00	0.00	1.39	2.78	0.00	0.00	0.00	0.00
Eustace †	1.13	0.00	1.25	0.86	1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fardin 2009	1.13	0.00	0.00	0.86	1.30	1.85	0.00	0.00	0.00	2.78	0.00	0.00	0.00	0.00
Fardin 2010	0.28	4.65	1.25	0.00	0.00	1.85	0.00	0.00	1.39	2.78	0.00	0.00	0.00	0.00
Fjeldbo †	0.28	0.00	1.25	0.00	0.00	1.85	0.00	0.00	0.00	2.78	25.00	0.00	0.00	0.00
Ghazoui †	0.56	0.00	1.25	0.86	0.00	1.85	1.25	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Ghorbel †	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Halle †	0.56	0.00	0.00	1.72	0.00	1.85	1.25	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Hu †	0.85	0.00	1.25	0.00	0.00	0.00	1.25	1.96	0.00	0.00	0.00	0.00	0.00	0.00
Jogi	0.28	2.33	0.00	0.86	0.00	0.00	0.00	1.96	0.00	0.00	0.00	0.00	0.00	0.00
Koong	0.85	0.00	2.50	0.86	0.00	0.00	1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lendhal	0.00	0.00	1.25	2.59	0.00	0.00	1.25	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Lin †	1.13	0.00	0.00	2.59	0.00	0.00	1.25	1.96	0.00	0.00	0.00	0.00	0.00	0.00
Manalo	0.85	4.65	0.00	1.72	0.00	0.00	1.25	0.00	0.00	2.78	0.00	25.00	0.00	0.00
Mense	1.69	0.00	0.00	0.86	0.00	0.00	0.00	0.00	1.39	2.78	0.00	0.00	0.00	0.00
Mo †	1.13	2.33	0.00	0.00	0.00	0.00	1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ning	0.00	0.00	0.00	0.00	1.30	1.85	0.00	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Peters	0.28	0.00	1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ragnum †	1.41	2.33	0.00	0.00	0.00	1.85	1.25	1.96	0.00	0.00	0.00	0.00	0.00	0.00
Seigneuric C.	0.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Seigneuric E0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.96	0.00	0.00	0.00	0.00	0.00	0.00
Seigneuric E2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Shi	0.28	2.33	0.00	0.00	0.00	0.00	0.00	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Shou *†	0.28	0.00	0.00	0.00	0.00	0.00	2.50	1.96	1.39	0.00	0.00	0.00	0.00	0.00
Suh *†	0.85	0.00	0.00	0.86	0.00	0.00	0.00	1.96	1.39	0.00	0.00	0.00	0.00	0.00
Sorensen	1.13	0.00	0.00	1.72	0.00	0.00	0.00	1.96	1.39	0.00	0.00	0.00	0.00	0.00
Starmans	0.28	0.00	0.00	2.59	0.00	0.00	1.25	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Sun *†	0.00	0.00	0.00	0.00	0.00	1.85	1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sung	0.56	0.00	0.00	0.86	1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tardon	1.13	2.33	0.00	3.45	1.30	1.85	1.25	1.96	0.00	0.00	0.00	25.00	0.00	0.00
Toustrup	1.97	0.00	2.50	2.59	0.00	0.00	1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trong *	0.85	2.33	0.00	0.86	0.00	0.00	1.25	0.00	2.78	0.00	0.00	0.00	0.00	0.00
Van Malenstein	0.85	2.33	1.25	0.86	0.00	0.00	0.00	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Wang 2005	1.13	0.00	0.00	0.86	0.00	1.85	1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Wang 2020 *†	1.13	0.00	0.00	3.45	0.00	1.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Winter †	0.56	0.00	0.00	1.72	1.30	0.00	1.25	0.00	0.00	2.78	0.00	0.00	0.00	0.00
Yang 2017 †	1.13	0.00	0.00	2.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yang Prostate *†	0.85	0.00	0.00	0.86	1.30	0.00	1.25	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Yang Sarcoma †	0.85	2.33	3.75	2.59	0.00	0.00	0.00	0.00	2.78	0.00	0.00	0.00	0.00	0.00
Ye	1.41	0.00	0.00	1.72	0.00	1.85	0.00	0.00	1.39	0.00	0.00	0.00	0.00	0.00
Zhang †	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Zou *†	0.85	2.33	2.50	0.00	0.00	0.00	1.25	0.00	1.39	0.00	0.00	0.00	0.00	0.00

Table 4.25: *Percentage accuracy of determining hypoxic samples from normoxic samples in experiments involving cancer cell lines/samples for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue) for all the cancer cell lines and the four scores: Median *a)*, Mean *b)*, GSVA *c)* and ssGSEA *d)*. Conventions as in *Table 4.2*.

Chapter 5: Performance of hypoxia gene expression signature scores in cells with common driver mutations under hypoxia and in microenvironmental conditions

5.1 Key mutations in cancer and hypoxia

Germline (mutations present since birth) and somatic (acquired mutations) form a key part in carcinogenesis. Common germline mutations that ultimately result in cancer include proteins such as BRCA1 and BRCA2. The prevalence of BRCA1 and BRCA2 mutations in the western population has been estimated as approximately 1 in 400 individuals[126,127] although different ethnic groups have been shown to have higher frequencies, e.g. within individuals of Ashkenazi Jewish descent estimates of around 1 in 40-50 individuals[128–130]. Both BRCA1 and BRCA2 are involved in DNA damage repair and are important tumour suppressor genes. Individuals harbouring BRCA1/2 mutations are at an increased risk of breast, ovarian, prostate and pancreatic tumours. Remarkably, carriers of mutations in BRCA2 have lifetime risk of developing pancreatic cancer between 1 in 10 to 1 in 20[131,132].

Although less common in the population, highly penetrant germline mutations of tumour suppressor genes and proto-oncogenes can also lead to cancer. One example is a mutation in the tumour suppressor gene retinoblastoma (RB1) that is a key regulator of the G1/S checkpoint in the cell cycle[133]. The protein transcribed from RB1, pRB, is active in quiescent cells (in G1 of the cell cycle) and cell cycle arrest. Indeed, hyperphosphorylation of pRB is needed at the G1/S transition to relieve inhibition of the transcription factor E2F to allow cell cycle progression[134]. Individuals carrying germline mutations in one copy of their RB1 gene are at markedly increased risk of developing malignant tumours of the retinal cells (retinoblastomas).

Germline mutations can influence tumour suppressor proteins depending on their chromosomal location. For instance, mutations in the protein coding region of the “guardian of the genome” TP53 can lead to the very penetrant Li-Fraumeni syndrome which is associated with an increased risk of a number of cancers including soft tissue sarcoma,

osteosarcoma, adrenocortical carcinoma and cancers of the breast, brain and bone marrow. Individuals with Li-Fraumeni syndrome have ~50% chance of developing cancer by 40 years of age and ~90% chance by 60 years of age[135]. However, other more common mutations in the 3'-UTR of TP53 have been found at 1% of the European population. These have been shown to reduce protein and mRNA of TP53 and are associated with an increased risk of non-melanomatous skin cancers (basal cell carcinomas) and malignancies of the brain (gliomas) in a number of different cohorts[136,137]. Perhaps being evolutionarily selected for as carriers have increased fat-free mass and height compared to those without it[137]. Also carriers also might have potentially preferentially altered oxidative phosphorylation, in line with observations seen in Li-Fraumeni syndrome[138].

Genome-wide approaches to associate germline variants with cancer incidence and profiling tumours have both yielded successes in identifying important cancer drivers and therapeutic targets. Whereas genome-wide association studies (GWASs) harness genomic sequencing and results on millions of SNPs and relate these to outcomes using general linear models. Importantly, GWAS undercovers loci that influence cancer risk in the general population, not just in rare diseases. The explosion of genomic data, the reduction in sequencing costs and associated growth in GWAS studies has led to a huge number of SNPs being associated with differential cancer risk. Associations to more common malignancies, such as lung and breast cancer have been more well studied[139,140] compared to rarer cancers. One interesting discovery is the SNP rs1051730[A] being strongly associated with lung cancer. rs1051730[A] is found within the subunit of the nicotinic acetylcholine receptor (neuronal acetylcholine receptor subunit alpha-3, CHRNA3) and was found to be strongly associated with lung cancer, nicotine dependence and smoking quantity. This SNP provided the first genetic link between tobacco addiction and lung cancer[141].

Somatic mutations can occur in any of the cells of the body except the germ cells and so are not passed onto offspring. They tend to occur following DNA damage secondary to events such as exposure to ionising radiation, exposure to mutagenic chemicals

(endogenous or exogenous), DNA replication errors and even infection by certain viruses. Such events cause DNA alterations including base substitutions, insertions and deletions (indels), chromosomal rearrangements and copy-number changes. Somatic mutations by some have been thought to underlie the majority of cancer cases in humans, although this has been widely debated[142]. Indeed, so called somatic "driver mutations" confer growth advantages and so are positively selected for in tumorigenesis. However, 'passenger' mutations also exist that do not confer potential survival advantages for the cancer cell and have been described as biologically neutral[143] and unpicking which mutations are "driver" and which are "passenger" has been a major challenge in cancer biology.

Tumour mutational profiling also is helpful in determining somatic mutations. Key projects focussing on somatic mutations in tumours include the International Cancer Genome Consortium project (ICGC), the Cancer Genome Atlas (TCGA) and the Catalogue Of Somatic Mutations In Cancer (COSMIC). In brief, the International Cancer Genome Consortium project spans 50 cancer types, including 25,000 cancer samples; the TCGA includes profiling of 20,000 primary cancer and matched normal samples in 33 cancer types and COSMIC combines these datasets and others to outline the mutational profile of over 1 million tumours[144]. Such efforts have allowed the estimation of frequency of mutations within different cancer types and helped formulate hypotheses surrounding repurposing drugs across cancer types.

Indeed, the most mutated gene in the human genome is the aforementioned TP53[145,146]. The main physiological role of the p53 tumour suppressor protein is to prevent the replications of cells whose DNA is damaged via different molecular mechanisms (*Fig. 5.1*). It has been termed a 'cellular gatekeeper'[147,148], maintaining the integrity of the genome in somatic cells. The concentration of the TP53 transcription factor is controlled and kept low by ubiquitination by MDM2 (Mouse Double Minute 2, human homolog), and subsequent proteolysis[149]. The p53 protein can sense when a DNA damage occurs either by its C-terminal domain or ATM (ATM phosphorylates and switches off MDM2[150]).

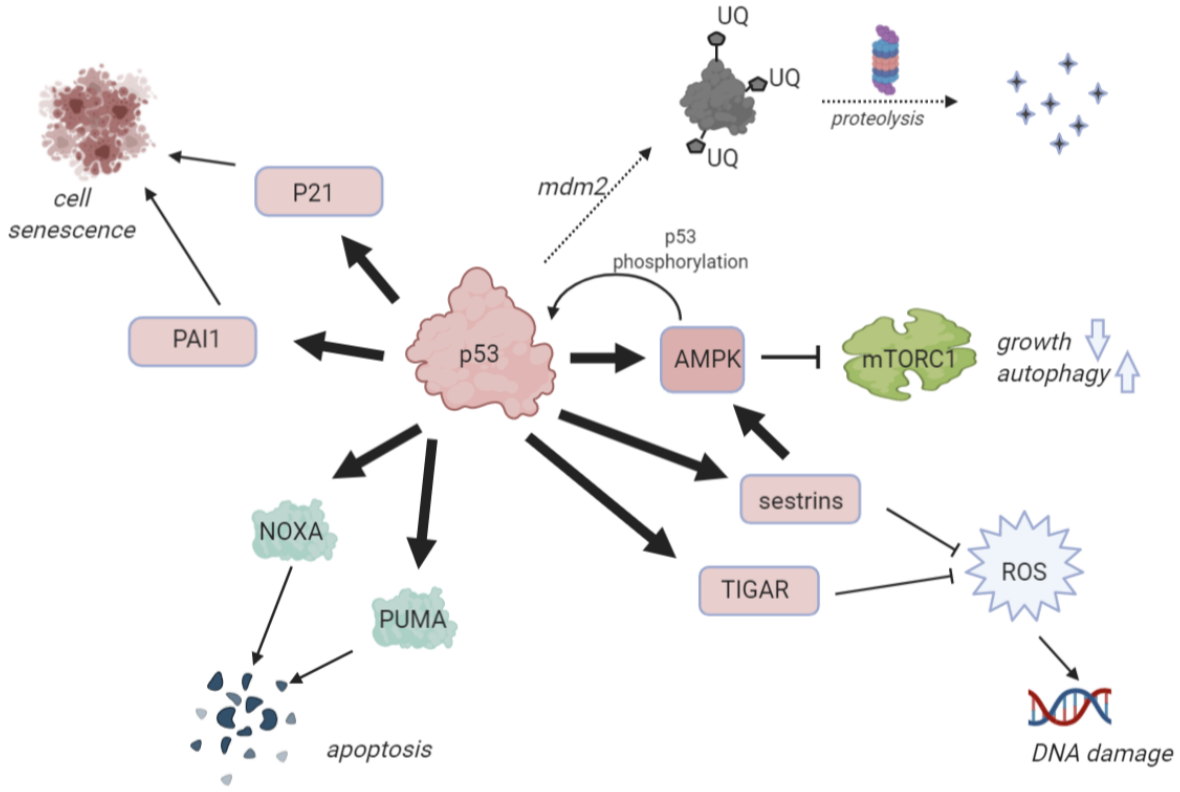


Figure 5.1: *Key TP53 cellular functions*

MDM2 ubiquitination reduces p53 cellular levels by proteolysis. If levels of TP53 rise, e.g. secondary to DNA damage, growth will be suppressed (e.g via mTOR) and apoptosis will be encouraged (e.g via induction of NOXA and PUMA). Figure created with BioRender, personal communication from Dr. B. Harris.

Another key mutation which is found in ~40% of renal cancers is that of the protein linked to the hypoxia response, VHL (function discussed in Chapter 1 and types of mutation detailed in *Fig. 5.2*). Looking at other key parts of the hypoxia response you also can see mutations. For instance, HIF-1α is found to be mutated in ~5% of endometrial cancers and the gene encoding PHD1, otherwise known as EGLN2, is found to be mutated in ~2%[144]. Also some mutations are thought to be enriched in hypoxic tumours.

In a recent paper, Bhandari et al found that driver mutations in TP53, MYC and PTEN are enriched in hypoxic tumours and stated that hypoxia plays a "critical role in shaping the genomic and evolutionary landscapes of cancer"[151]. They also interestingly report that hypoxic tumours which also have a PTEN mutation tend to have a polyclonal architecture and which confer a more aggressive phenotype. In this analysis the authors use median Buffa hypoxia signature to define hypoxic samples in the TCGA and a subset of cancers from the International Cancer Genome Consortium project (the Pan-Cancer Analysis of Whole Genomes dataset) but there was no detailed discussion surrounding the controversies surrounding summarising hypoxia signatures or comprehensive comparison of signatures (three were tested) leaving key questions unanswered which hopefully this work will help address. Also the hypoxia signature itself might perform differently in different mutational profiles and thus change ultimate conclusions, e.g. under certain mutational profiles the gene expression of some elements of the signature might be altered by elements other than hypoxia. Thus, as a next step to assess the utility of hypoxia signatures and their scores this study moves onto assessing the performance in experiments using cell lines with knockouts of key genes found in cancer and indeed within the response hypoxia pathway.

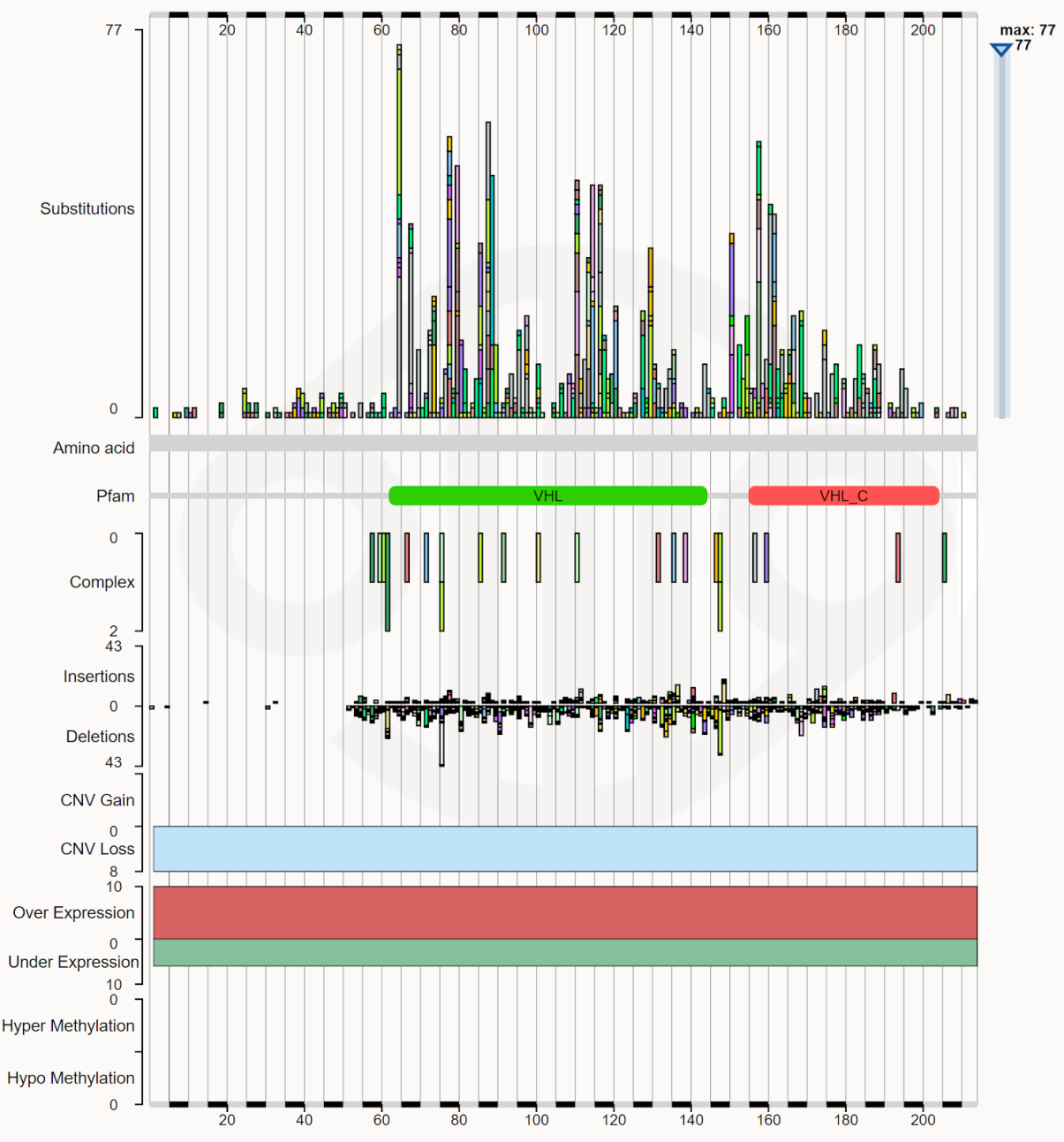


Figure 5.2: *Graphical view of mutations across VHL in renal cancer from Cosmic database*

Mutations are displayed at the amino acid level across the full length of the gene. This default peptide view shows a histogram of single base substitutions, colour coded by residue according to the colour scheme used in Ensembl. Under this is shown the amino acid

sequence and the Pfam protein structures, followed by complex mutations and insertions and deletions. The figure and its description are adapted from the Cosmic database.

5.2 The performance of hypoxia signatures in TP53 KO experiments

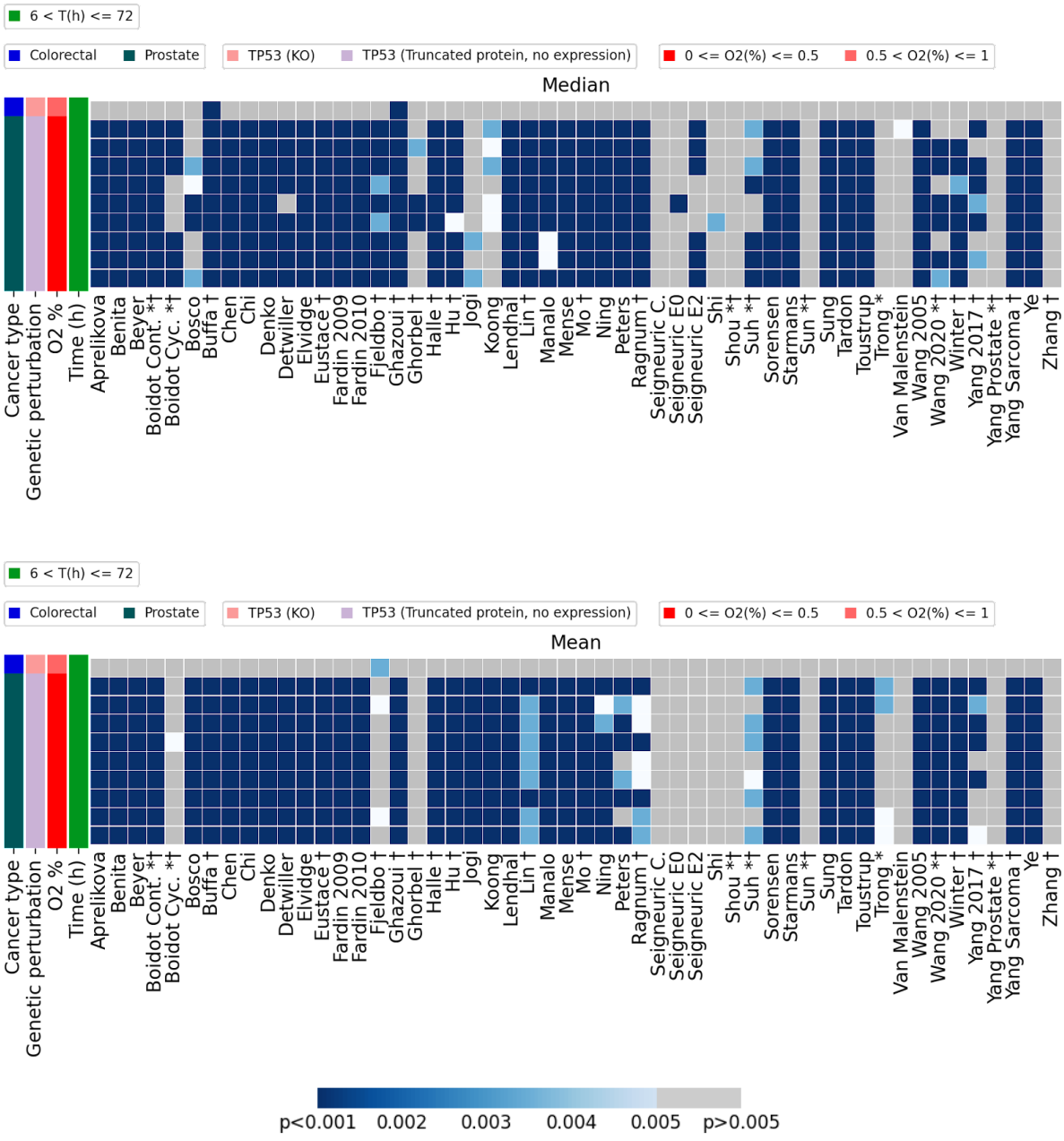
Two cancer cell lines (PC-3 and HCT116) were identified that had gene expression data in hypoxia and normoxia with no TP53 protein expression from the Gene Expression Omnibus (as in Methods). The performance of the 53 signatures at defining hypoxia in these two cell lines using gene expression data was assessed from two datasets *a)* GSE80657 (stop codon in TP53 mRNA resulting in no protein expression in PC-3, a prostate cancer cell line) - nine pairwise combinations, *b)* GSE42416 (recombinant adeno-associated virus TP53 knockout in HCT116, a colon cancer cell line) - one pairwise combination (*Table 5.1*). While acknowledging that this may not reflect most in vivo situations of carcinogenesis—where, barring germline mutations (e.g. in Li-Fraumeni syndrome/carriers of the rs78378222[C] SNP[137]), most normal cells maintain normal levels of TP53.

Mean and median performed well. Most signatures reached 90-100% accuracy using one of these three scores but Sengeiuric common, Shou, Sun and Zhang achieved 0% accuracy on all scores (*Table 5.2*). With TP53 mutations being the most common recorded in cancer cells, it is reassuring that these hypoxia signatures seem to work even in the presence of no TP53 expression and perhaps reflects that fact that some signatures may have been derived from cells with known/unknown aberrations of TP53. The best performing score from Chapter 4.12, Aprelikova using GSVA, showed 100% accuracy in the TP53 KO experiments. ssGSEA was the worst performing score (*Fig. 5.3*, last panel).

Cell line/samples	Number of Hypoxic samples (KO)	Number of hypoxic samples (P53 truncated protein, no expression)	Number of Normoxic samples (no P53 expression)	Pairs
PC-3	0	3	3	9
HCT116	1	0	1	1
Total	**1**	**3**	**4**	**10**

Table 5.1: *P53 truncated protein cancer cell lines used in hypoxia experiments in GEO*

Details of the hypoxia P53 truncated protein versus normoxia experiments in GEO using renal cancer cell lines.


6 < T(h) <= 72
Colorectal
Prostate
TP53 (KO)
TP53 (Truncated protein, no expression)
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
Median
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fieldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup *
Trong
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †

6 < T(h) <= 72
Colorectal
Prostate
TP53 (KO)
TP53 (Truncated protein, no expression)
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
Mean
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fieldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup *
Trong
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

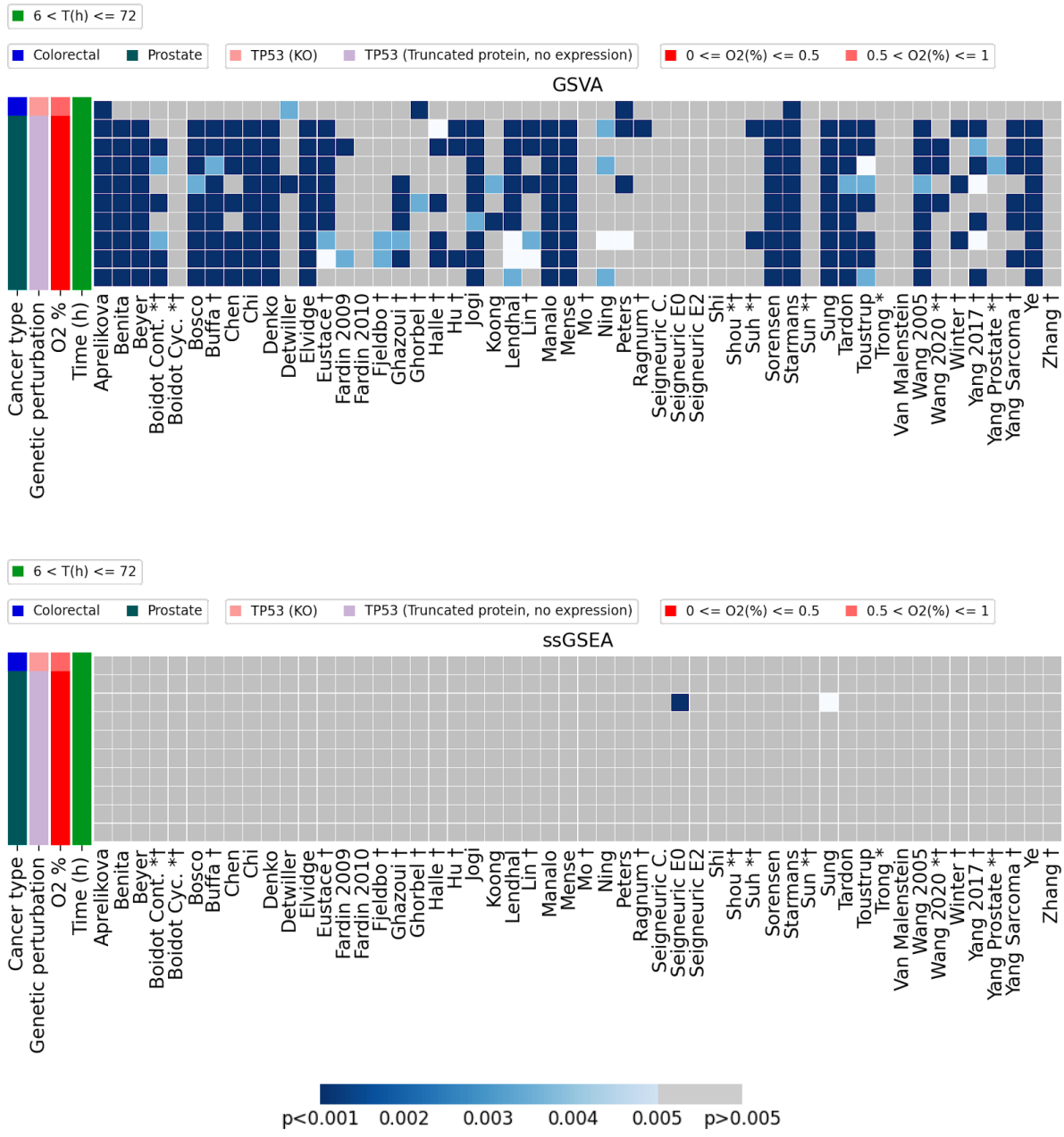


Figure 5.3: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in cancer cell lines with no TP53 expression*

TP53 mutations are particularly common in cancer and this shows that even in the scenario of complete protein depletion many signatures are still effective. The first bar on the left indicates the cancer type of the samples. The other three bars report the genetic perturbation, percentage of oxygen and time (in hours) under hypoxia for the hypoxic

sample. For simplicity I will refer to the hypoxic and normoxic samples as S1 and S2 respectively.

	Median	Mean	GSVA	ssGSEA
Aprelikova	90.00	90.00	100.00	0.00
Benita	90.00	90.00	90.00	0.00
Beyer	90.00	90.00	90.00	0.00
Boidot Cont. *†	90.00	90.00	60.00	0.00
Boidot Cyc. *†	60.00	10.00	0.00	0.00
Bosco	50.00	90.00	90.00	0.00
Buffa †	100.00	90.00	90.00	0.00
Chen	90.00	90.00	70.00	0.00
Chi	90.00	90.00	90.00	0.00
Denko	90.00	90.00	90.00	0.00
Detwiller	80.00	90.00	20.00	0.00
Elvidge	90.00	90.00	90.00	0.00
Eustace †	90.00	90.00	80.00	0.00
Fardin 2009	90.00	90.00	20.00	0.00
Fardin 2010	90.00	90.00	0.00	0.00
Fjeldbo †	90.00	30.00	20.00	0.00
Ghazoui †	100.00	90.00	50.00	0.00
Ghorbel †	30.00	0.00	20.00	0.00
Halle †	90.00	90.00	50.00	0.00
Hu †	90.00	90.00	30.00	0.00
Jogi	20.00	90.00	90.00	0.00
Koong	50.00	90.00	20.00	0.00
Lendhal	90.00	90.00	90.00	0.00
Lin †	90.00	90.00	50.00	0.00
Manalo	90.00	90.00	90.00	0.00
Mense	90.00	90.00	90.00	0.00
Mo †	90.00	90.00	0.00	0.00
Ning	90.00	90.00	40.00	0.00
Peters	90.00	70.00	40.00	0.00
Ragnum †	90.00	90.00	10.00	0.00
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	10.00	0.00	0.00	10.00
Seigneuric E2	60.00	0.00	0.00	0.00
Shi	10.00	0.00	0.00	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	60.00	60.00	20.00	0.00
Sorensen	90.00	90.00	90.00	0.00
Starmans	90.00	90.00	100.00	0.00
Sun *†	0.00	0.00	0.00	0.00
Sung	90.00	90.00	90.00	10.00
Tardon	90.00	90.00	90.00	0.00
Toustrup	90.00	90.00	80.00	0.00
Trong *	0.00	40.00	0.00	0.00
Van Malenstein	10.00	0.00	0.00	0.00
Wang 2005	90.00	90.00	90.00	0.00
Wang 2020 *†	60.00	90.00	30.00	0.00
Winter †	80.00	90.00	30.00	0.00
Yang 2017 †	60.00	60.00	70.00	0.00
Yang Prostate *†	0.00	0.00	10.00	0.00
Yang Sarcoma †	90.00	90.00	60.00	0.00
Ye	90.00	90.00	90.00	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	20.00	0.00	40.00	0.00

Table 5.2: *Percentage accuracy of determining hypoxic samples from normoxic samples in cancer cell lines with no TP53 expression for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

5.3 The performance of hypoxia signatures in VHL mutated cells and when VHL is reintroduced

VHL mutations are commonly found in renal cancer cell lines and indeed these were the only cell lines identified with VHL mutations subjected to hypoxia. The renal cancer cell lines which had gene expression data in hypoxia and normoxia were identified from the Gene Expression Omnibus (detailed in Chapter 2.4). This yielded three cell lines: A498, RCC4 and 786-O. All three cell lines harbour a VHL mutation[152], although the existence of the VHL mutation in 786-O has been debated[153]. A498 was the most common cell line used (*Table 5.3*). With a VHL mutation present, one may expect HIF-1α activation irrespective of hypoxic status; thus this may affect hypoxia signature accuracy (e.g. HIF-1α regulated genes may well be upregulated in both normoxia and hypoxic environments).

I then assessed the performance of the 53 signatures at defining hypoxia using five datasets from GEO (GSE107848, GSE117775, GSE42416, GSE85353, GSE65168). Overall, signatures performed worse across the renal cancer cell lines compared to their performance in other tissue types (*Fig. 5.4* and *Table 5.4*, compared to all sections in Chapter 4). This is likely because the majority of the signatures contain HIF-1α regulated genes. Such genes are upregulated in the context of VHL loss-of-function mutations as HIF-1α which is stabilised in the absence of VHL. The best performing signature and score was the nasopharyngeal cell line-derived signature, Sung (90 gene symbols) using the GVSA score which could accurately differentiate between hypoxic and normoxic samples in 81.25% of cases (*Table 5.4*). Lendhal is the only signature that incorporated a VHL mutated cell line into its genesis (RCC4). Surprisingly, Lendhal only managed 56.25% accuracy using the GVSA score and other scores achieved less than 35% accuracy.

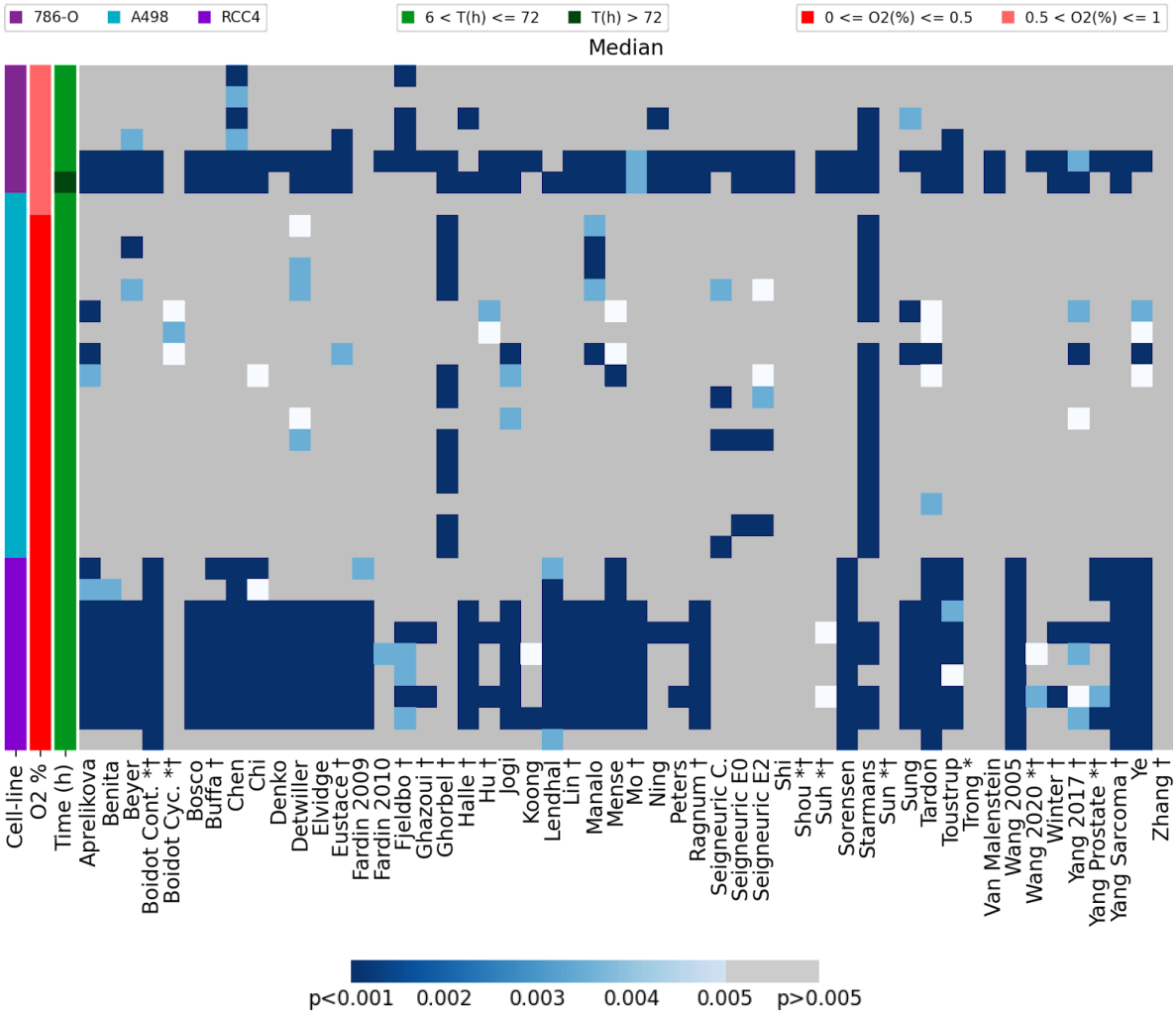
Also, the performance of the signatures within the 786-O renal cancer cell line when VHL is reintroduced was explored, although pairwise combinations of hypoxic and normoxic samples were small (4 pairwise combinations from GEO Series GSE65168, *Fig. 5.5*)[154]. 31

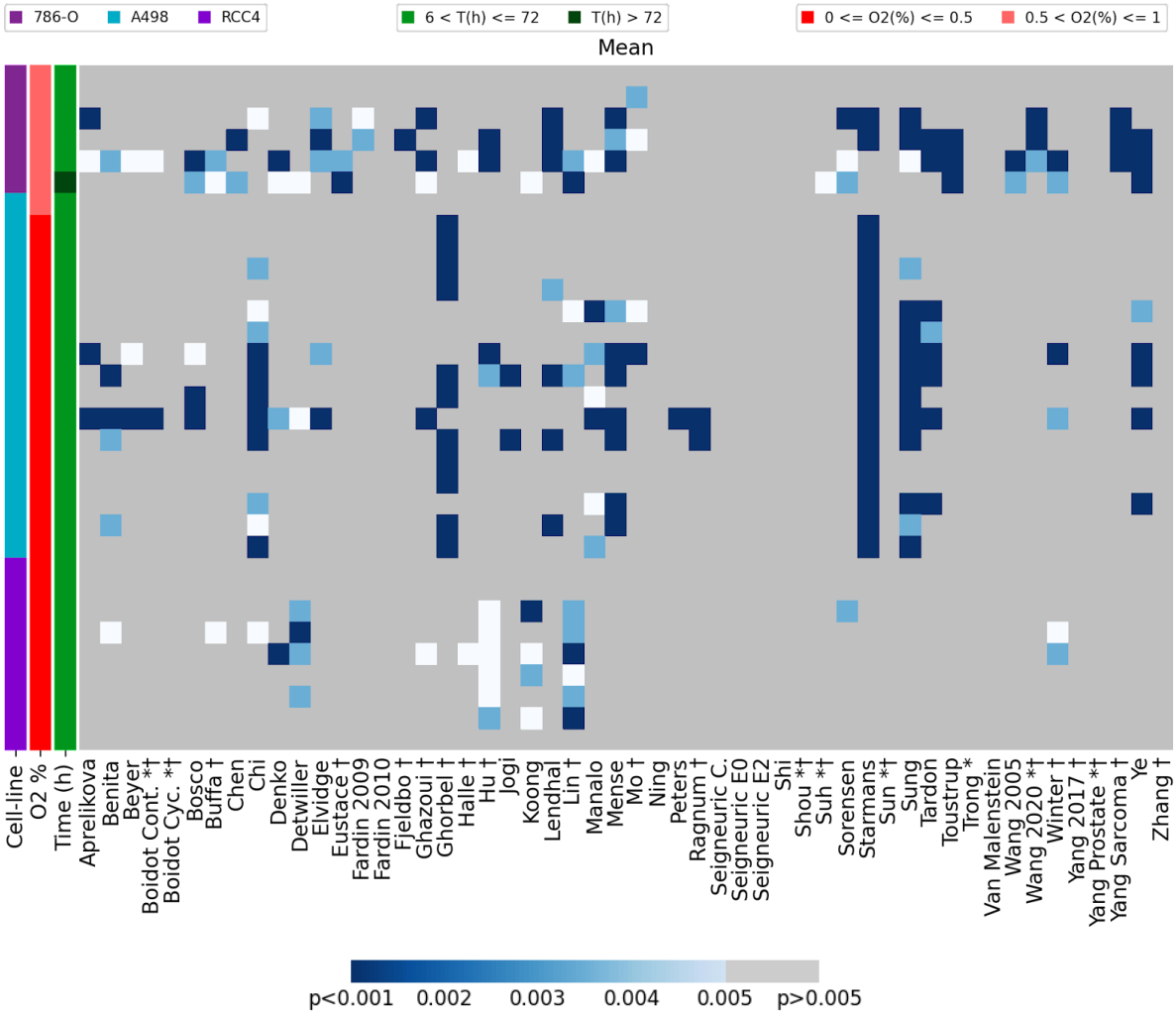
of the 53 signatures reached 100% accuracy in mean and nine did so using the median score. The best performing signature across all cancer types, Aprelikova using the GSVA score, only correctly classified 50% of samples using the GSVA but did reach 100% in using the mean (*Table 5.5*). This perhaps suggests even in the presence of a functional VHL, renal cancer cell response to hypoxia might be different, so perhaps highlight the need for a renal specific hypoxia signature.

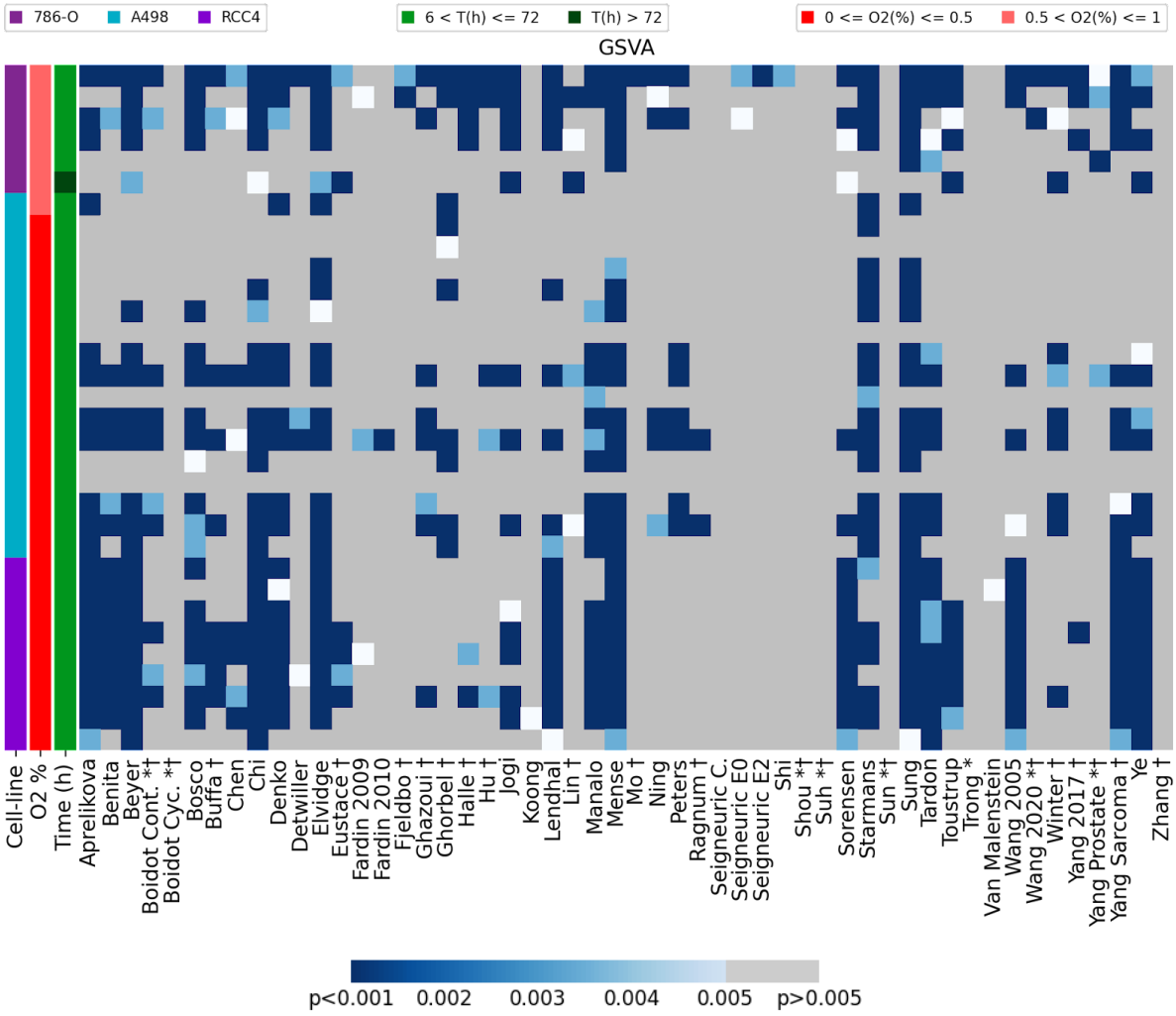
Cell line/samples	Number of Hypoxic samples	Number of Normoxic samples	Pairs
A498	5	5	17
RCC4	3	3	9
786-O	4	3	6
Total	**12**	**11**	**32**

Table 5.3: *Renal cancer cell lines used in hypoxia experiments in GEO*

Details of the hypoxia versus normoxia experiments in GEO using renal cancer cell lines.


786-O
A498
RCC4
6 < T(h) <= 72
T(h) > 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
Median
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


786-O
A498
RCC4
6 < T(h) <= 72
T(h) > 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
Mean
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


786-O
A498
RCC4
6 < T(h) <= 72
T(h) > 72
0 <= O2(%) <= 0.5
0.5 < O2(%) <= 1
GSVA
Cell-line
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

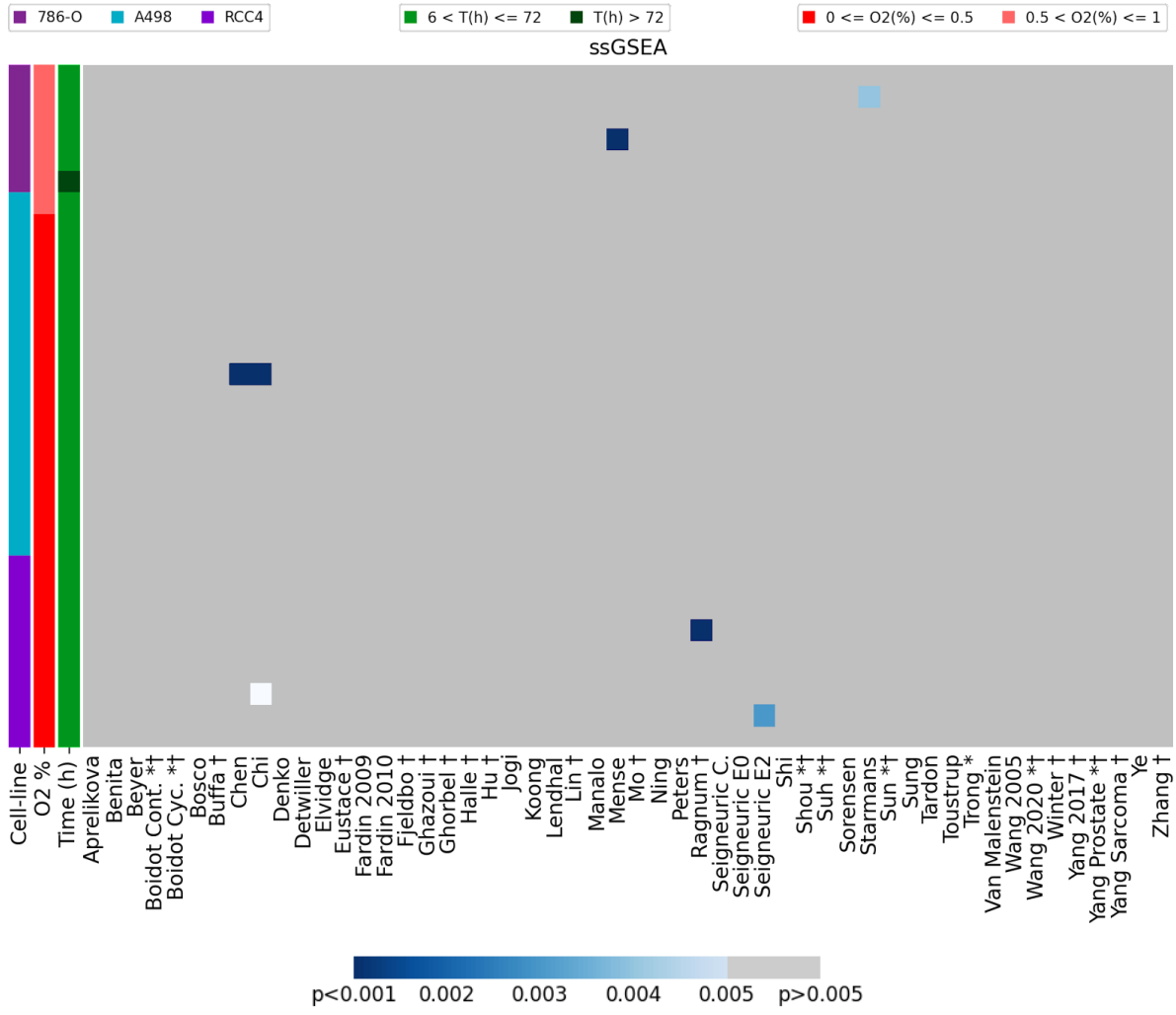


Figure 5.4: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in renal cancer*

Comparison of the scoring methods of the 53 hypoxia signatures across all renal cancer hypoxia experiments identified. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	40.62	12.50	62.50	0.00
Benita	28.12	18.75	46.88	0.00
Beyer	34.38	9.38	68.75	0.00
Boidot Cont. *†	34.38	6.25	31.25	0.00
Boidot Cyc. *†	9.38	0.00	0.00	0.00
Bosco	25.00	15.62	62.50	0.00
Buffa †	28.12	9.38	28.12	0.00
Chen	43.75	6.25	25.00	3.12
Chi	34.38	40.62	75.00	6.25
Denko	21.88	12.50	56.25	0.00
Detwiller	40.62	18.75	12.50	0.00
Elvidge	25.00	15.62	75.00	0.00
Eustace †	31.25	6.25	18.75	0.00
Fardin 2009	21.88	6.25	9.38	0.00
Fardin 2010	6.25	0.00	3.12	0.00
Fjeldbo †	28.12	3.12	6.25	0.00
Ghazoui †	9.38	15.62	25.00	0.00
Ghorbel †	40.62	34.38	31.25	0.00
Halle †	25.00	6.25	18.75	0.00
Hu †	18.75	31.25	15.62	0.00
Jogi	34.38	6.25	40.62	0.00
Koong	9.38	15.62	3.12	0.00
Lendhal	31.25	21.88	56.25	0.00
Lin †	25.00	31.25	15.62	0.00
Manalo	40.62	21.88	56.25	0.00
Mense	40.62	31.25	78.12	3.12
Mo †	25.00	12.50	3.12	0.00
Ning	12.50	0.00	18.75	0.00
Peters	12.50	3.12	25.00	0.00
Ragnum †	25.00	6.25	6.25	3.12
Seigneuric C.	15.62	0.00	0.00	0.00
Seigneuric E0	12.50	0.00	6.25	0.00
Seigneuric E2	21.88	0.00	3.12	3.12
Shi	6.25	0.00	3.12	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	12.50	3.12	0.00	0.00
Sorensen	34.38	12.50	46.88	0.00
Starmans	71.88	56.25	71.88	3.12
Sun *†	0.00	0.00	0.00	0.00
Sung	31.25	43.75	81.25	0.00
Tardon	50.00	25.00	59.38	0.00
Toustrup	34.38	9.38	34.38	0.00
Trong *	0.00	0.00	0.00	0.00
Van Malenstein	6.25	0.00	3.12	0.00
Wang 2005	28.12	6.25	43.75	0.00
Wang 2020 *†	9.38	9.38	6.25	0.00
Winter †	12.50	18.75	31.25	0.00
Yang 2017 †	28.12	0.00	12.50	0.00
Yang Prostate *†	18.75	0.00	12.50	0.00
Yang Sarcoma †	34.38	9.38	59.38	0.00
Ye	43.75	25.00	59.38	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	0.00	0.00	0.00

Table 5.4: *Percentage accuracy of determining hypoxic samples from normoxic samples in renal cancer for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

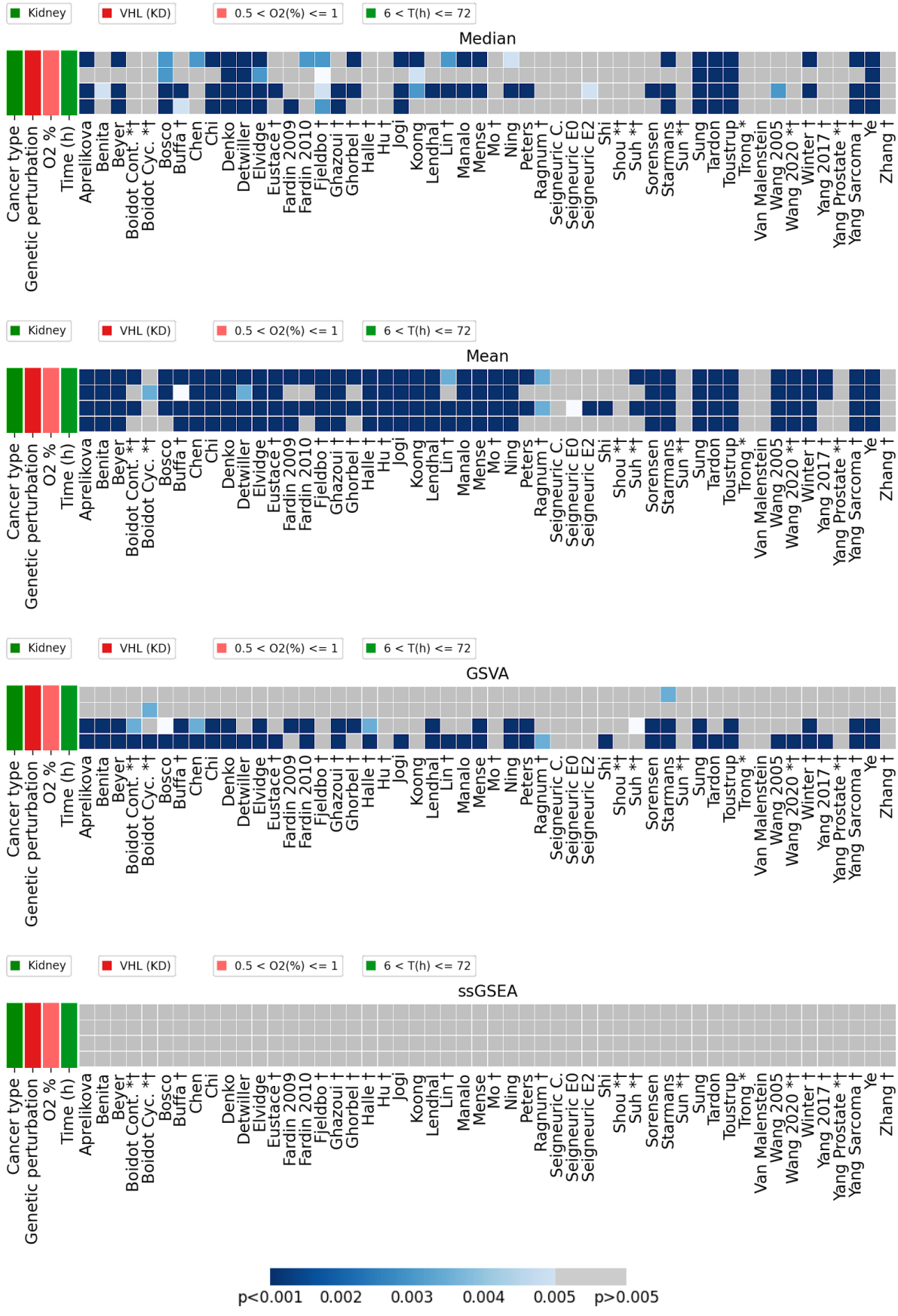

Kidney
VHL (KD)
0.5 < O2(%) <= 1
6 < T(h) <= 72
Median
Mean
GSVA
ssGSEA
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

Figure 5.5: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in a VHL-negative renal cancer cell line (786-O) and where VHL was subsequently reintroduced .*

Comparison of the scoring methods of the 53 hypoxia signatures across GSE65168. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	75.00	100.00	50.00	0.00
Benita	25.00	100.00	50.00	0.00
Beyer	75.00	100.00	50.00	0.00
Boidot Cont. *†	0.00	50.00	50.00	0.00
Boidot Cyc. *†	0.00	25.00	50.00	0.00
Bosco	100.00	75.00	50.00	0.00
Buffa †	50.00	100.00	50.00	0.00
Chen	25.00	100.00	50.00	0.00
Chi	75.00	100.00	50.00	0.00
Denko	100.00	100.00	50.00	0.00
Detwiller	100.00	100.00	25.00	0.00
Elvidge	100.00	100.00	50.00	0.00
Eustace †	25.00	100.00	25.00	0.00
Fardin 2009	25.00	75.00	25.00	0.00
Fardin 2010	25.00	50.00	50.00	0.00
Fjeldbo †	100.00	100.00	0.00	0.00
Ghazoui †	50.00	100.00	50.00	0.00
Ghorbel †	50.00	50.00	25.00	0.00
Halle †	0.00	100.00	50.00	0.00
Hu †	0.00	100.00	0.00	0.00
Jogi	75.00	100.00	25.00	0.00
Koong	75.00	100.00	0.00	0.00
Lendhal	25.00	100.00	50.00	0.00
Lin †	50.00	75.00	25.00	0.00
Manalo	50.00	100.00	25.00	0.00
Mense	50.00	100.00	50.00	0.00
Mo †	0.00	100.00	0.00	0.00
Ning	50.00	100.00	50.00	0.00
Peters	25.00	50.00	50.00	0.00
Ragnum †	0.00	50.00	25.00	0.00
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	0.00	25.00	0.00	0.00
Seigneuric E2	25.00	25.00	0.00	0.00
Shi	0.00	25.00	25.00	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	0.00	50.00	25.00	0.00
Sorensen	25.00	100.00	50.00	0.00
Starmans	75.00	100.00	75.00	0.00
Sun *†	0.00	0.00	0.00	0.00
Sung	100.00	100.00	50.00	0.00
Tardon	100.00	100.00	25.00	0.00
Toustrup	100.00	100.00	50.00	0.00
Trong *	0.00	0.00	0.00	0.00
Van Malenstein	0.00	0.00	0.00	0.00
Wang 2005	25.00	100.00	25.00	0.00
Wang 2020 *†	0.00	100.00	25.00	0.00
Winter †	50.00	100.00	50.00	0.00
Yang 2017 †	0.00	50.00	25.00	0.00
Yang Prostate *†	0.00	0.00	0.00	0.00
Yang Sarcoma †	75.00	100.00	50.00	0.00
Ye	100.00	100.00	50.00	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	25.00	0.00	0.00

Table 5.5: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in a VHL-negative renal cancer cell line (786-O) and where VHL was subsequently reintroduced .*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

5.4 Cell lines with HIF perturbation under hypoxia

Investigating the efficacy of hypoxia signatures in the condition of hypoxia-inducible factor knockout is a worthwhile endeavour to test dependence of hypoxia signatures on hypoxia inducible factors (e.g. HIF-1a). This is useful as some tumours may harbour gain/loss of function mutations in hypoxia-inducible factors. Several experiments involving perturbation of the HIF system were identified including *a)* HIF-1a knockdown in hypoxia vs normoxia (wt), *b)* HIF-1b knockout in hypoxia vs normoxia (wt), *c)* HIF-2a knockdown in hypoxia vs normoxia (wt) and *d)* HIF-1a + HIF-2a knockdown or knockout vs normoxia (wt) which were identified from GEO. These encompassed a range of cell lines: HIF-1a knockdown in hypoxia vs normoxia (501mel, LN229, MCF-7 and DAOY), HIF-1b knockout in hypoxia vs normoxia (MCF-7), HIF-2a knockdown in hypoxia vs normoxia (LN229, DAOY) and HIF-1a + HIF-2a knockdown or knockout in hypoxia vs normoxia (MCF-7, MDA-MB-231).

In the HIF-1a knockdown in hypoxia vs normoxia (37 pairwise combinations, 21 of which being the melanoma cell line 501mel, *Table 5.6*) it is possible to observe a drop in accuracy in the majority of signatures. However, some signatures still perform well (*Fig 5.6*), for instance Ye using the median score despite the different cell lines, oxygen tensions and HIF-1a knockdown still has an impressive 91.89% accuracy (*Table 5.7*). Ye was developed from 31 breast cancer cell lines under 1% oxygen for 24 hours and the 42-gene signature was not developed under the conditions of HIF-1a knockdown. Aprelikova, using a GSVA score, achieved ~88% accuracy in identifying hypoxia in all types in hypoxia vs normoxia (Chapter 4.12) performed with an accuracy of just 32.43% in HIF-1a knockdown.

In the HIF-1b knockout vs normoxia experiment (MCF-7 cells, 72 pairs, *Table 5.8*) the median score performs well in the majority of signatures (*Fig. 5.7*). Indeed, 31 signatures achieved 100% accuracy and the signatures that performed best in breast cancer (median scores in Benita, Sorensen and Yang Sarcoma) and the overall best signature (Aprelikova

GSVA) all classify with 100% accuracy (*Table 5.9*). Interestingly, all other scores performed quite poorly overall with some notable exceptions, for instance Lin had 100% accuracy across two scores - median and mean.

In HIF-2a knockdown in hypoxia vs normoxia in one glioblastoma cell line and one medulloblastoma cell line (LN229 and DAOY, *Table 5.10*) made little difference to signature performance (*Fig. 5.8*). The best performing signatures in central nervous system derived cell lines: Benita, Beyer, Chi, Denko, Fjeldbo, Mense, Sung, Tardon and Yang Sarcoma (mean) and Sung (median) all identified hypoxic samples in the HIF-2a knockdown in hypoxia vs normoxia with 100% accuracy, aside from Mense that reached 90% using the mean score. Also Aprelikova GSVA reached 100% (*Table 5.11*). These results suggest HIF-2a knockdown makes minimal difference to effective hypoxia signature performance. Therefore, one may expect any effect seen in HIF-1a + HIF-2a knockdown/knockout experiments in hypoxia vs normoxia down to the knockdown of HIF-1a.

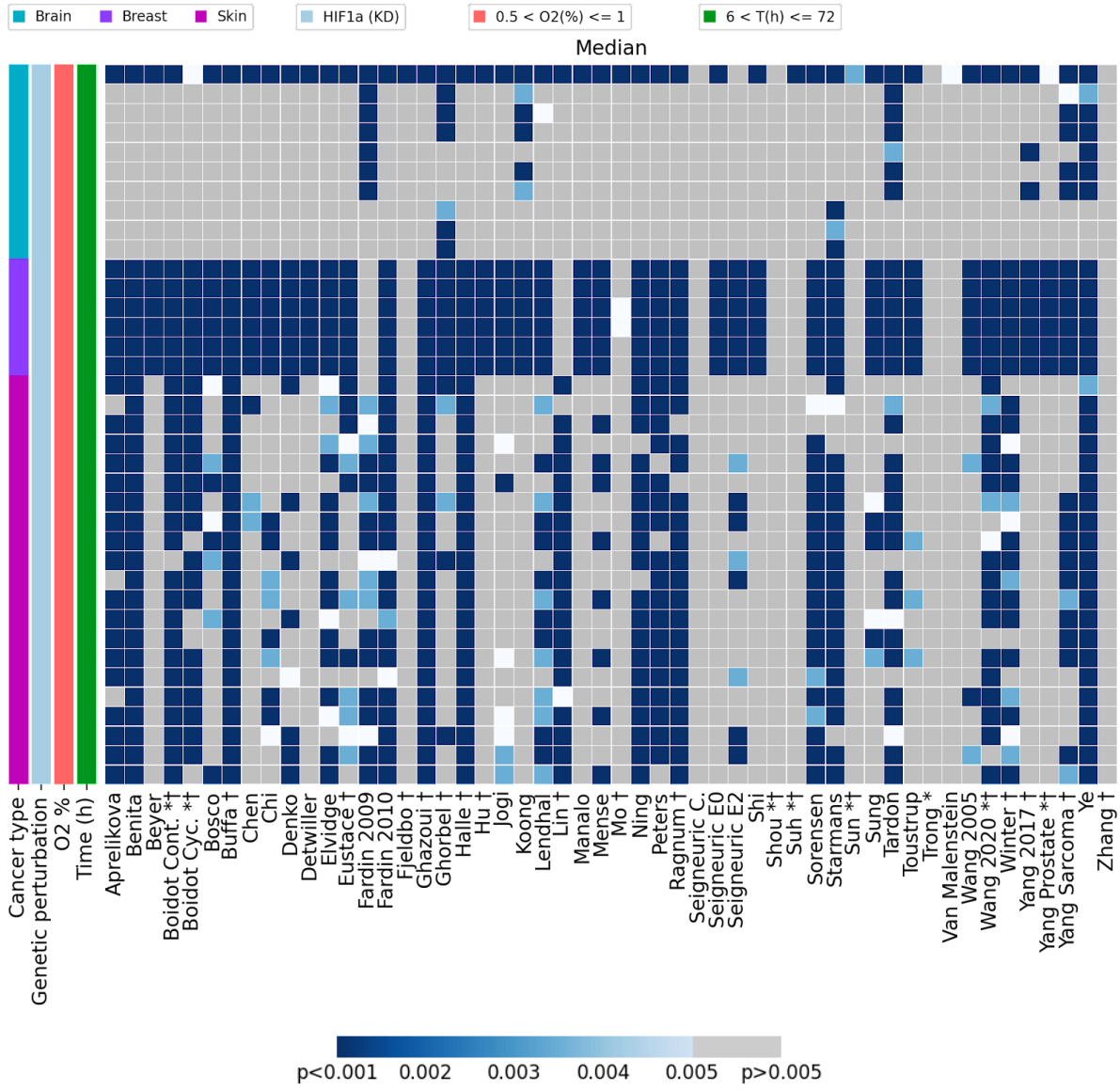
I went on to investigate the performance of the 53 hypoxia signatures in two datasets from breast cancer cell lines (MCF-7, GSE3188, MDA-MB-231, GSE108833) with HIF-1a + HIF-2a knockdown or knockouted (*Table 5.12*). In GSE108833, HIF-1a + HIF-2a was knockouted in MDA-MB-231 cells using the CRISPR/Cas9 technique[155] (ER-, PR-, HER2- cell line, four pairwise combinations) and in GSE3188, HIF-1a + HIF-2a was knockdown in MCF-7 cells using short interfering RNA-based suppression[59] (ER+, PR+, HER2-, six pairwise combinations). ssGSEA and GVSA perform poorly and median appears the most effective overall score (*Fig. 5.9*). Six signatures spread across median and mean reached 100% accuracy (median: Mense, Starmans and Yang Prostate, mean: Denko, Peters, Starmans, and Wang 2015). Ye, the most effective signature in solely HIF-1a knockdown reached only 60% in median and mean. The best performing signatures in breast cancer, Benita, Sorensen and Yang Sarcoma using the median score, decrease in performance under this condition, with 60% accuracy using the median score achieved. The overall best signature across cell lines (Aprelikova GSVA) identified in Chapter 4.12, only reaches 10%

(*Table 5.13*). Looking at the median and mean scores, most signatures appear to work in the knockdown experiments but not in the knockout experiments. This may suggest only a small amount of HIF-1a/HIF-2a is needed to be effective and/or perhaps the knockdown was not particularly effective, although the western blots in the paper appear to suggest it worked well.

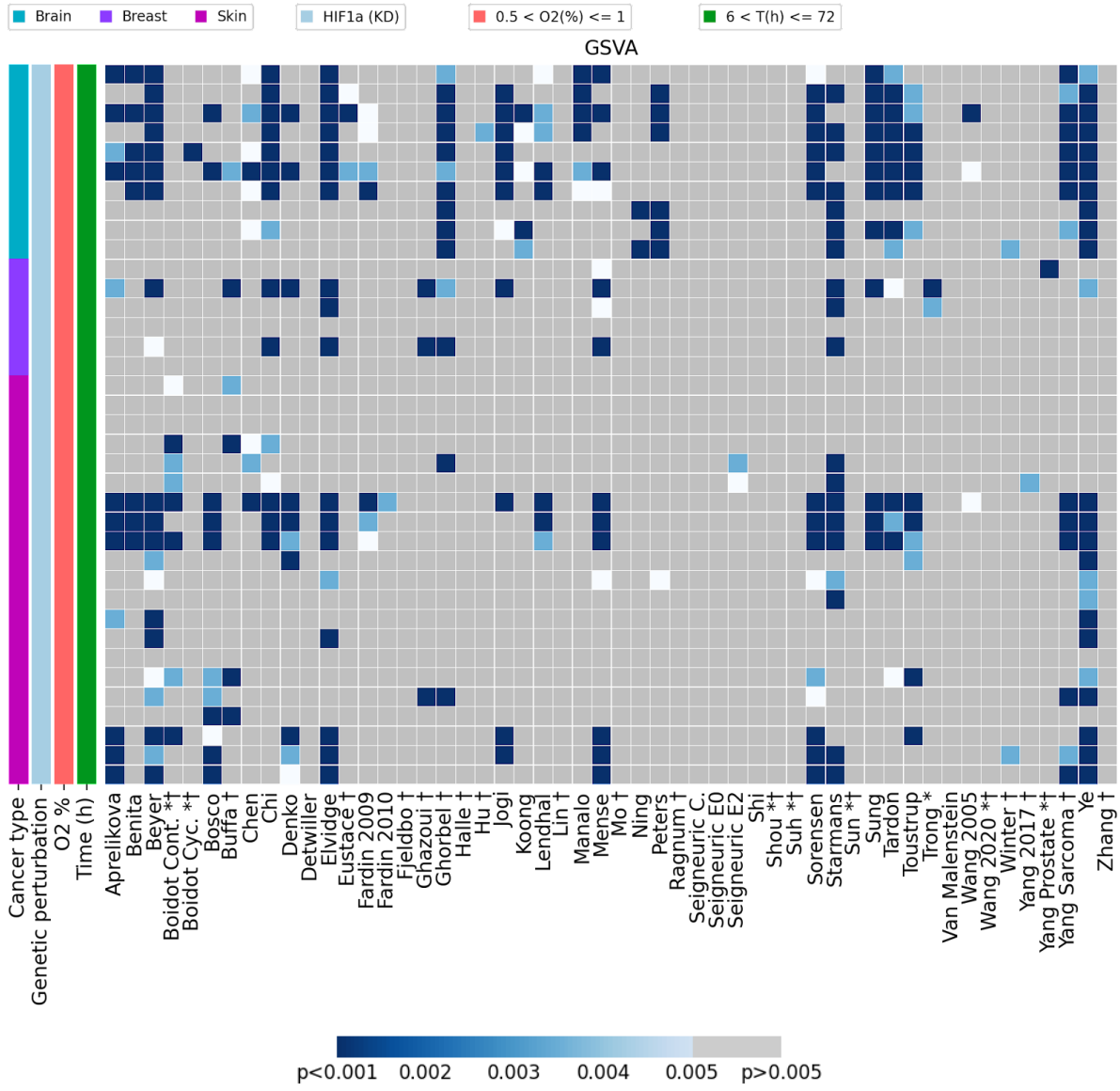
Cell line/samples	Number of Hypoxic HIF1a KD samples	Number of Normoxic samples	Pairs
501mel	7	3	21
LN229	3	3	9
MCF-7	3	2	6
DAOY	1	1	1
Total	14	9	37

Table 5.6: *HIF1a knocked down cell lines used in hypoxia experiments in GEO*

Details of the hypoxia HIF1a KD versus normoxia experiments in GEO


Brain
Breast
Skin
HIF1a (KD)
0.5 < O2(%) <= 1
6 < T(h) <= 72
Median
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


Brain
Breast
Skin
HIF1a (KD)
0.5 < O2(%) <= 1
6 < T(h) <= 72
Mean
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


Brain
Breast
Skin
HIF1a (KD)
0.5 < O2(%) <= 1
6 < T(h) <= 72
GSVA
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge †
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

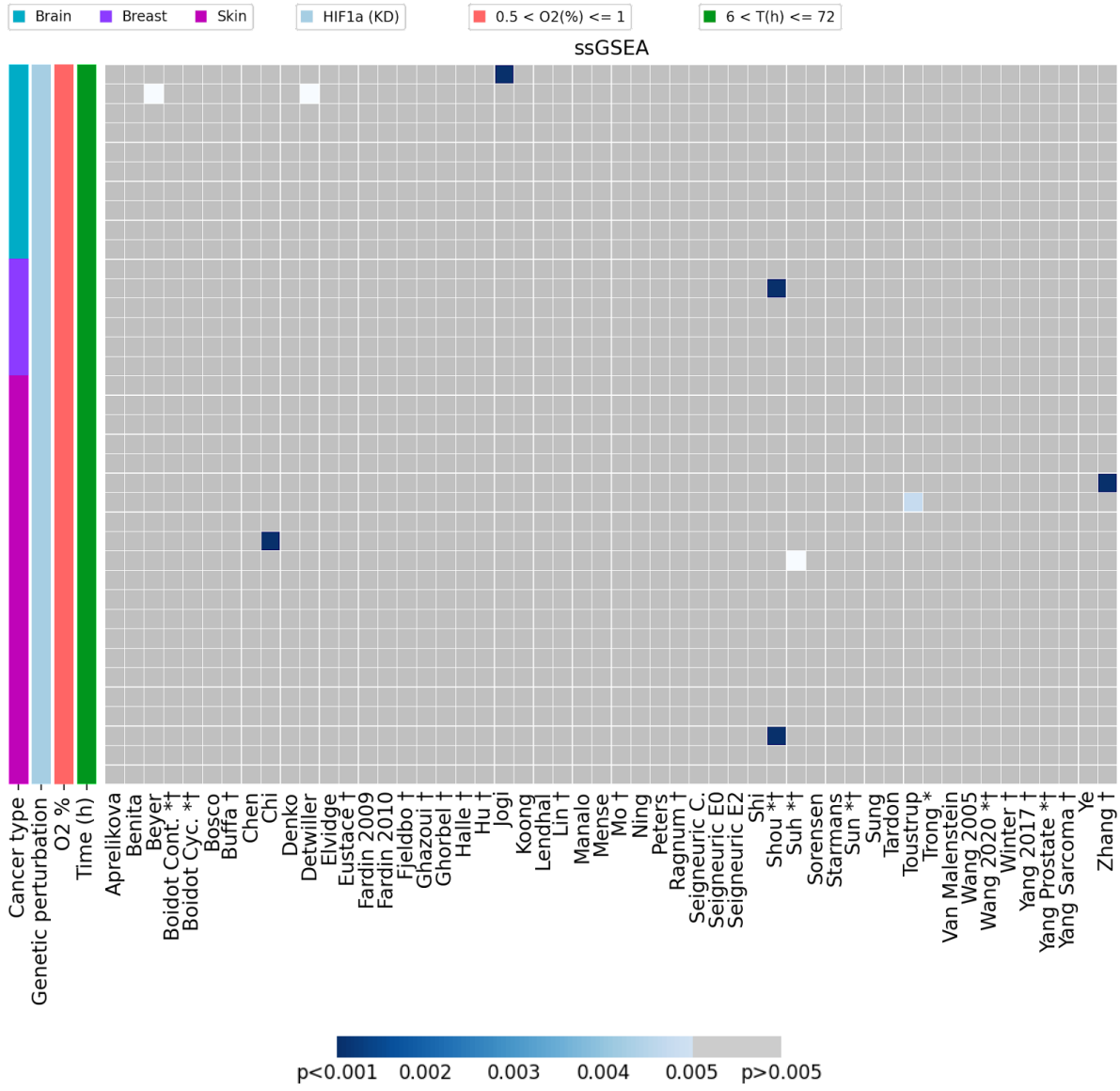


Figure 5.6: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in HIF1a knocked down cancer cell lines exposed to hypoxia*

Comparison of the scoring methods of the 53 hypoxia signatures across four HIF1a knocked down cell lines exposed to hypoxia. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	67.57	18.92	32.43	0.00
Benita	75.68	24.32	21.62	0.00
Beyer	18.92	27.03	56.76	2.70
Boidot Cont. *†	72.97	16.22	21.62	0.00
Boidot Cyc. *†	64.86	16.22	2.70	0.00
Bosco	40.54	18.92	29.73	0.00
Buffa †	75.68	24.32	16.22	0.00
Chen	27.03	32.43	24.32	0.00
Chi	43.24	21.62	40.54	2.70
Denko	40.54	18.92	27.03	0.00
Detwiller	18.92	18.92	0.00	2.70
Elvidge	62.16	21.62	48.65	0.00
Eustace †	51.35	18.92	8.11	0.00
Fardin 2009	67.57	18.92	18.92	0.00
Fardin 2010	75.68	16.22	2.70	0.00
Fjeldbo †	2.70	0.00	0.00	0.00
Ghazoui †	75.68	18.92	8.11	0.00
Ghorbel †	48.65	40.54	37.84	0.00
Halle †	75.68	18.92	0.00	0.00
Hu †	18.92	24.32	2.70	0.00
Jogi	37.84	18.92	29.73	2.70
Koong	32.43	27.03	13.51	0.00
Lendhal	62.16	18.92	21.62	0.00
Lin †	56.76	29.73	0.00	0.00
Manalo	18.92	2.70	16.22	0.00
Mense	40.54	18.92	40.54	0.00
Mo †	8.11	2.70	0.00	0.00
Ning	70.27	18.92	5.41	0.00
Peters	67.57	35.14	18.92	0.00
Ragnum †	70.27	24.32	0.00	0.00
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	18.92	0.00	0.00	0.00
Seigneuric E2	37.84	29.73	5.41	0.00
Shi	18.92	0.00	0.00	0.00
Shou *†	0.00	0.00	0.00	5.41
Suh *†	2.70	2.70	0.00	2.70
Sorensen	70.27	35.14	43.24	0.00
Starmans	75.68	35.14	51.35	0.00
Sun *†	2.70	0.00	0.00	0.00
Sung	35.14	29.73	32.43	0.00
Tardon	81.08	18.92	37.84	0.00
Toustrup	27.03	35.14	35.14	2.70
Trong *	0.00	0.00	5.41	0.00
Van Malenstein	2.70	0.00	0.00	0.00
Wang 2005	27.03	21.62	8.11	0.00
Wang 2020 *†	67.57	0.00	0.00	0.00
Winter †	67.57	18.92	5.41	0.00
Yang 2017 †	24.32	2.70	2.70	0.00
Yang Prostate *†	18.92	0.00	2.70	0.00
Yang Sarcoma †	62.16	18.92	37.84	0.00
Ye	91.89	35.14	64.86	0.00
Zhang †	0.00	0.00	0.00	2.70
Zou *†	18.92	0.00	0.00	0.00

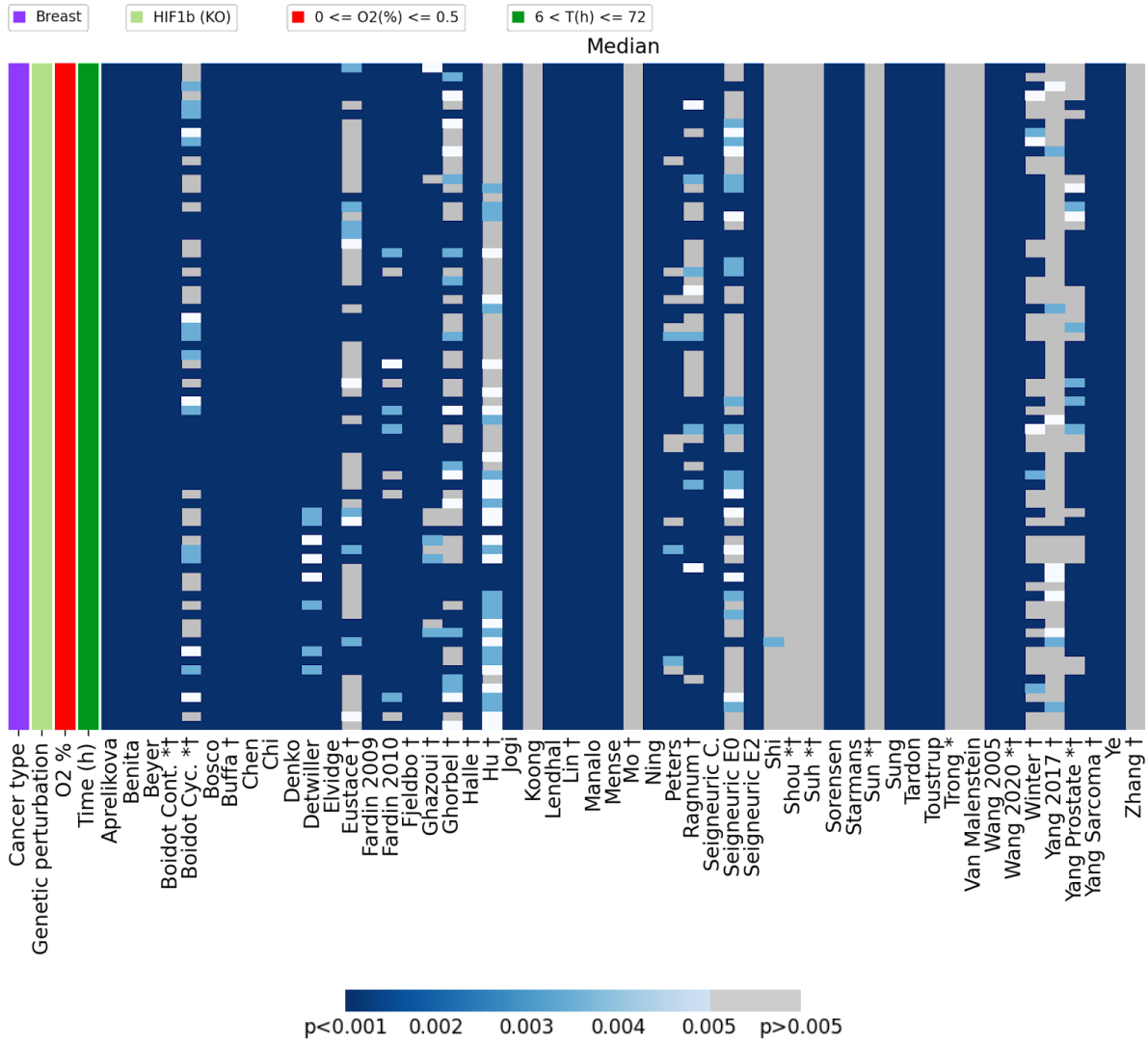
Table 5.7: *Percentage accuracy of determining hypoxic samples from normoxic samples in HIF1a knocked down cell lines across four hypoxia scores*

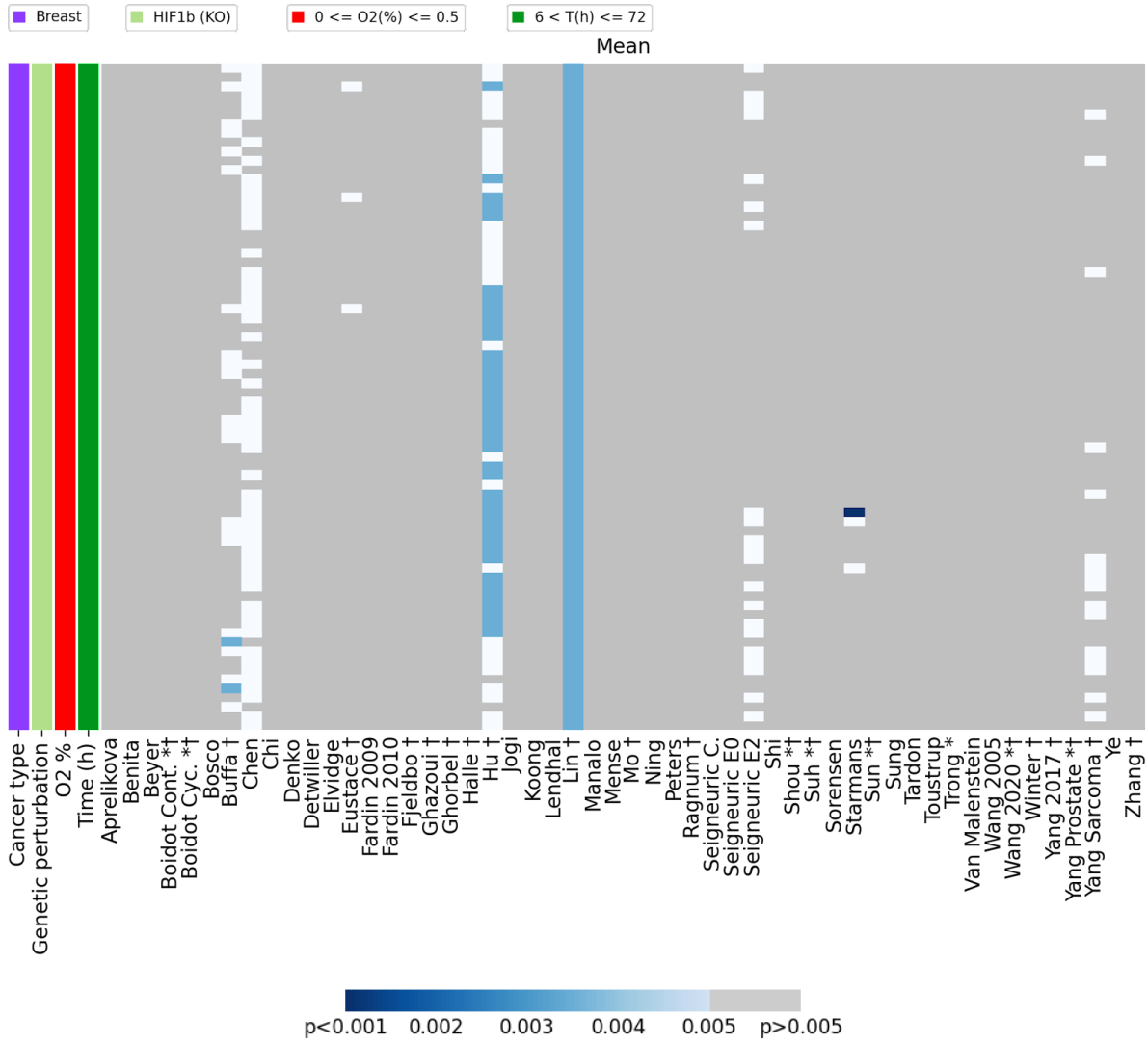
The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

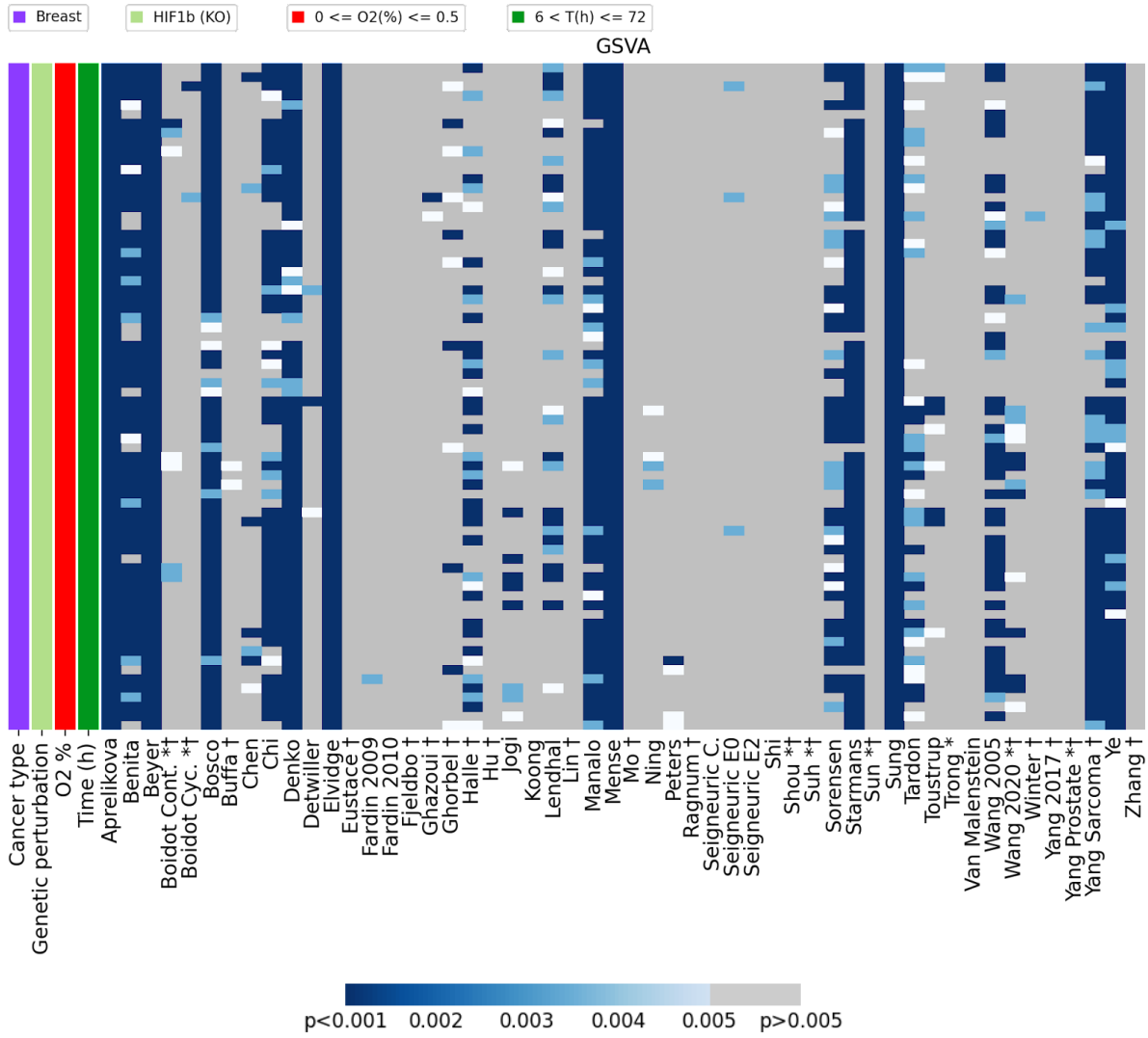
Cell line/samples	Number of Hypoxic samples with HIF1b KO	Number of Normoxic samples	Pairs
MCF-7	6	12	72

Table 5.8: *HIF1b knocked out cell lines used in hypoxia experiments in GEO*

Details of the hypoxia HIF1b knocked out versus normoxia experiments in GEO


Breast
HIF1b (KO)
0 <= O2(%) <= 0.5
6 < T(h) <= 72
Median
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


Breast
HIF1b (KO)
0 <= O2(%) <= 0.5
6 < T(h) <= 72
Mean
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


Breast
HIF1b (KO)
0 <= O2(%) <= 0.5
6 < T(h) <= 72
GSVA
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

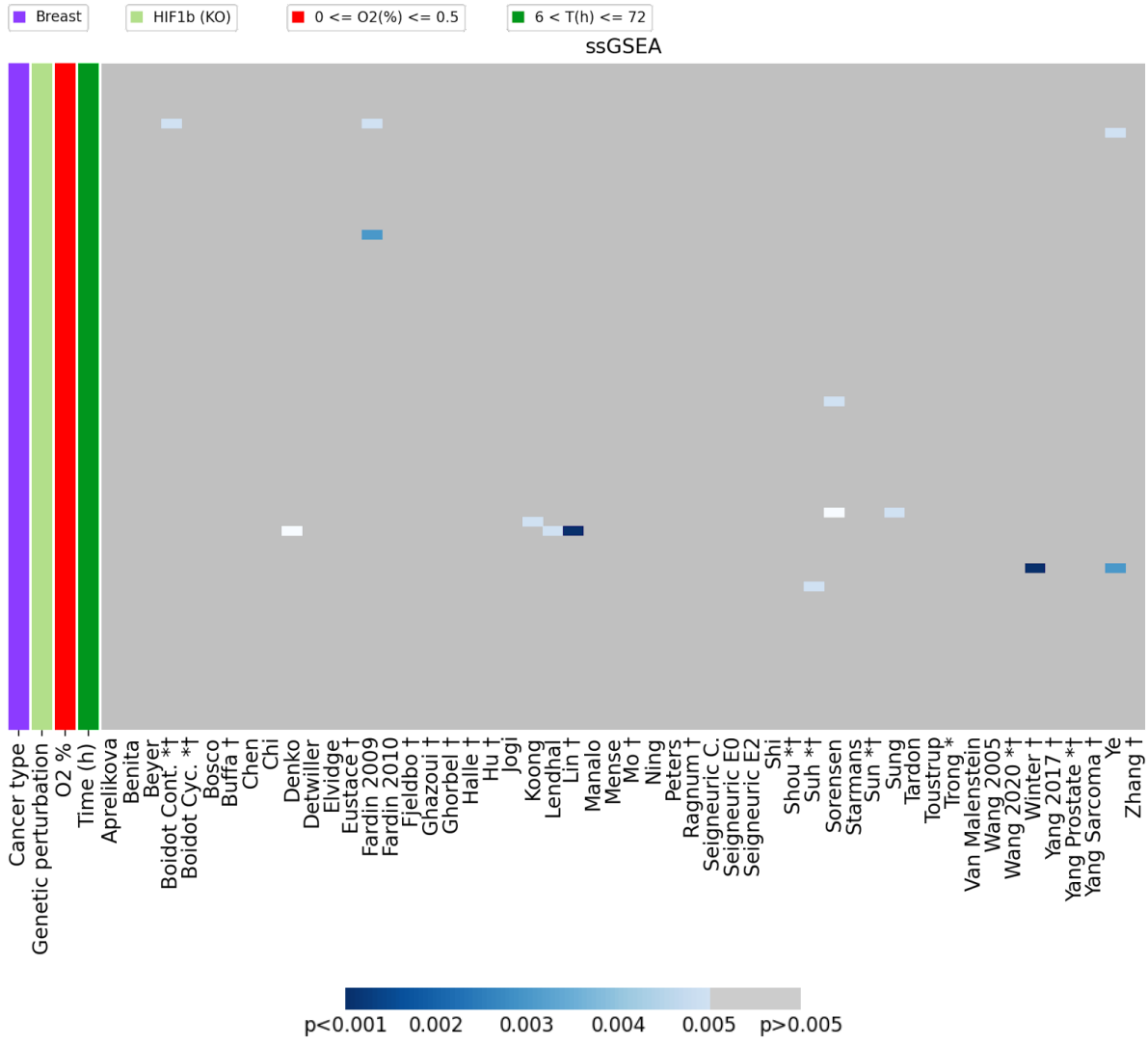


Figure 5.7: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in HIF1b knocked out cancer cell lines exposed to hypoxia*

Comparison of the scoring methods of the 53 hypoxia signatures on HIF1b knocked out cell lines exposed to hypoxia. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	100.00	0.00	100.00	0.00
Benita	100.00	0.00	86.11	0.00
Beyer	100.00	0.00	100.00	0.00
Boidot Cont. *†	100.00	0.00	9.72	1.39
Boidot Cyc. *†	66.67	0.00	2.78	0.00
Bosco	100.00	0.00	97.22	0.00
Buffa †	100.00	30.56	2.78	0.00
Chen	100.00	75.00	9.72	0.00
Chi	100.00	0.00	75.00	0.00
Denko	100.00	0.00	97.22	1.39
Detwiller	100.00	0.00	4.17	0.00
Elvidge	100.00	0.00	100.00	0.00
Eustace †	48.61	4.17	0.00	0.00
Fardin 2009	100.00	0.00	1.39	2.78
Fardin 2010	93.06	0.00	0.00	0.00
Fjeldbo †	100.00	0.00	0.00	0.00
Ghazoui †	93.06	0.00	2.78	0.00
Ghorbel †	61.11	0.00	15.28	0.00
Halle †	100.00	0.00	52.78	0.00
Hu †	52.78	95.83	0.00	0.00
Jogi	100.00	0.00	12.50	0.00
Koong	0.00	0.00	0.00	1.39
Lendhal	100.00	0.00	41.67	1.39
Lin †	100.00	100.00	0.00	1.39
Manalo	100.00	0.00	88.89	0.00
Mense	100.00	0.00	100.00	0.00
Mo †	0.00	0.00	0.00	0.00
Ning	100.00	0.00	5.56	0.00
Peters	88.89	0.00	5.56	0.00
Ragnum †	73.61	0.00	0.00	0.00
Seigneuric C.	100.00	0.00	0.00	0.00
Seigneuric E0	51.39	0.00	4.17	0.00
Seigneuric E2	100.00	29.17	0.00	0.00
Shi	1.39	0.00	0.00	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	0.00	0.00	0.00	1.39
Sorensen	100.00	0.00	59.72	2.78
Starmans	100.00	4.17	93.06	0.00
Sun *†	0.00	0.00	0.00	0.00
Sung	100.00	0.00	100.00	1.39
Tardon	100.00	0.00	50.00	0.00
Toustrup	100.00	0.00	12.50	0.00
Trong *	0.00	0.00	0.00	0.00
Van Malenstein	0.00	0.00	0.00	0.00
Wang 2005	100.00	0.00	65.28	0.00
Wang 2020 *†	100.00	0.00	19.44	0.00
Winter †	52.78	0.00	1.39	1.39
Yang 2017 †	15.28	0.00	0.00	0.00
Yang Prostate *†	72.22	0.00	0.00	0.00
Yang Sarcoma †	100.00	22.22	83.33	0.00
Ye	100.00	0.00	97.22	2.78
Zhang †	0.00	0.00	0.00	0.00
Zou *†	58.33	0.00	0.00	0.00

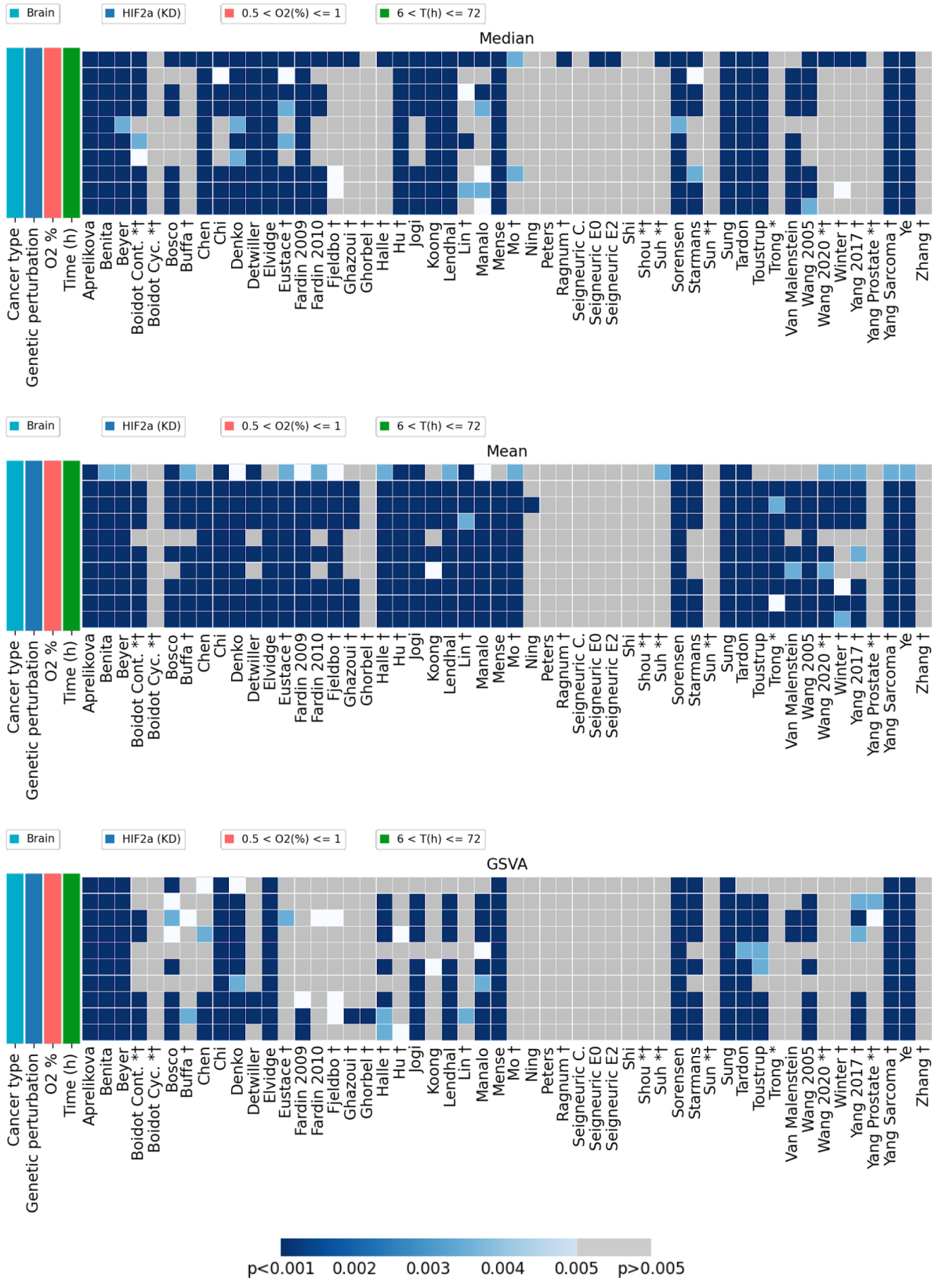
Table 5.9: *Percentage accuracy of determining hypoxic samples from normoxic samples in HIF1b knocked out cell lines across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

Cell line/samples	Number of Normoxic samples	Number of Hypoxic HIF2a KD samples	Pairs
LN229	3	3	9
DAOY	1	1	1
Total	4	4	10

Table 5.10: *HIF2a knocked down cell lines used in hypoxia experiments in GEO*

Details of the hypoxia HIF2a knocked down versus normoxia experiments in GEO


Brain
HIF2a (KD)
0.5 < O2(%) <= 1
6 < T(h) <= 72
Median
Mean
GSVA
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

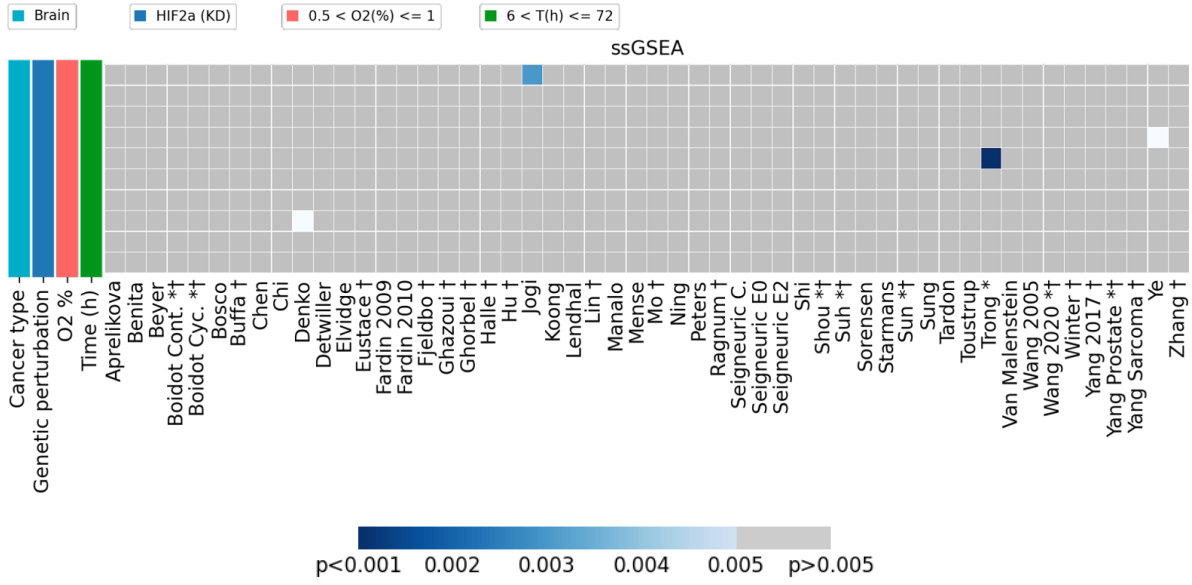


Figure 5.8: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in HIF2a knocked down cancer cell lines exposed to hypoxia*

Comparison of the scoring methods of the 53 hypoxia signatures on HIF2a knocked out cell lines exposed to hypoxia. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	100.00	100.00	100.00	0.00
Benita	100.00	100.00	100.00	0.00
Beyer	100.00	100.00	100.00	0.00
Boidot Cont. *†	90.00	80.00	50.00	0.00
Boidot Cyc. *†	0.00	0.00	0.00	0.00
Bosco	60.00	80.00	80.00	0.00
Buffa †	10.00	90.00	20.00	0.00
Chen	100.00	90.00	50.00	0.00
Chi	70.00	100.00	100.00	0.00
Denko	100.00	100.00	100.00	10.00
Detwiller	100.00	80.00	20.00	0.00
Elvidge	100.00	90.00	100.00	0.00
Eustace †	80.00	100.00	10.00	0.00
Fardin 2009	100.00	100.00	30.00	0.00
Fardin 2010	60.00	80.00	10.00	0.00
Fjeldbo †	30.00	100.00	30.00	0.00
Ghazoui †	10.00	60.00	10.00	0.00
Ghorbel †	0.00	0.00	10.00	0.00
Halle †	10.00	100.00	70.00	0.00
Hu †	100.00	100.00	20.00	0.00
Jogi	70.00	100.00	80.00	10.00
Koong	100.00	80.00	10.00	0.00
Lendhal	100.00	100.00	80.00	0.00
Lin †	40.00	100.00	10.00	0.00
Manalo	60.00	100.00	90.00	0.00
Mense	100.00	90.00	100.00	0.00
Mo †	20.00	100.00	0.00	0.00
Ning	0.00	10.00	0.00	0.00
Peters	0.00	0.00	0.00	0.00
Ragnum †	10.00	0.00	0.00	0.00
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	10.00	0.00	0.00	0.00
Seigneuric E2	10.00	0.00	0.00	0.00
Shi	0.00	0.00	0.00	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	10.00	10.00	0.00	0.00
Sorensen	100.00	100.00	100.00	0.00
Starmans	70.00	70.00	80.00	0.00
Sun *†	0.00	0.00	0.00	0.00
Sung	100.00	100.00	100.00	0.00
Tardon	100.00	100.00	90.00	0.00
Toustrup	100.00	90.00	80.00	0.00
Trong *	0.00	90.00	0.00	10.00
Van Malenstein	80.00	80.00	20.00	0.00
Wang 2005	70.00	90.00	70.00	0.00
Wang 2020 *†	10.00	90.00	0.00	0.00
Winter †	20.00	70.00	0.00	0.00
Yang 2017 †	10.00	80.00	60.00	0.00
Yang Prostate *†	0.00	0.00	20.00	0.00
Yang Sarcoma †	100.00	100.00	100.00	0.00
Ye	100.00	100.00	100.00	10.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	10.00	10.00	0.00	0.00

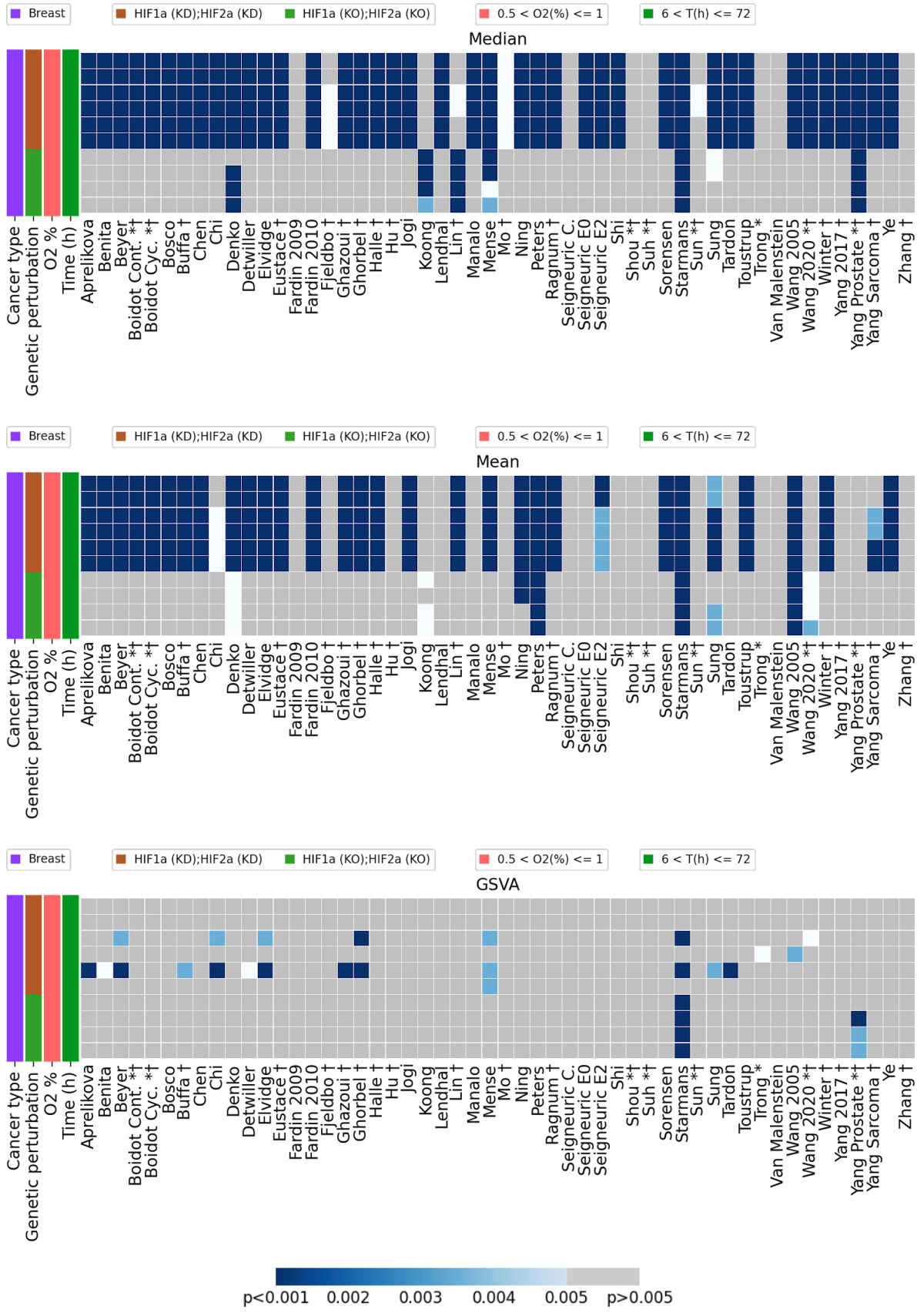
Table 5.11: *Percentage accuracy of determining hypoxic samples from normoxic samples in HIF2a knocked down cell lines across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

Cell line/samples	Number of Hypoxic samples with HIF1a and HIF2a KO	Number of Hypoxic samples with HIF1a and HIF2a KD	Number of Normoxic samples	Pairs
MCF-7	0	3	2	6
MDA-MB-231	2	0	2	4
Total	2	3	4	10

Table 5.12: *HIF1a and HIF2a KO/KD cell lines used in hypoxia experiments in GEO*

Details of the hypoxia HIF1a and HIF2a KO/KD versus normoxia experiments in GEO


Breast
HIF1a (KD);HIF2a (KD)
HIF1a (KO);HIF2a (KO)
0.5 < O2(%) <= 1
6 < T(h) <= 72
Median
Mean
GSVA
Cancer type
Genetic perturbation
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

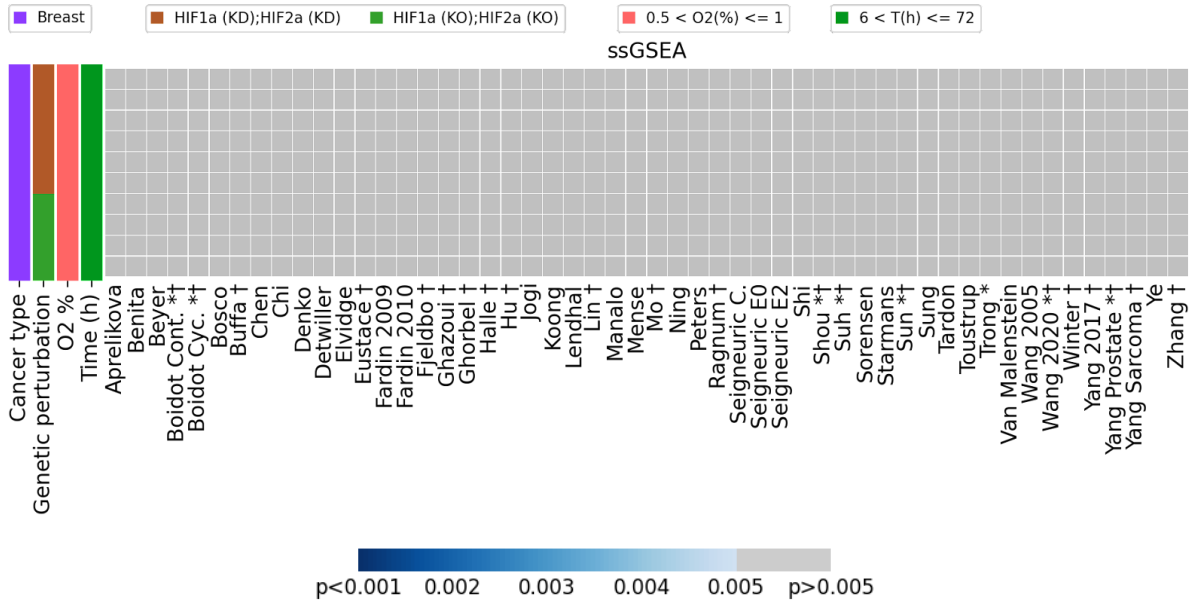


Figure 5.9: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in HIF1a and HIF2a KO/KD cancer cell lines exposed to hypoxia*

Comparison of the scoring methods of the 53 hypoxia signatures on HIF1a and HIF2a KO/KD cell lines exposed to hypoxia. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	60.00	60.00	10.00	0.00
Benita	60.00	60.00	10.00	0.00
Beyer	60.00	60.00	20.00	0.00
Boidot Cont. *†	60.00	60.00	0.00	0.00
Boidot Cyc. *†	60.00	60.00	0.00	0.00
Bosco	60.00	60.00	0.00	0.00
Buffa †	60.00	60.00	10.00	0.00
Chen	60.00	60.00	0.00	0.00
Chi	60.00	40.00	20.00	0.00
Denko	90.00	100.00	0.00	0.00
Detwiller	60.00	60.00	10.00	0.00
Elvidge	60.00	60.00	20.00	0.00
Eustace †	60.00	60.00	0.00	0.00
Fardin 2009	0.00	0.00	0.00	0.00
Fardin 2010	60.00	60.00	0.00	0.00
Fjeldbo †	40.00	0.00	0.00	0.00
Ghazoui †	60.00	60.00	10.00	0.00
Ghorbel †	60.00	60.00	20.00	0.00
Halle †	60.00	60.00	0.00	0.00
Hu †	60.00	0.00	0.00	0.00
Jogi	60.00	60.00	0.00	0.00
Koong	40.00	30.00	0.00	0.00
Lendhal	60.00	0.00	0.00	0.00
Lin †	60.00	60.00	0.00	0.00
Manalo	60.00	0.00	0.00	0.00
Mense	100.00	60.00	30.00	0.00
Mo †	60.00	0.00	0.00	0.00
Ning	60.00	80.00	0.00	0.00
Peters	60.00	100.00	0.00	0.00
Ragnum †	60.00	60.00	0.00	0.00
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	60.00	0.00	0.00	0.00
Seigneuric E2	60.00	60.00	0.00	0.00
Shi	60.00	0.00	0.00	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	0.00	0.00	0.00	0.00
Sorensen	60.00	60.00	0.00	0.00
Starmans	100.00	100.00	60.00	0.00
Sun *†	20.00	0.00	0.00	0.00
Sung	80.00	80.00	10.00	0.00
Tardon	60.00	0.00	10.00	0.00
Toustrup	60.00	60.00	0.00	0.00
Trong *	0.00	0.00	10.00	0.00
Van Malenstein	0.00	0.00	0.00	0.00
Wang 2005	60.00	100.00	10.00	0.00
Wang 2020 *†	60.00	40.00	10.00	0.00
Winter †	60.00	60.00	0.00	0.00
Yang 2017 †	60.00	0.00	0.00	0.00
Yang Prostate *†	100.00	0.00	30.00	0.00
Yang Sarcoma †	60.00	40.00	0.00	0.00
Ye	60.00	60.00	0.00	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	40.00	0.00	0.00	0.00

Table 5.13: *Percentage accuracy of determining hypoxic samples from normoxic samples in HIF1a and HIF2a knocked down cell lines across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

5.5 Assessing the effectiveness of mimicked hypoxia

Three cancer cell lines (MCF-7, MB231RN-LM and PANC-1) which had gene expression data in mimicked hypoxia and normoxia were identified from the Gene Expression Omnibus (as per methods). Mimicked hypoxia was achieved using either Cobalt(II) chloride ($CoCL_2$, a chemical inducer of HIF-1a[156] used in GSE45362 & GSE82104) or Dimethyloxalylglycine, N-(Methoxyoxoacetyl)-glycine methyl ester (DMOG, a competitive prolylhydroxylase inhibitor which increases HIF-1a activation[157] used in GSE1388 and GSE85353).

I then assessed the performance of the 53 signatures at defining hypoxia in these three cell lines. Again, ssGSEA appeared the worst performing score. Median score appeared the most effective (*Fig. 5.10*, top panel). Sixteen signatures achieved 100% accuracy in identifying mimicked hypoxia using the median: Aprelikova, Benita, Boidot Continous, Chen, Chi, Elvidge, Halle, Lin, Mense, Sorensen, Starmans, Sung, Tardon, Toustrup, Yang Sarcoma and Ye (*Fig. 5.10*). Indeed, the best performing breast and pancreatic cancer signatures Benita, Sorensen and Yang sarcoma showed 100% accuracy identifying mimicked hypoxia using the median score. This suggests these signatures depend on HIF-1a activation and can be used to identify cells under mimicked hypoxia and perhaps that the mimicked hypoxic state reflects, at least partially, the gene expression profiles seen in cells in a hypoxic chamber (at least in these two cancer types). Several signatures (eight) using GVSA also reached 100% accuracy: Aprelikova, Benita, Beyer, Chi, Elvidge, Sorensen, Sung and Ye. The worst performing signatures were familiar: Shou and Seigneuric common achieved 0% accuracy on all scores (*Table 5.15*).

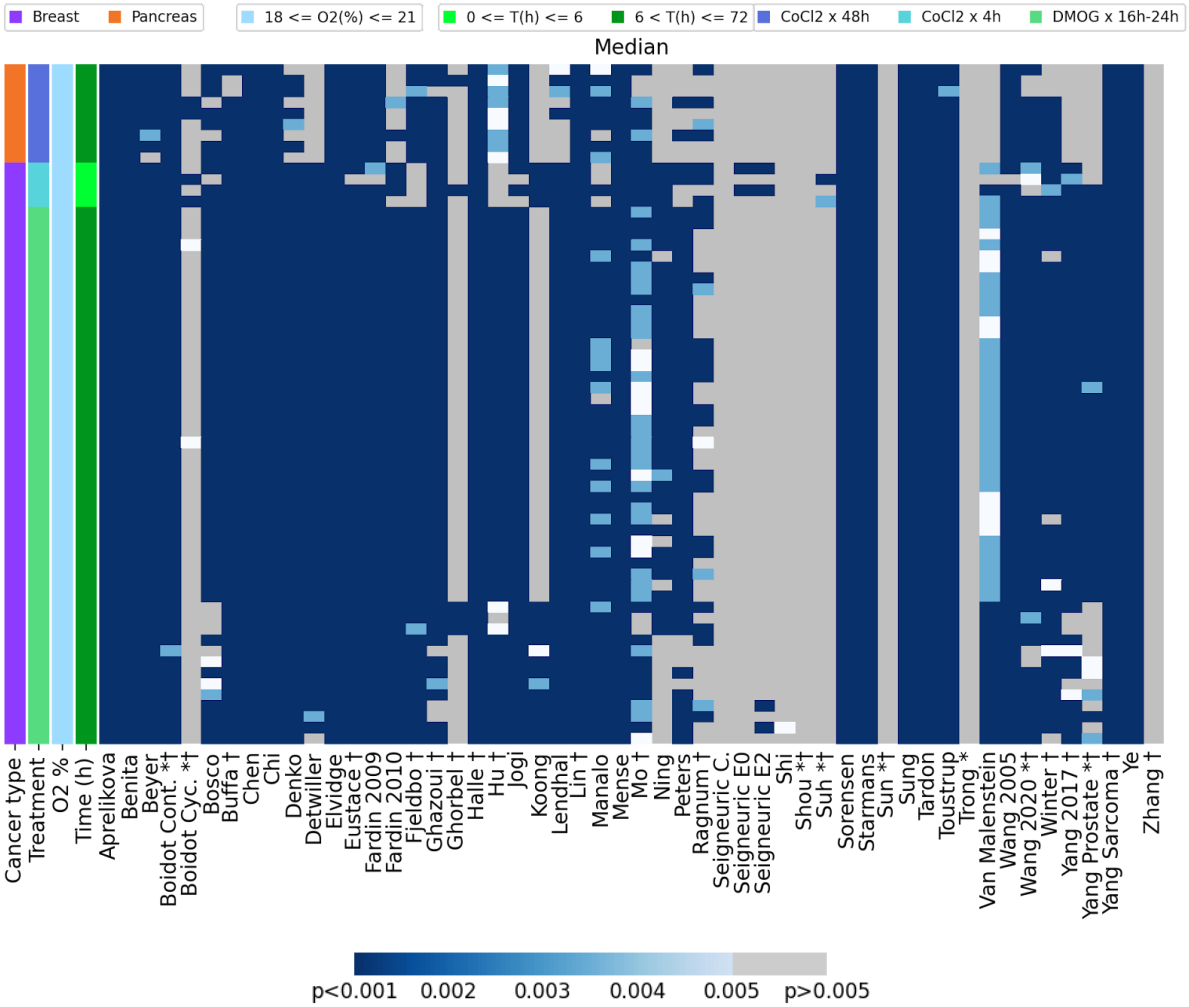
I also had the opportunity to investigate which hypoxia signatures, if any, may differentiate between mimicked and true hypoxia in breast cancer cells (MCF-7, five normoxic samples, five hypoxia samples and 13 pairwise combinations). Interestingly, Benita, Ghazoui, Lendhal and Ye using the mean score had the ability to differentiate true

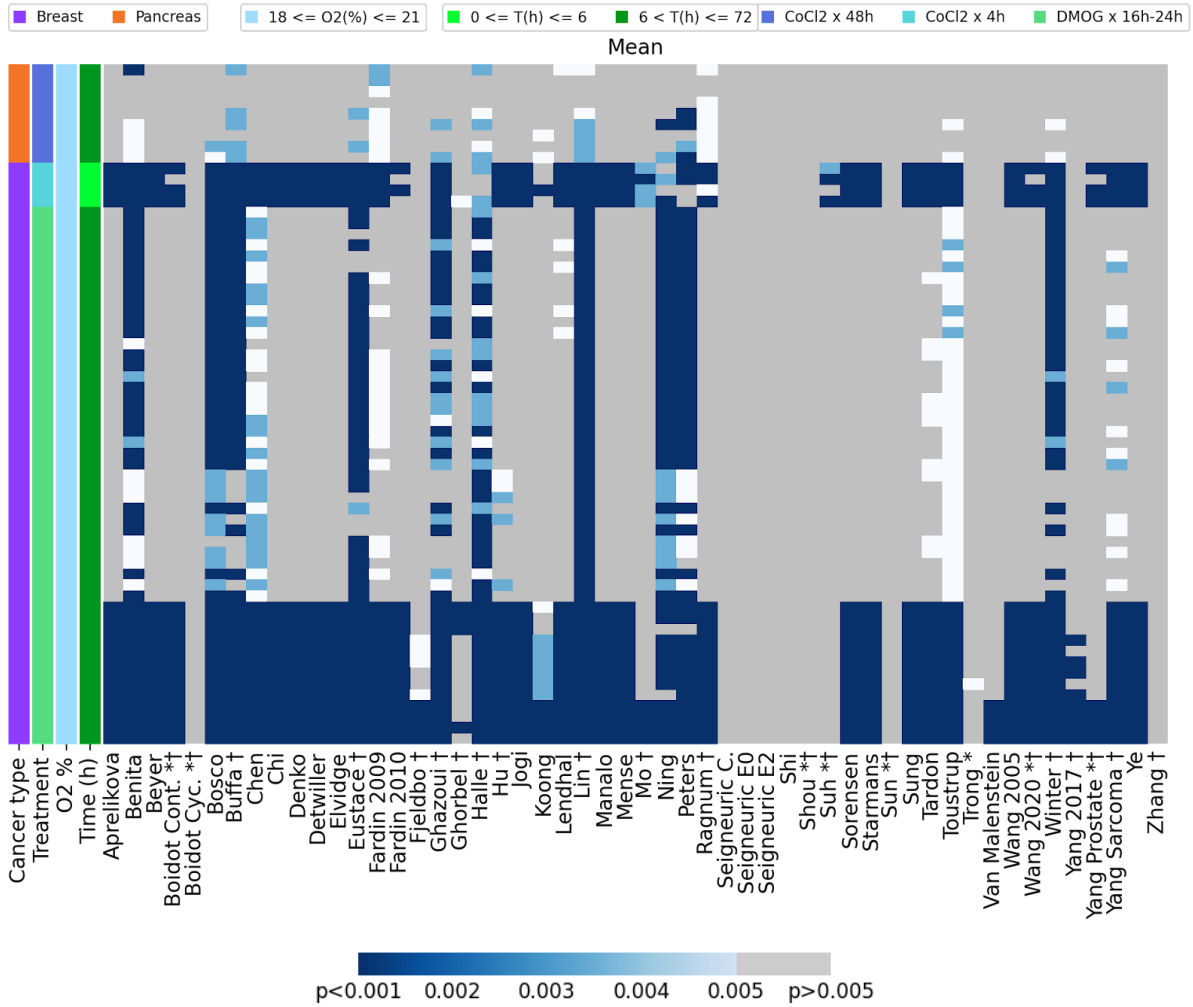
hypoxia and mimicked hypoxia with 100% accuracy *(Table 5.16*). This suggests although there are similarities, there are some significant differences in gene expression between hypoxic and mimicked hypoxic samples. The aforementioned gene expression signatures can be used to distinguish between the two conditions.

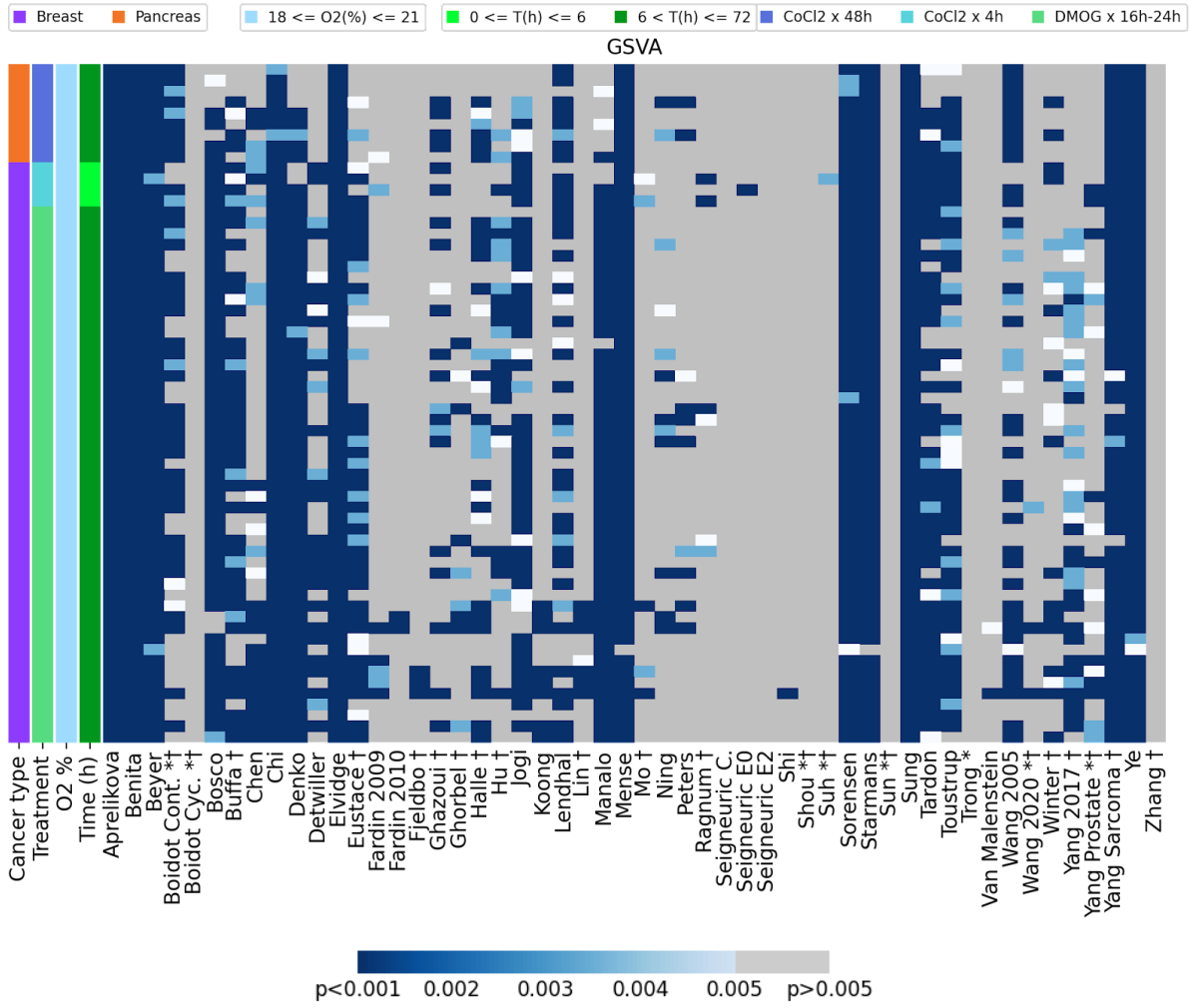
Cell line/samples	Number of Mimicked hypoxic samples	Number of Normoxic samples	Pairs
MCF-7	11	11	49
PANC-1	3	3	9
MB231RN-LM	2	2	4
Total	**16**	**16**	**62**

Table 5.14: *Cell lines used in mimicked in hypoxia experiments in GEO*

Details of cell lines in mimicked hypoxia versus normoxia experiments in GEO.


Breast
Pancreas
18 <= O2(%) <= 21
0 <= T(h) <= 6
6 < T(h) <= 72
CoCl2 x 48h
CoCl2 x 4h
DMOG x 16h-24h
Median
Cancer type
Treatment
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


Breast
Pancreas
18 <= O2(%) <= 21
0 <= T(h) <= 6
6 < T(h) <= 72
CoCl2 x 48h
CoCl2 x 4h
DMOG x 16h-24h
Mean
Cancer type
Treatment
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


Breast
Pancreas
18 <= O2(%) <= 21
0 <= T(h) <= 6
6 < T(h) <= 72
CoCl2 x 48h
CoCl2 x 4h
DMOG x 16h-24h
GSVA
Cancer type
Treatment
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

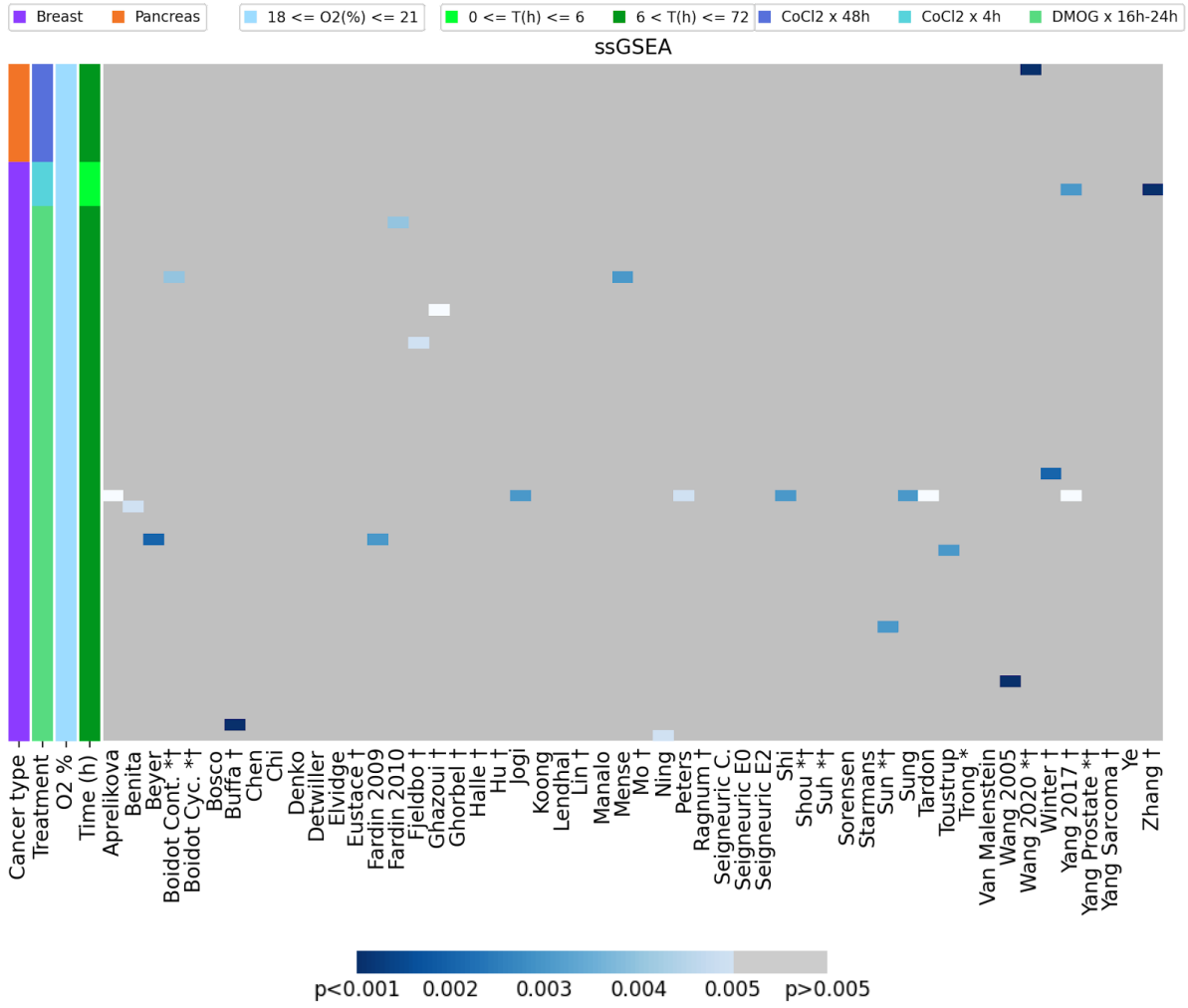


Figure 5.10: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in cancer cell lines exposed to mimicked hypoxia*

Comparison of the scoring methods of the 53 hypoxia signatures across three cell lines exposed to mimicked hypoxia. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	100.00	27.42	100.00	1.61
Benita	100.00	93.55	100.00	1.61
Beyer	98.39	27.42	100.00	1.61
Boidot Cont. *†	100.00	25.81	67.74	1.61
Boidot Cyc. *†	9.68	0.00	0.00	0.00
Bosco	88.71	87.10	90.32	0.00
Buffa †	96.77	80.65	72.58	1.61
Chen	100.00	82.26	45.16	0.00
Chi	100.00	27.42	100.00	0.00
Denko	93.55	27.42	90.32	0.00
Detwiller	83.87	27.42	46.77	0.00
Elvidge	100.00	27.42	100.00	0.00
Eustace †	98.39	79.03	70.97	0.00
Fardin 2009	98.39	64.52	12.90	1.61
Fardin 2010	87.10	24.19	3.23	1.61
Fjeldbo †	93.55	12.90	4.84	1.61
Ghazoui †	93.55	77.42	33.87	1.61
Ghorbel †	8.06	8.06	14.52	0.00
Halle †	100.00	90.32	51.61	0.00
Hu †	91.94	35.48	38.71	0.00
Jogi	96.77	27.42	80.65	1.61
Koong	27.42	22.58	12.90	0.00
Lendhal	91.94	35.48	66.13	0.00
Lin †	100.00	95.16	11.29	0.00
Manalo	91.94	27.42	85.48	0.00
Mense	100.00	27.42	98.39	1.61
Mo †	91.94	12.90	8.06	0.00
Ning	62.90	82.26	19.35	1.61
Peters	79.03	85.48	16.13	1.61
Ragnum †	45.16	38.71	9.68	0.00
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	3.23	0.00	1.61	0.00
Seigneuric E2	6.45	0.00	0.00	0.00
Shi	1.61	0.00	1.61	1.61
Shou *†	0.00	0.00	0.00	0.00
Suh *†	3.23	4.84	1.61	0.00
Sorensen	100.00	27.42	100.00	0.00
Starmans	100.00	27.42	98.39	0.00
Sun *†	0.00	0.00	0.00	1.61
Sung	100.00	27.42	100.00	1.61
Tardon	100.00	43.55	82.26	1.61
Toustrup	100.00	88.71	82.26	1.61
Trong *	0.00	1.61	0.00	0.00
Van Malenstein	83.87	6.45	3.23	0.00
Wang 2005	98.39	27.42	64.52	1.61
Wang 2020 *†	91.94	25.81	4.84	1.61
Winter †	90.32	75.81	33.87	1.61
Yang 2017 †	80.65	12.90	58.06	3.23
Yang Prostate *†	69.35	11.29	27.42	0.00
Yang Sarcoma †	100.00	50.00	96.77	0.00
Ye	100.00	27.42	100.00	0.00
Zhang †	0.00	0.00	0.00	1.61
Zou *†	6.45	6.45	1.61	0.00

Table 5.15: *Percentage accuracy of determining normoxic samples from mimicked hypoxia for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	46.15	23.08	46.15	0.00
Benita	53.85	100.00	61.54	0.00
Beyer	38.46	38.46	0.00	0.00
Boidot Cont. *†	69.23	69.23	46.15	0.00
Boidot Cyc. *†	23.08	23.08	23.08	0.00
Bosco	23.08	69.23	53.85	0.00
Buffa †	92.31	92.31	69.23	0.00
Chen	61.54	84.62	38.46	0.00
Chi	38.46	84.62	38.46	0.00
Denko	30.77	69.23	30.77	0.00
Detwiller	69.23	84.62	15.38	0.00
Elvidge	84.62	53.85	69.23	0.00
Eustace †	46.15	84.62	30.77	0.00
Fardin 2009	7.69	0.00	0.00	0.00
Fardin 2010	30.77	46.15	15.38	0.00
Fjeldbo †	38.46	7.69	0.00	0.00
Ghazoui †	69.23	100.00	46.15	0.00
Ghorbel †	53.85	38.46	30.77	0.00
Halle †	46.15	46.15	7.69	0.00
Hu †	7.69	0.00	0.00	0.00
Jogi	53.85	76.92	38.46	0.00
Koong	15.38	7.69	0.00	0.00
Lendhal	53.85	100.00	30.77	0.00
Lin †	61.54	53.85	15.38	7.69
Manalo	7.69	61.54	38.46	0.00
Mense	38.46	76.92	46.15	0.00
Mo †	23.08	7.69	0.00	0.00
Ning	69.23	69.23	30.77	0.00
Peters	92.31	92.31	69.23	0.00
Ragnum †	69.23	30.77	23.08	0.00
Seigneuric C.	0.00	0.00	0.00	0.00
Seigneuric E0	0.00	0.00	0.00	0.00
Seigneuric E2	15.38	30.77	0.00	0.00
Shi	0.00	0.00	7.69	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	15.38	7.69	0.00	0.00
Sorensen	76.92	84.62	69.23	0.00
Starmans	92.31	76.92	84.62	7.69
Sun *†	0.00	0.00	0.00	0.00
Sung	69.23	69.23	53.85	0.00
Tardon	61.54	92.31	46.15	7.69
Toustrup	76.92	92.31	61.54	0.00
Trong *	0.00	0.00	0.00	0.00
Van Malenstein	46.15	23.08	15.38	0.00
Wang 2005	53.85	69.23	0.00	0.00
Wang 2020 *†	15.38	46.15	15.38	0.00
Winter †	38.46	76.92	61.54	0.00
Yang 2017 †	0.00	30.77	15.38	0.00
Yang Prostate *†	0.00	0.00	7.69	0.00
Yang Sarcoma †	92.31	92.31	23.08	0.00
Ye	76.92	100.00	61.54	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	0.00	0.00	0.00

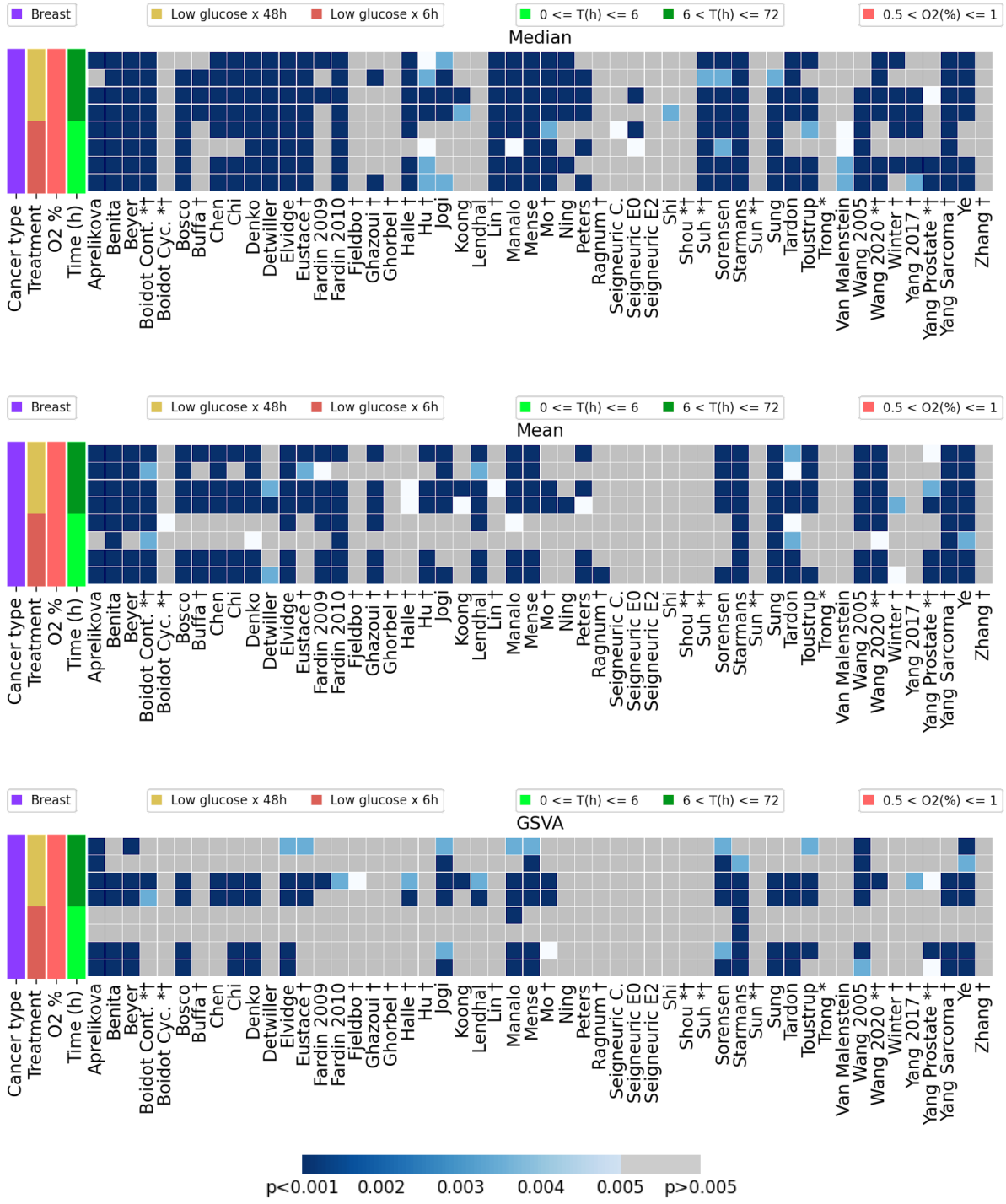
Table 5.16: *Percentage accuracy of differentiating between true hypoxia and mimicked hypoxia using the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

5.6 The effect of low glucose media on hypoxia signature performance

One breast cancer cell line was identified that had gene expression data in hypoxia and low glucose media compared to normoxia (lung metastatic subline of MDA-231 [LM], four hypoxic samples in low glucose media and two normoxic with normal glucose controls for a total of eight pairs). In this experiment, cells were exposed to hypoxia (1% O2) with low (1 mg/ml) glucose for 6 or 48 hours[158] (GSE107300).

I assessed the performance of the 53 signatures at defining hypoxia in this condition. ssGSEA performed poorly. Median score was the most effective (*Fig. 5.11*, top panel) with 21 signatures achieving 100% accuracy (*Table 5.17*). The best performing breast cancer signatures using the median, Benita, Sorensen and Yang sarcoma, again all hit 100% accuracy suggesting they are not affected by reducing the glucose concentration to 1 mg/ml environments. Eight signatures also had 100% accuracy using mean, including five which also had 100% accuracy using the median (Benita, Boidot continuous, Starmans, Tardon and Yang Sarcoma). Aprelikova GSVA had 75% accuracy for hypoxia in the low glucose condition.


Breast
Low glucose x 48h
Low glucose x 6h
0 <= T(h) <= 6
6 < T(h) <= 72
0.5 < O2(%) <= 1
Median
Mean
GSVA
Cancer type
Treatment
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup *
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

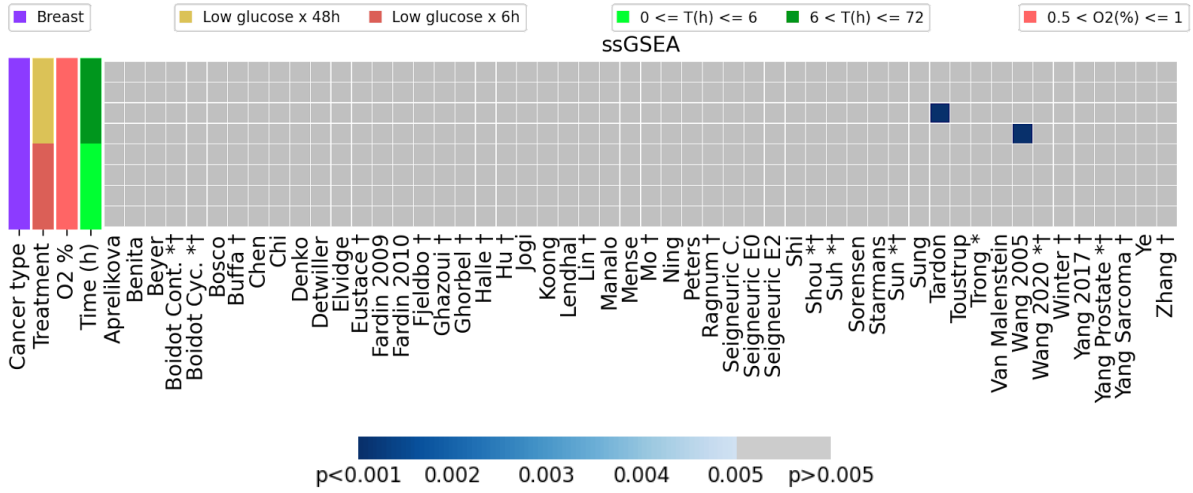


Figure 5.11: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in cancer cell lines which are in low glucose media*

Comparison of the scoring methods of the 53 hypoxia signatures across all cell lines grown in low glucose media (1 mg/ml). Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	87.50	87.50	75.00	0.00
Benita	100.00	100.00	50.00	0.00
Beyer	100.00	87.50	62.50	0.00
Boidot Cont. *†	100.00	100.00	25.00	0.00
Boidot Cyc. *†	0.00	12.50	0.00	0.00
Bosco	87.50	75.00	50.00	0.00
Buffa †	37.50	62.50	0.00	0.00
Chen	87.50	75.00	25.00	0.00
Chi	87.50	62.50	50.00	0.00
Denko	100.00	87.50	50.00	0.00
Detwiller	100.00	37.50	0.00	0.00
Elvidge	100.00	87.50	62.50	0.00
Eustace †	100.00	50.00	37.50	0.00
Fardin 2009	25.00	87.50	12.50	0.00
Fardin 2010	100.00	87.50	12.50	0.00
Fjeldbo †	0.00	0.00	12.50	0.00
Ghazoui †	25.00	75.00	0.00	0.00
Ghorbel †	0.00	0.00	0.00	0.00
Halle †	87.50	25.00	25.00	0.00
Hu †	87.50	62.50	0.00	0.00
Jogi	62.50	62.50	75.00	0.00
Koong	25.00	25.00	12.50	0.00
Lendhal	0.00	87.50	25.00	0.00
Lin †	100.00	12.50	0.00	0.00
Manalo	100.00	87.50	75.00	0.00
Mense	100.00	75.00	75.00	0.00
Mo †	100.00	25.00	37.50	0.00
Ning	62.50	12.50	0.00	0.00
Peters	62.50	62.50	0.00	0.00
Ragnum †	0.00	12.50	0.00	0.00
Seigneuric C.	12.50	0.00	0.00	0.00
Seigneuric E0	37.50	0.00	0.00	0.00
Seigneuric E2	0.00	0.00	0.00	0.00
Shi	12.50	0.00	0.00	0.00
Shou *†	0.00	0.00	0.00	0.00
Suh *†	100.00	0.00	0.00	0.00
Sorensen	100.00	62.50	75.00	0.00
Starmans	100.00	100.00	87.50	0.00
Sun *†	0.00	0.00	0.00	0.00
Sung	87.50	100.00	50.00	0.00
Tardon	75.00	100.00	50.00	12.50
Toustrup	75.00	62.50	50.00	0.00
Trong *	0.00	0.00	0.00	0.00
Van Malenstein	50.00	0.00	0.00	0.00
Wang 2005	75.00	87.50	75.00	12.50
Wang 2020 *†	75.00	100.00	12.50	0.00
Winter †	62.50	25.00	0.00	0.00
Yang 2017 †	62.50	0.00	12.50	0.00
Yang Prostate *†	37.50	62.50	37.50	0.00
Yang Sarcoma †	100.00	100.00	50.00	0.00
Ye	75.00	100.00	75.00	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	0.00	0.00	0.00

Table 5.17: *Percentage accuracy of determining hypoxic samples from normoxic samples in low glucose media for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

5.7 The performance of hypoxia signatures in cycling hypoxia

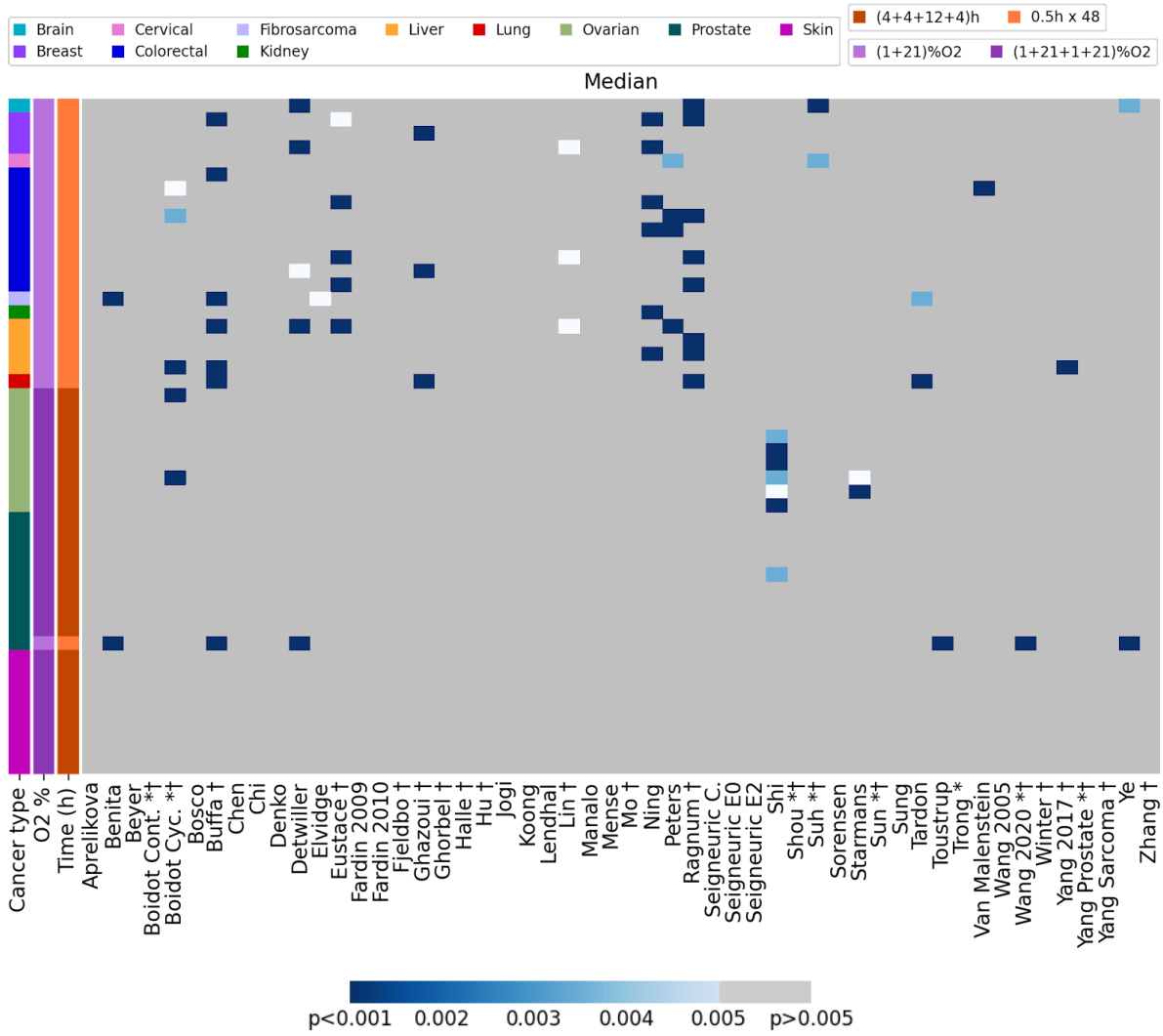
The erratic blood supply in tumours due to aberrant vessel growth has been hypothesised to lead to intermittent periods of hypoxia. Therefore, researchers have tried to match this approach in the laboratory by exposing cells to normal and low oxygen tensions in a cyclic fashion. This approach has been shown to give different gene expression profiles as opposed to exposure to just hypoxia alone[103]. Twenty-one cancer cell lines (*Table 5.18*) from a range of tissues (prostate, ovarian, melanoma, breast [4 lines], lung, colon [6 lines], cervical, brain, liver [4 lines], kidney, fibrosarcoma) that had gene expression data in cycling hypoxia and paired samples in continuous normoxia were identified from the Gene Expression Omnibus (as per methods). Cycling hypoxia protocols (see legend, *Fig. 5.12*) varied but data were combined to assess signature and summary score performance.

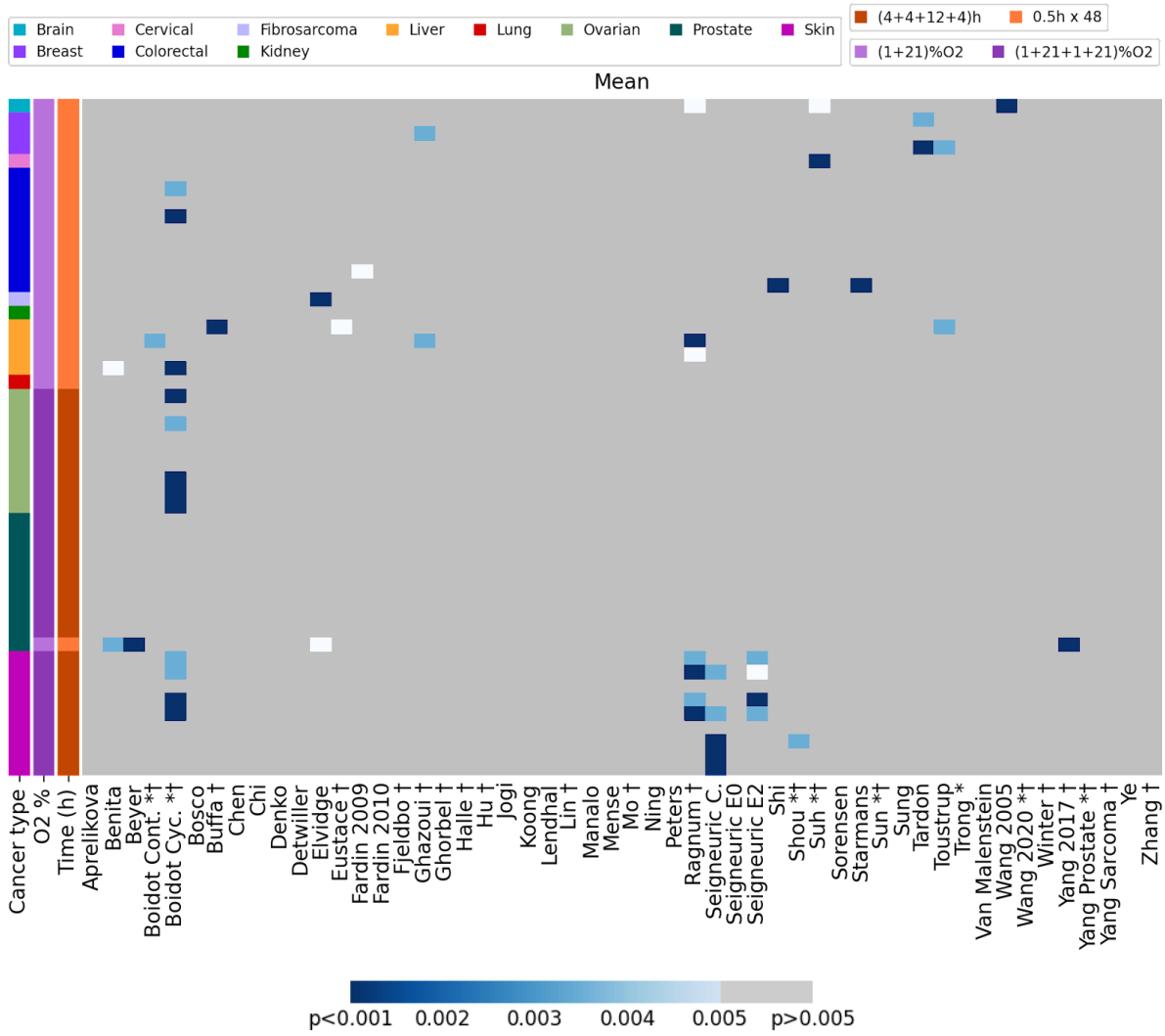
In the main, signatures performed worse in cycling hypoxia. ssGSEA was the worst performing score. Interestingly, using the GSVA score, the signature that reached the highest accuracy of 36.73% was Boidot Cyclic Hypoxia that was developed using cell line data based on experiments of cycling hypoxia (*Table 5.19*). But overall, the results suggest that these signatures are not reliable in the context of cycling hypoxia in the laboratory.

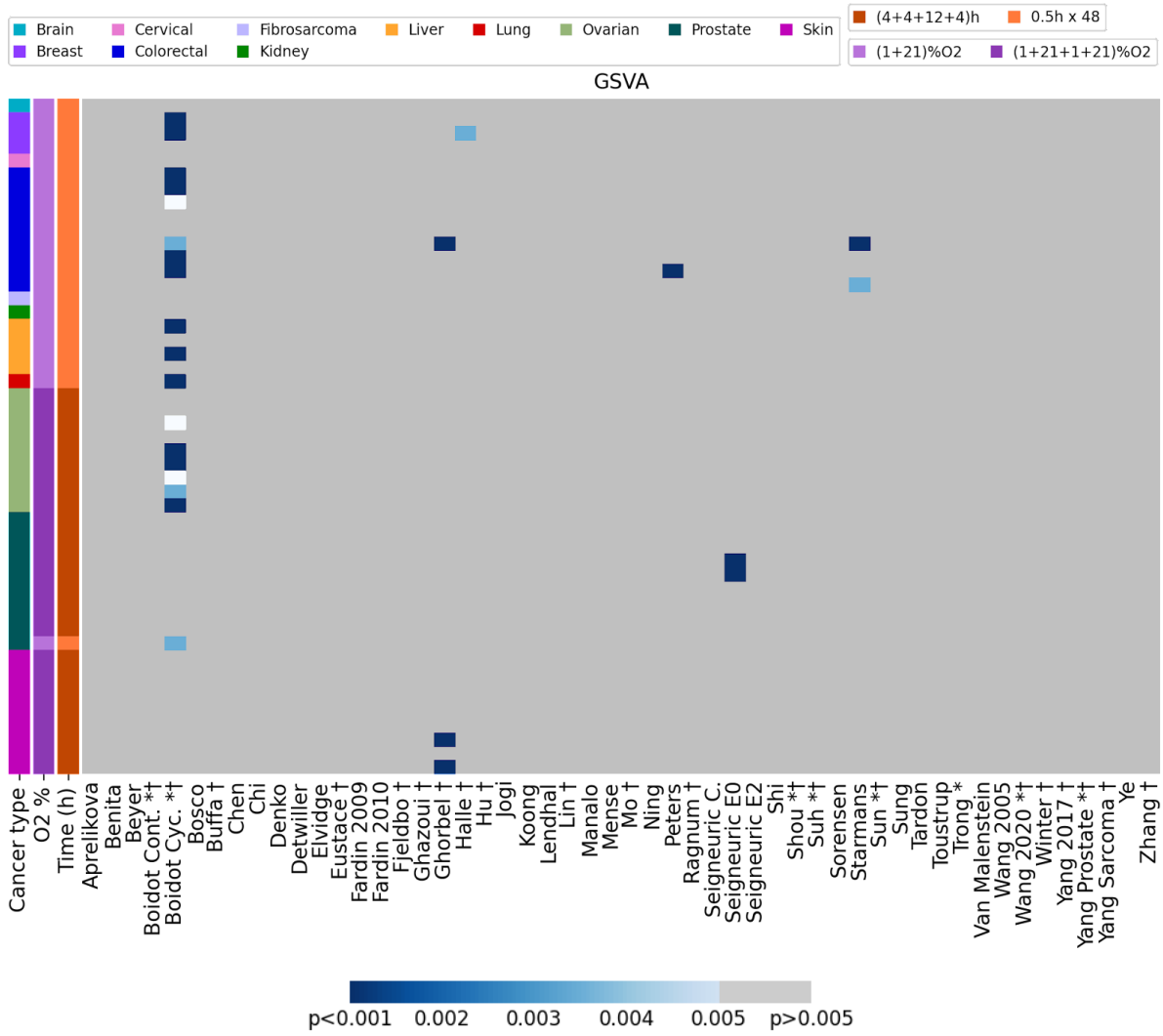
Cell line/samples	Number of Normoxic samples	Number of samples under cycling hypoxia	Pairs
PC-3	4	4	10
WM793B	3	3	9
SK-OV-3	3	3	9
HCT116	2	2	4
MDA-MB-231	1	1	1
SiHa	1	1	1
WiDr	1	1	1
U373	1	1	1
T-47D	1	1	1
SK-HEP-1	1	1	1
PLC-PRF-5	1	1	1
A498	1	1	1
A549	1	1	1
LoVo	1	1	1
HepG2	1	1	1
Hep3B	1	1	1
HT29	1	1	1
HT1080	1	1	1
HCT-15	1	1	1
COLO-205	1	1	1
MCF-7	1	1	1
Total	29	29	49

Table 5.18: *Cell lines used in normoxia vs. cycling hypoxia experiments in GEO*

Details of cell lines in normoxia vs cycling hypoxia experiments in GEO.


Brain
Breast
Cervical
Colorectal
Fibrosarcoma
Kidney
Liver
Lung
Ovarian
Prostate
Skin
(4+4+12+4)h
0.5h x 48
(1+21)%O2
(1+21+1+21)%O2
Median
Cancer type
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont.*†
Boidot Cyc.*†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou*†
Suh*†
Sorensen
Starmans
Sun*†
Sung
Tardon
Toustrup
Trong*
Van Malenstein
Wang 2005
Wang 2020*†
Winter †
Yang 2017 †
Yang Prostate*†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


Brain
Breast
Cervical
Colorectal
Fibrosarcoma
Kidney
Liver
Lung
Ovarian
Prostate
Skin
(4+4+12+4)h
0.5h x 48
(1+21)%O2
(1+21+1+21)%O2
Mean
Cancer type
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont.*†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


Brain
Breast
Cervical
Colorectal
Fibrosarcoma
Kidney
Liver
Lung
Ovarian
Prostate
Skin
(4+4+12+4)h
0.5h x 48
(1+21)%O2
(1+21+1+21)%O2
GSVA
Cancer type
O2 %
Time (h)
Aprelikova
Benita
Beyer
Boidot Cont.*†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

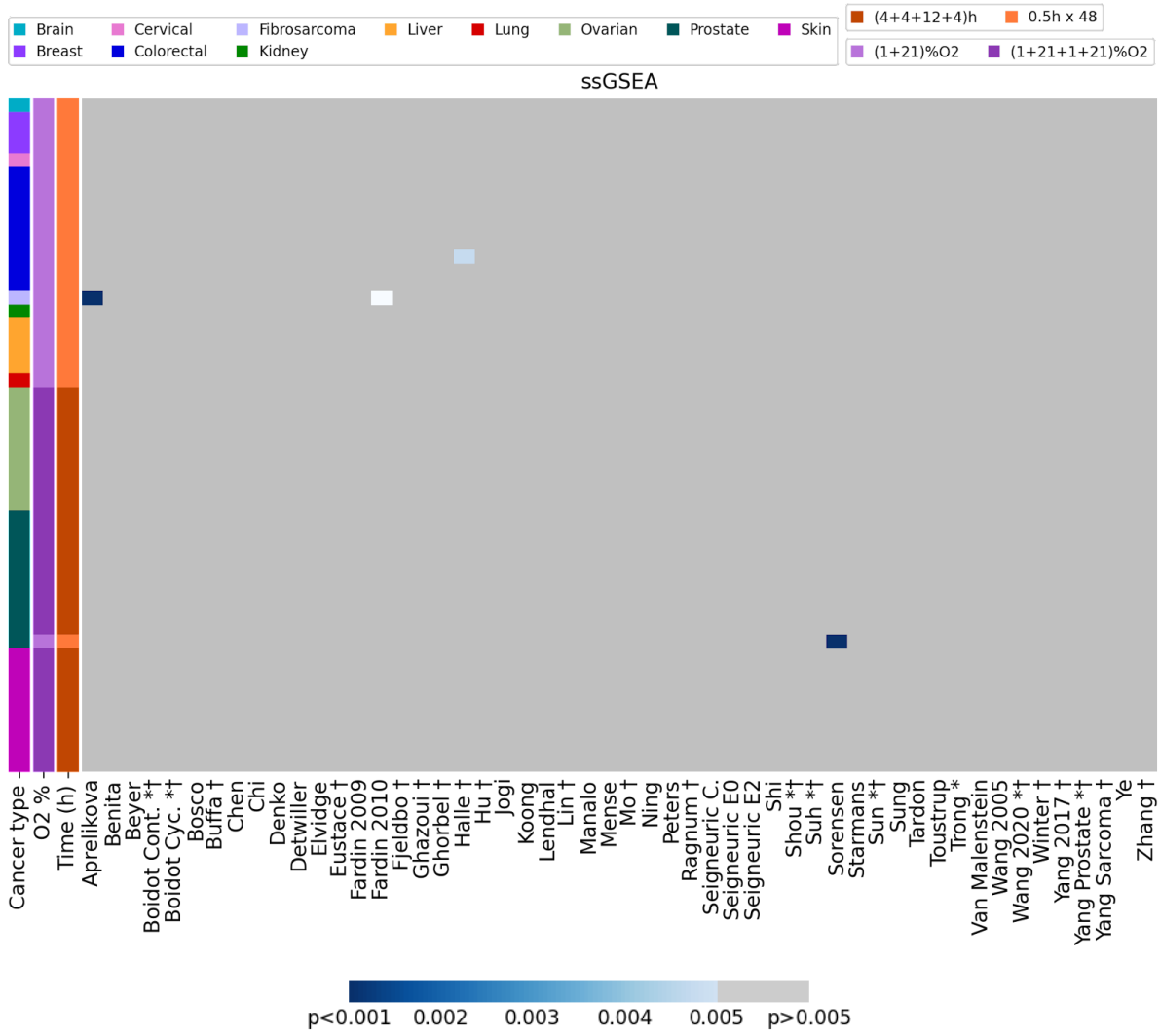


Figure 5.12: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in cancer cell lines which have been exposed to cycles of hypoxia and compared to normoxia*

Comparison of the scoring methods of the 53 hypoxia signatures exposed to cycles of hypoxia. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	0.00	0.00	0.00	2.04
Benita	4.08	4.08	0.00	0.00
Beyer	0.00	2.04	0.00	0.00
Boidot Cont. *†	0.00	2.04	0.00	0.00
Boidot Cyc. *†	10.20	24.49	36.73	0.00
Bosco	0.00	0.00	0.00	0.00
Buffa †	14.29	2.04	0.00	0.00
Chen	0.00	0.00	0.00	0.00
Chi	0.00	0.00	0.00	0.00
Denko	0.00	0.00	0.00	0.00
Detwiller	10.20	0.00	0.00	0.00
Elvidge	2.04	4.08	0.00	0.00
Eustace †	10.20	2.04	0.00	0.00
Fardin 2009	0.00	2.04	0.00	0.00
Fardin 2010	0.00	0.00	0.00	2.04
Fjeldbo †	0.00	0.00	0.00	0.00
Ghazoui †	6.12	4.08	0.00	0.00
Ghorbel †	0.00	0.00	6.12	0.00
Halle †	0.00	0.00	2.04	2.04
Hu †	0.00	0.00	0.00	0.00
Jogi	0.00	0.00	0.00	0.00
Koong	0.00	0.00	0.00	0.00
Lendhal	0.00	0.00	0.00	0.00
Lin †	6.12	0.00	0.00	0.00
Manalo	0.00	0.00	0.00	0.00
Mense	0.00	0.00	0.00	0.00
Mo †	0.00	0.00	0.00	0.00
Ning	12.24	0.00	0.00	0.00
Peters	8.16	0.00	2.04	0.00
Ragnum †	16.33	14.29	0.00	0.00
Seigneuric C.	0.00	10.20	0.00	0.00
Seigneuric E0	0.00	0.00	4.08	0.00
Seigneuric E2	0.00	8.16	0.00	0.00
Shi	14.29	2.04	0.00	0.00
Shou *†	0.00	2.04	0.00	0.00
Suh *†	4.08	4.08	0.00	0.00
Sorensen	0.00	0.00	0.00	2.04
Starmans	4.08	2.04	4.08	0.00
Sun *†	0.00	0.00	0.00	0.00
Sung	0.00	0.00	0.00	0.00
Tardon	4.08	4.08	0.00	0.00
Toustrup	2.04	4.08	0.00	0.00
Trong *	0.00	0.00	0.00	0.00
Van Malenstein	2.04	0.00	0.00	0.00
Wang 2005	0.00	2.04	0.00	0.00
Wang 2020 *†	2.04	0.00	0.00	0.00
Winter †	0.00	0.00	0.00	0.00
Yang 2017 †	2.04	2.04	0.00	0.00
Yang Prostate *†	0.00	0.00	0.00	0.00
Yang Sarcoma †	0.00	0.00	0.00	0.00
Ye	4.08	0.00	0.00	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	0.00	0.00	0.00	0.00

Table 5.19: *Percentage accuracy of determining hypoxic samples from normoxic samples in cells exposed to cycling hypoxia versus normoxia for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

5.8 The best performing signature score across all cell lines and experimental conditions

To conclude this chapter, an assessment will be conducted on the performance of hypoxia gene expression signatures on an unprecedented scale. This will involve examining the conditions discussed in this chapter, as well as the different cancer types explored in Chapter 4.12. To recap, the analysis in 4.12, saw the Aprelikova signature using the GVSA score being the best performing score across the cancer types, with only poor performance on lung cancer (which was mainly *ex vivo* samples). Cell lines with VHL mutations in this analysis were not included as one would expect this might influence signature performance, as demonstrated in Chapter 5.3. However, the analysis was able to identify signatures with good performance in this condition. When VHL was introduced to one of these cell lines, some hypoxia signatures did seem to work, although the sample size was extremely limited. With this evidence, it is prudent to suggest tissue specific signatures for renal cancer cell lines as well as highlighting the preference of establishing VHL mutation status before applying a hypoxia signature.

In terms of potential microenvironmental conditions recreated at the benchside, low glucose did not hugely affect performance. However, one experiment in cycling hypoxia did yield poor performance of all signatures using the scores that had worked well/moderately well in other areas (median, mean and GVSA). In mimicked hypoxia, sees some signatures highlighted previously as performing well, but also uncovers gene expression differences between these two states. Therefore in the global analysis, since the goal is to find a signature that would be universal, the immortalised normal cell lines were also included. However, the following were excluded:

1) Lung and renal cancer samples, as they appear to respond to hypoxia differently compared to the other cancer types
2) Mimicked hypoxia, as this does not occur in the clinic

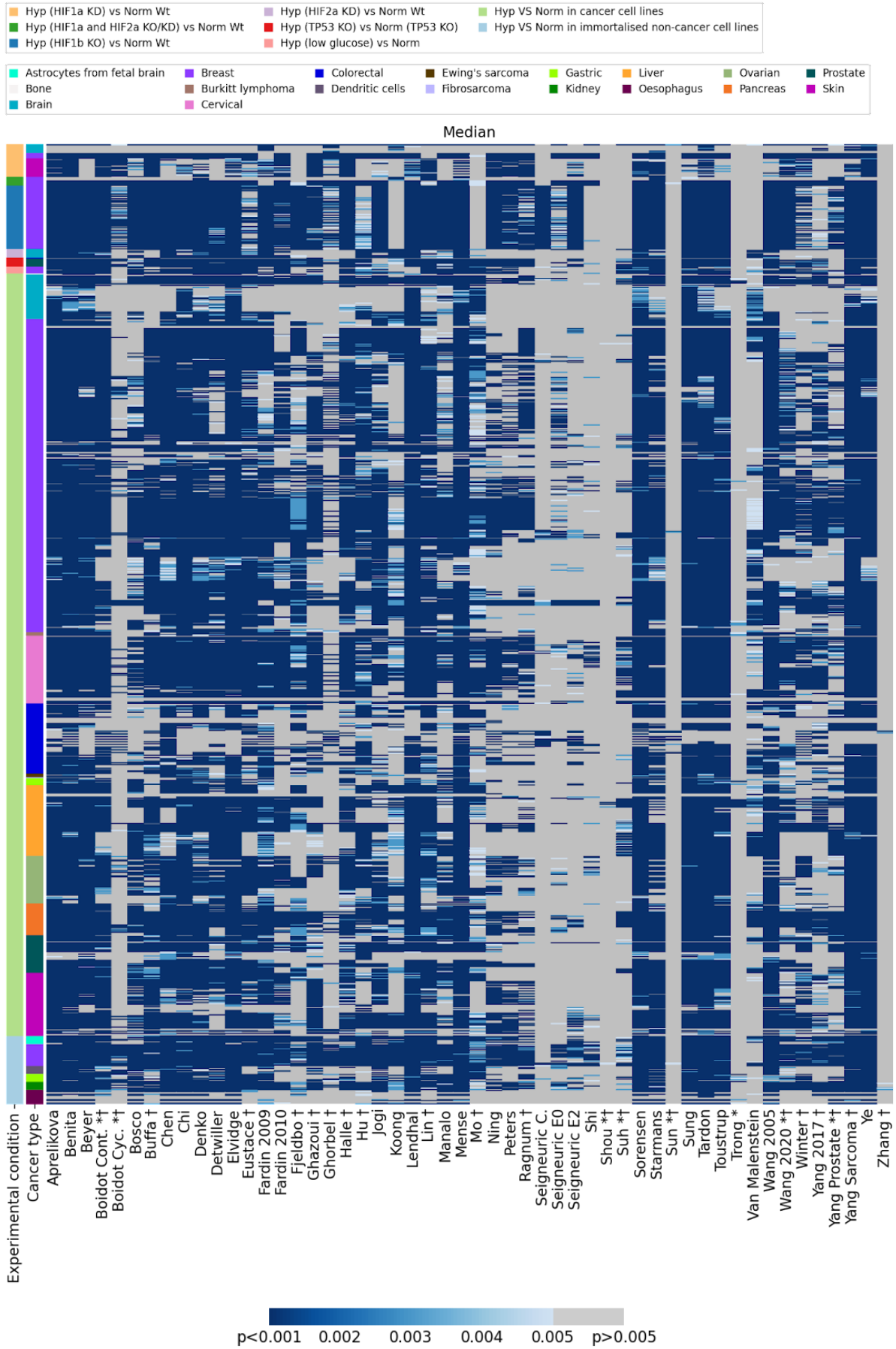
3) Cycling hypoxia experiments, as they show conflicting results depending on the experimental conditions used and it is unknown how representative they are of *in vivo* tumours

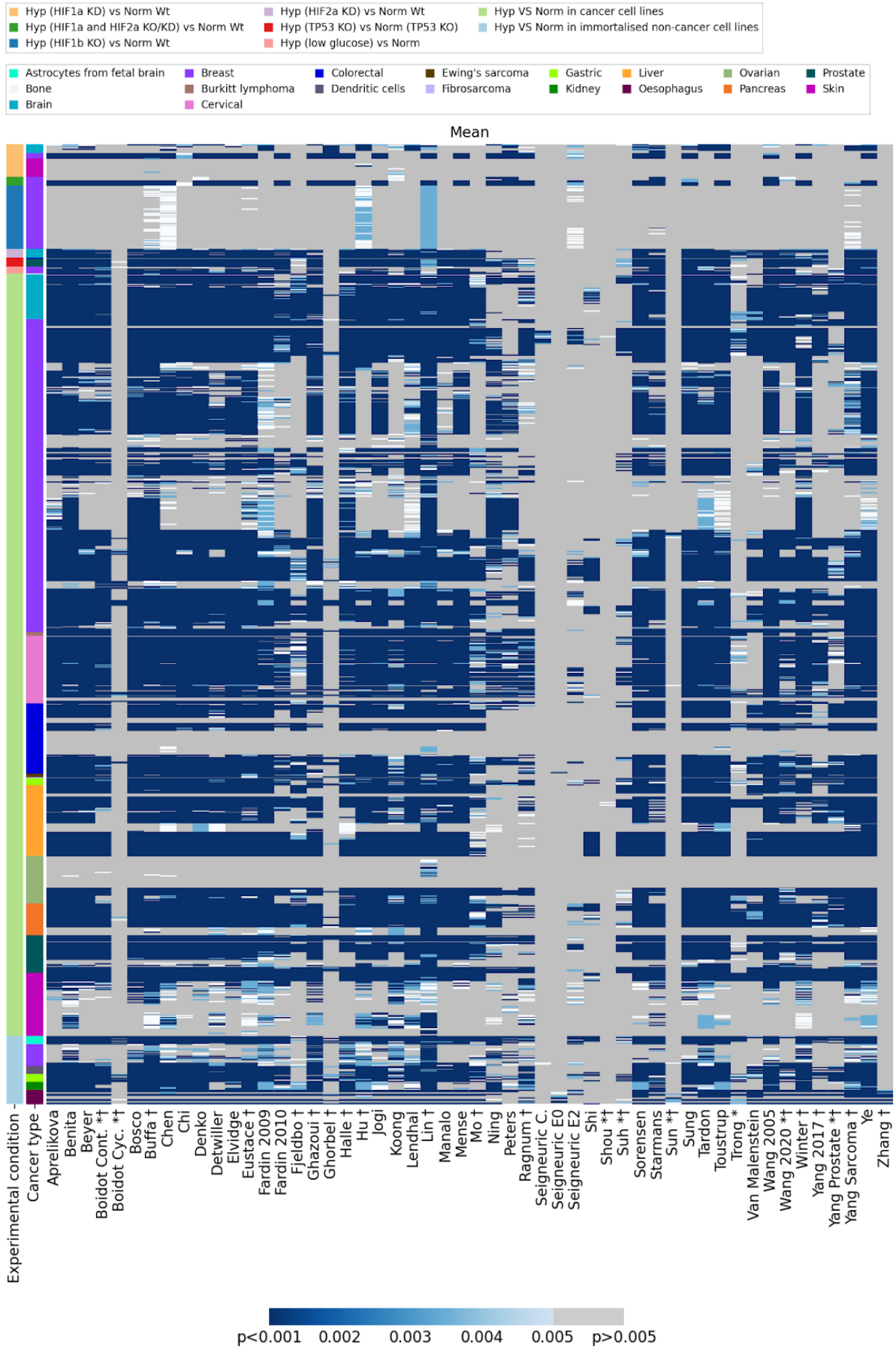
This approach yields an impressive 92.84% accuracy in identifying hypoxic samples using the Sorensen signature and median score (*Fig. 5.13*, *Table 5.20*) across 1090 pairwise combinations from 98 different cell lines (*Appendix 2, Supplementary Table S3*). The samples in which the signature does not perform well are detailed in *Fig. 5.14*. Notably, the signature fails on some HIF-1a knockdown experiments, mainly in the glioblastoma cell line (LN229, nine pairs) but also in a number of other experiments which other signatures also struggle to define as hypoxia using the median score. Finally, in lung using Starmans or Mense with mean score is recommended (although bear in mind *ex vivo* samples here made up most of the cohort) and in renal cell lines, Sung using the GVSA score to investigate hypoxic status seems prudent.

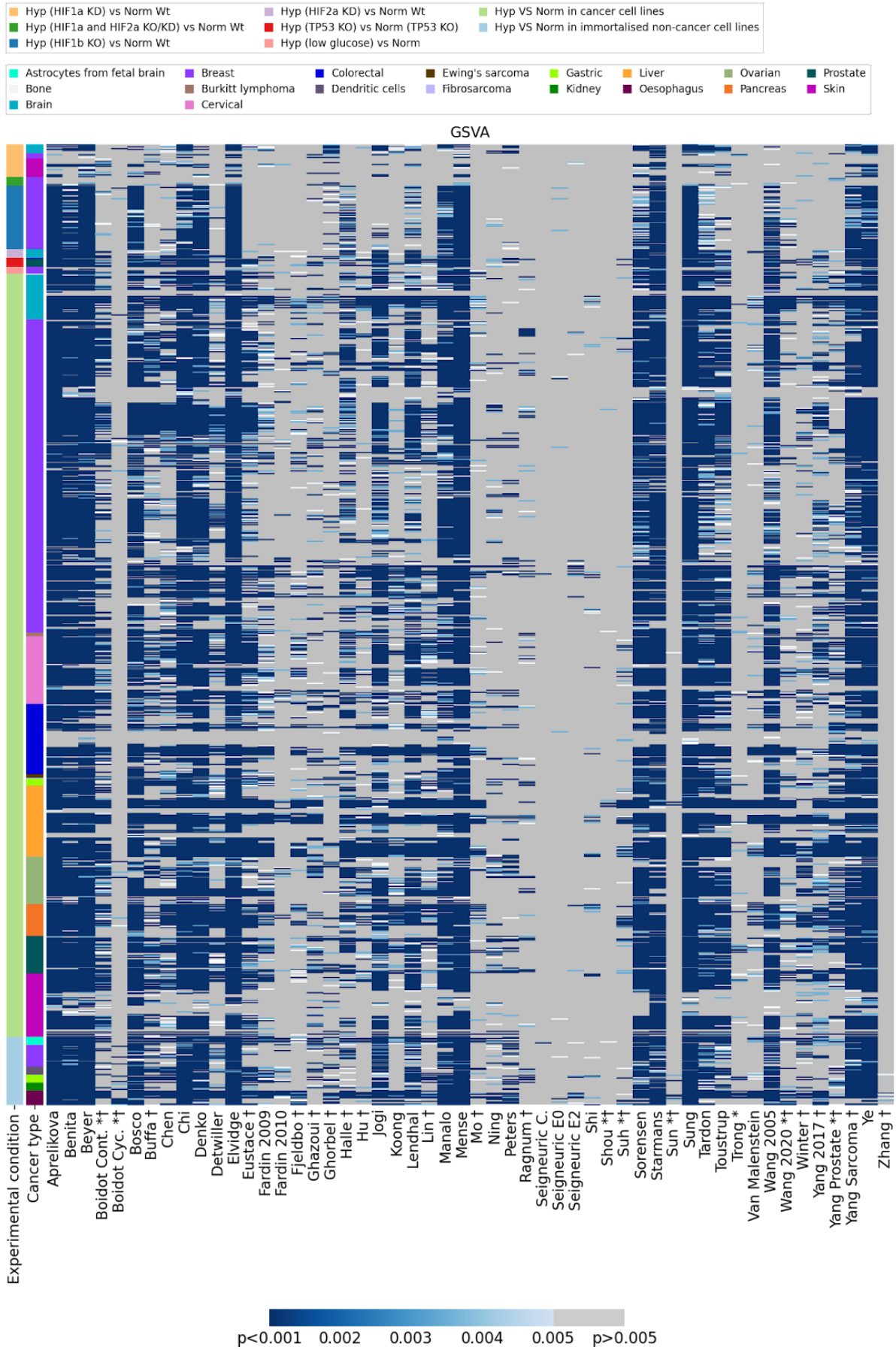
The origin of the best performing signature, Sorensen, comes from five different cell lines, four head and neck (FaDuDD, UTSCC5, UTSCC14 and UTSCC15) and one cervical (SiHa)[70]. The aim of the study was to identify a robust hypoxia profile unaffected by pH across cell lines, as the authors had previously shown that hypoxic gene expression was suppressed by low extracellular pH[159,160]. Therefore, in this work, these five lines were exposed to different oxygen concentrations (0%, 0.1%, 1.5% or 21% oxygen) and different pHs (7.5 or 6.3) and gene expression analysed with the microarray platform, Affymetrix Human Genome U133 Plus 2.0. To identify hypoxia inducible pH independent genes expression had to be at least 2-fold upregulated in 0.1% oxygen at both normal and low pH, and there had to be at least a 2-fold difference between 5% O2 pH 7.5 and 0.1% O2 pH 6.3. Furthermore, expression at 5% O2 pH 6.3 should not exceed 50% of the difference between 5% O2 pH 7.5 and 0.1% O2 pH 7.5. Common genes meeting these criteria were identified from each cell line but there was no commonality when the five cell lines were considered, however SiHa was excluded, and this yielded a common set of 37 microarray probesets,

representing 27 different genes[70]. This approach suggests that perhaps controlling for pH is important for deriving hypoxia signatures, a factor not overly focussed on in most analyses.

After determining the most effective universal signature through this extensive analysis of all publicly available hypoxic and matched normoxic cell line and ex vivo samples, the study is well-positioned to address the next challenge: exploring the utility of hypoxia signatures in a large dataset of primary cancer and normal tissue samples across multiple tissue types, using The Cancer Genome Atlas (TCGA).


Hyp (HIF1a KD) vs Norm Wt
Hyp (HIF1a and HIF2a KO/KD) vs Norm Wt
Hyp (HIF1b KO) vs Norm Wt
Hyp (HIF2a KD) vs Norm Wt
Hyp (TP53 KO) vs Norm (TP53 KO)
Hyp (low glucose) vs Norm
Hyp VS Norm in cancer cell lines
Hyp VS Norm in immortalised non-cancer cell lines
Astrocytes from fetal brain
Bone
Brain
Breast
Burkitt lymphoma
Cervical
Colorectal
Dendritic cells
Ewing's sarcoma
Fibrosarcoma
Gastric
Kidney
Liver
Oesophagus
Ovarian
Pancreas
Prostate
Skin
Median
Experimental condition
Cancer type
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


Hyp (HIF1a KD) vs Norm Wt
Hyp (HIF1a and HIF2a KO/KD) vs Norm Wt
Hyp (HIF1b KO) vs Norm Wt
Hyp (HIF2a KD) vs Norm Wt
Hyp (TP53 KO) vs Norm (TP53 KO)
Hyp (low glucose) vs Norm
Hyp VS Norm in cancer cell lines
Hyp VS Norm in immortalised non-cancer cell lines
Astrocytes from fetal brain
Bone
Brain
Breast
Burkitt lymphoma
Cervical
Colorectal
Dendritic cells
Ewing's sarcoma
Fibrosarcoma
Gastric
Kidney
Liver
Oesophagus
Ovarian
Pancreas
Prostate
Skin
Mean
Experimental condition
Cancer type
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005


Hyp (HIF1a KD) vs Norm Wt
Hyp (HIF1a and HIF2a KO/KD) vs Norm Wt
Hyp (HIF1b KO) vs Norm Wt
Hyp (HIF2a KD) vs Norm Wt
Hyp (TP53 KO) vs Norm (TP53 KO)
Hyp (low glucose) vs Norm
Hyp VS Norm in cancer cell lines
Hyp VS Norm in immortalised non-cancer cell lines
Astrocytes from fetal brain
Bone
Brain
Breast
Burkitt lymphoma
Cervical
Colorectal
Dendritic cells
Ewing's sarcoma
Fibrosarcoma
Gastric
Kidney
Liver
Oesophagus
Ovarian
Pancreas
Prostate
Skin
GSVA
Experimental condition
Cancer type
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

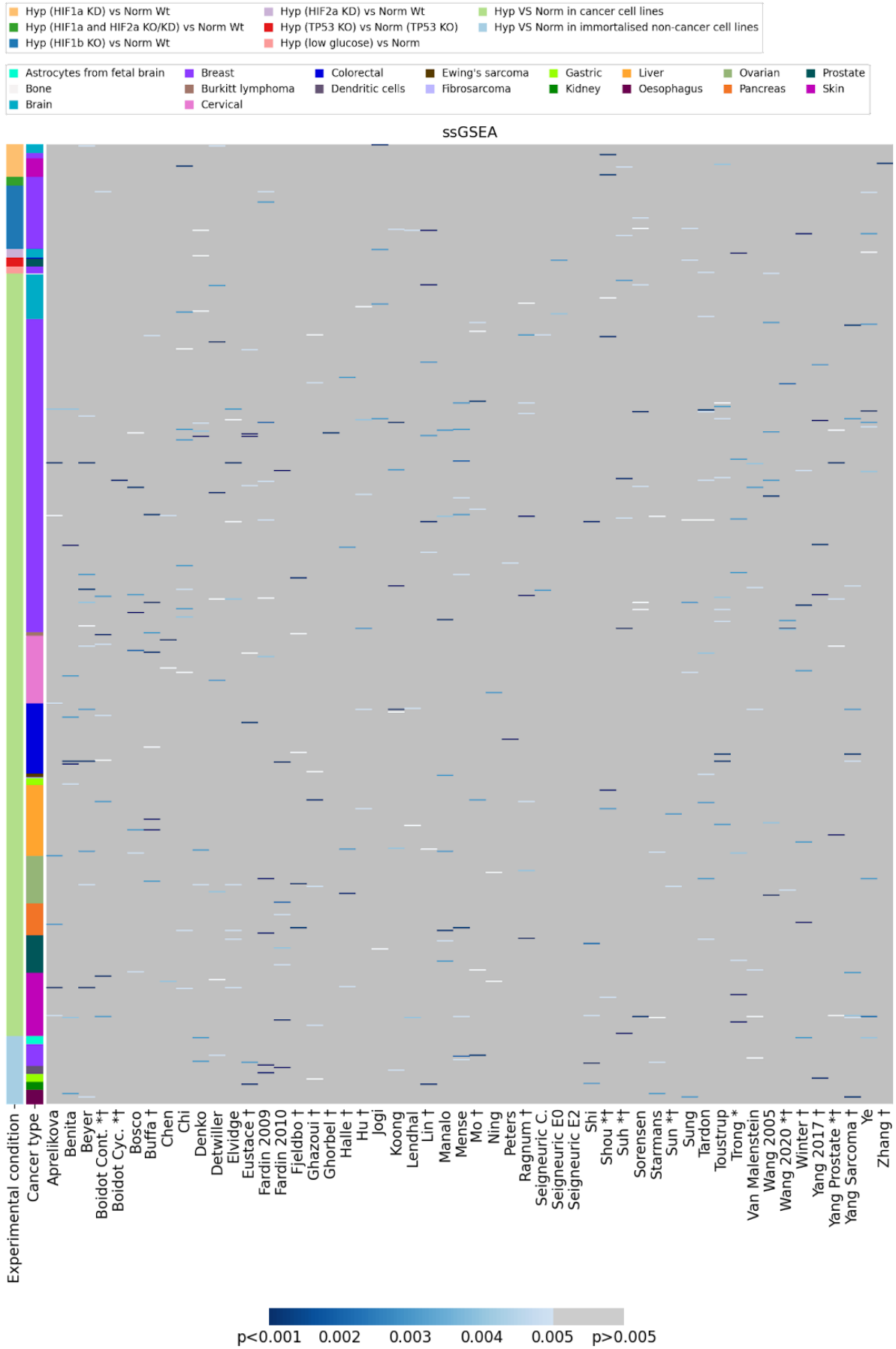

Hyp (HIF1a KD) vs Norm Wt
Hyp (HIF1a and HIF2a KO/KD) vs Norm Wt
Hyp (HIF1b KO) vs Norm Wt
Hyp (HIF2a KD) vs Norm Wt
Hyp (TP53 KO) vs Norm (TP53 KO)
Hyp (low glucose) vs Norm
Hyp VS Norm in cancer cell lines
Hyp VS Norm in immortalised non-cancer cell lines
Astrocytes from fetal brain
Bone
Brain
Breast
Burkitt lymphoma
Cervical
Colorectal
Dendritic cells
Ewing's sarcoma
Fibrosarcoma
Gastric
Kidney
Liver
Oesophagus
Ovarian
Pancreas
Prostate
Skin
ssGSEA
Experimental condition
Cancer type
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

Figure 5.13: *Comparison of the four hypoxia summary scores across the 53 published hypoxia signatures in cancer across all cell lines and experiments*

Comparison of the scoring methods of the 53 hypoxia signatures across all cell lines and experimental conditions including immortalised non-cancer cell lines. Conventions as in *Fig. 4.1*.

	Median	Mean	GSVA	ssGSEA
Aprelikova	89.63	65.50	89.27	0.73
Benita	91.10	73.03	80.73	0.83
Beyer	85.41	61.65	86.42	1.28
Boidot Cont. *†	78.44	57.71	44.13	0.83
Boidot Cyc. *†	20.09	5.05	2.84	0.09
Bosco	59.17	68.72	76.70	0.73
Buffa †	74.04	74.95	38.53	0.83
Chen	81.93	74.04	44.31	0.37
Chi	85.60	62.75	80.92	1.10
Denko	81.01	64.59	79.27	0.92
Detwiller	70.28	58.53	35.32	0.83
Elvidge	89.72	62.11	87.52	0.83
Eustace †	72.94	68.62	46.42	0.73
Fardin 2009	77.43	68.35	31.74	1.01
Fardin 2010	56.79	46.42	7.80	0.73
Fjeldbo †	57.43	37.71	17.98	0.46
Ghazoui †	57.43	63.21	19.08	0.64
Ghorbel †	21.56	7.52	13.03	0.09
Halle †	71.74	63.58	46.33	0.46
Hu †	67.71	65.41	19.54	0.55
Jogi	67.89	61.65	61.83	0.46
Koong	41.28	52.66	22.66	0.73
Lendhal	89.54	65.32	61.19	0.37
Lin †	80.18	84.95	27.61	0.73
Manalo	56.79	51.65	72.57	0.73
Mense	89.91	61.65	85.23	0.92
Mo †	52.48	40.18	12.66	0.64
Ning	47.34	38.81	11.65	0.28
Peters	37.43	27.52	11.01	0.18
Ragnum †	43.03	36.79	7.43	0.83
Seigneuric C.	11.38	1.47	0.18	0.18
Seigneuric E0	18.17	0.83	1.01	0.18
Seigneuric E2	22.11	14.95	2.39	0.00
Shi	9.27	10.73	3.94	0.46
Shou *†	0.56	0.28	0.84	0.65
Suh *†	19.45	19.08	7.43	0.64
Sorensen	92.84	63.94	76.61	0.73
Starmans	79.08	57.80	81.74	0.46
Sun *†	1.38	1.28	0.83	0.18
Sung	90.00	64.86	86.61	0.55
Tardon	92.39	70.73	71.47	1.01
Toustrup	87.61	67.34	61.19	1.01
Trong *	3.21	17.43	4.77	0.73
Van Malenstein	37.61	32.57	12.75	0.64
Wang 2005	81.65	65.41	70.37	0.64
Wang 2020 *†	56.33	46.51	19.54	0.46
Winter †	47.52	62.48	21.56	0.64
Yang 2017 †	46.15	36.88	36.97	0.46
Yang Prostate *†	42.02	34.86	32.02	0.55
Yang Sarcoma †	92.11	67.80	77.61	0.92
Ye	91.10	72.39	85.23	1.01
Zhang †	0.55	1.10	0.55	0.09
Zou *†	18.35	16.33	5.05	0.73

Table 5.20: *Percentage accuracy of determining hypoxic samples from normoxic samples in all cell lines and experimental conditions*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in *Table 4.2*.

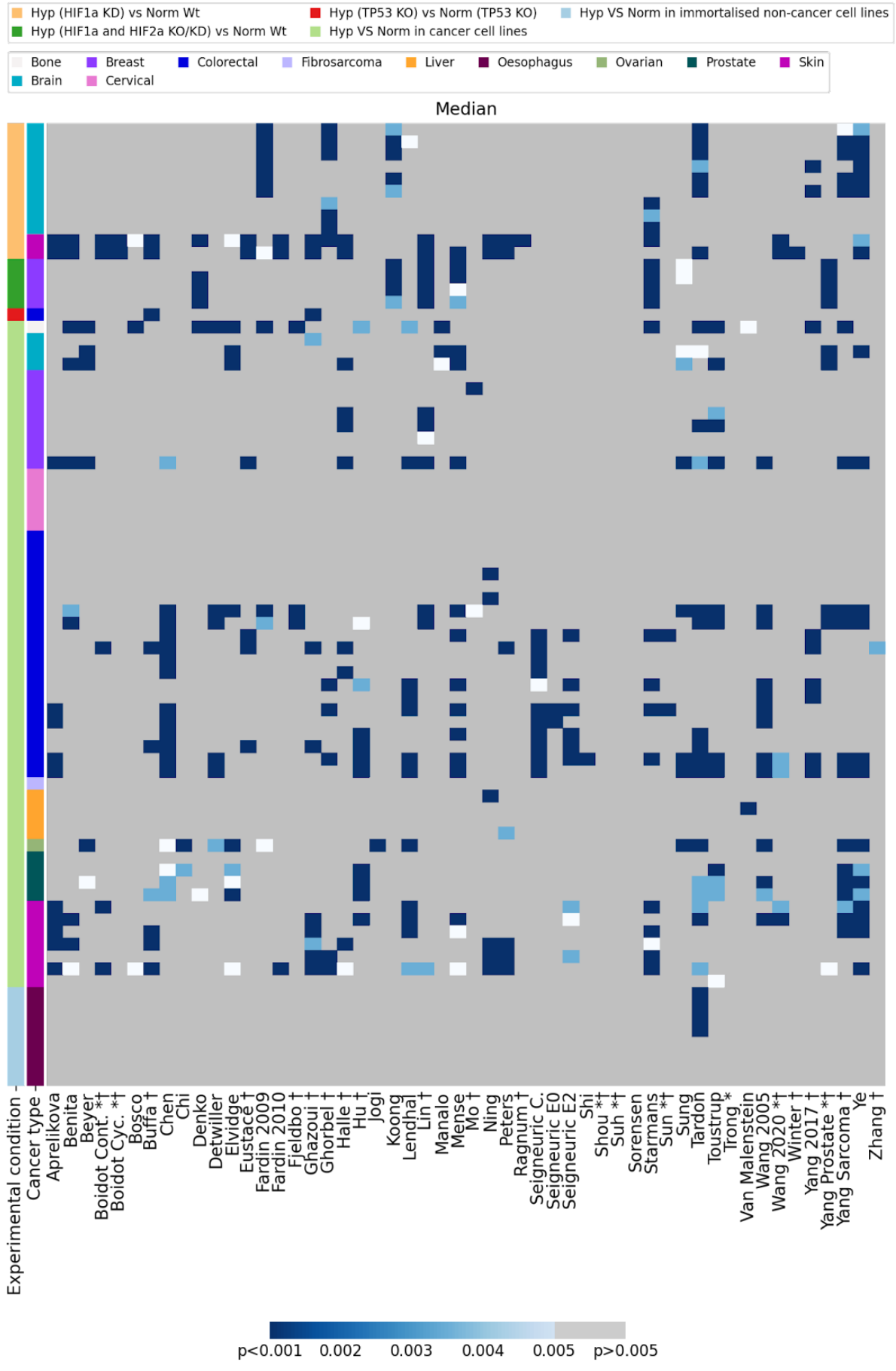

Hyp (HIF1a KD) vs Norm Wt
Hyp (HIF1a and HIF2a KO/KD) vs Norm Wt
Hyp (TP53 KO) vs Norm (TP53 KO)
Hyp VS Norm in cancer cell lines
Hyp VS Norm in immortalised non-cancer cell lines
Bone
Brain
Breast
Cervical
Colorectal
Fibrosarcoma
Liver
Oesophagus
Ovarian
Prostate
Skin
Median
Experimental condition
Cancer type
Aprelikova
Benita
Beyer
Boidot Cont. *†
Boidot Cyc. *†
Bosco
Buffa †
Chen
Chi
Denko
Detwiller
Elvidge
Eustace †
Fardin 2009
Fardin 2010
Fjeldbo †
Ghazoui †
Ghorbel †
Halle †
Hu †
Jogi
Koong
Lendhal
Lin †
Manalo
Mense
Mo †
Ning
Peters
Ragnum †
Seigneuric C.
Seigneuric E0
Seigneuric E2
Shi
Shou *†
Suh *†
Sorensen
Starmans
Sun *†
Sung
Tardon
Toustrup
Trong *
Van Malenstein
Wang 2005
Wang 2020 *†
Winter †
Yang 2017 †
Yang Prostate *†
Yang Sarcoma †
Ye
Zhang †
p<0.001
0.002
0.003
0.004
0.005
p>0.005

Figure 5.14: *Representation of all the pairs in the study that were non-significant using median score and Sorensen hypoxia signature.*

The heatmap represents where using the median score with Sorensen signature was ineffective. This was particularly the case in colorectal cancer (20 pairs). Conventions as in *Fig. 4.1*.

Chapter 6: Hypoxia signatures in clinical data

6.1 Exploring clinical data

As is common in science, transfer from the benchside to the clinic is a complex road. Many hurdles exist, not least the phenotypic and genetic differences seen between cancer cell line experiments in the laboratory and the complexity of the three dimensional tumour microenvironment in rodents, let alone humans. Many promising diagnostic and therapeutic advances have not traversed from bench to clinic because model systems are imperfect representations of humans as organisms. Indeed, to minimise and even avoid differences between model systems and humans, some of the signatures identified by the literature review have their roots in clinical samples rather than cellular models, e.g. Buffa.

With the obvious phenotypic differences between cellular models and human tumours, it is logical to check, and perhaps not wise to assume, that effective hypoxia signatures and scores that work well in cell lines also are the most appropriate in human tumours. However, this is not a simple task. Firstly, there are debates surrounding the most effective type of hypoxia measurements with the advances in modern technology, as perhaps the “gold standard” oxygen electrode has been surpassed and there are no large cohort studies across tumour types to establish the most effective method. Thus the common way to say if a hypoxia signature is effective is by looking at prognosis in clinical cohorts, i.e. if effective those patients with a higher score have a worse prognosis. Although useful, this does not prove the signatures are truly measuring hypoxia. At present there are no large datasets that have gene expression data, prognostic information as well as hypoxia measurements. This is an obviously important area for further grant applications and ongoing work.

The current study therefore is limited to the validation of hypoxia signature to datasets that have prognostic information only. However, for this task I chose one of the largest dataset globally: The Cancer Genome Atlas (TCGA). This not only includes profiling

of 20,000 primary cancers, it also includes matched normal adjacent tissue from the same patients, which is often omitted from most studies. Previous work looking at hypoxia signatures has focussed on primary tumour samples only. However, there are hints from the cell line work in this thesis that this might not be the most appropriate approach. So the goal is to minimise these potential errors by looking at the differences with a respective reference cohort of normal samples. This difference would have been missed if hypoxic samples are looked at in isolation. A similar phenomenon may well be true in clinical samples without comparison to a reference, e.g. a normal matched tissue sample. However, other studies of hypoxia signatures have looked at cancer samples alone, without comparison to presumably normoxic matched normal tissue samples.

This, combined with the results presented in Chapter 4 and 5, laid the foundation to construct a list of criteria of a likely effective signature/score combination in clinical samples, bearing in mind the limitation of the ability to correlate with other measures of oxygen tension. These are as follows:

If a hypoxia signature and score is working effectively in clinical samples you may expect it might:

- **Step 1**: have scores higher than in the normal tissue (as discussed above)
- **Step 2**: outperform RGS of the same length in simulations when comparing normal tissues and tumour tissues
- **Step 3**: show prognostic efficacy

Therefore, this list of three criteria will be taken to explore the all aforementioned signatures and scores to identify the likely most effective hypoxia signatures in clinical samples. It will be particularly interesting to see if the top performing signatures from the controlled experiments using cell lines also appear effective in the TGCA.

6.2 TCGA data cleaning

The make up of the TCGA dataset is discussed in detail in Methods. However, it is important to note that my analysis is focussed on:

1) solid tumours
    a) Hypoxia is unlikely in blood cancers by their very nature, particularly leukaemia
2) without VHL mutations
    a) Many hypoxia signatures rely on HIF-1a targets, which are upregulated in the absence of hypoxia when VHL is mutated. Therefore, one may expect hypoxia signatures to not perform as well if VHL is mutated, as seen in the cell line work. Therefore the following analysis excluded the tumour type in which this is common, renal cancer (similar to the cell line analysis).
3) with more than 30 samples
    a) This allowed enough power for survival and comparative analyses.

Fig. 6.1 details which tumour types were considered in the subsequent analyses.

a)

Study name	Cancer code
Breast invasive carcinoma	BRCA
Colon adenocarcinoma	COAD
Head and Neck squamous cell carcinoma	HNSC
Liver hepatocellular carcinoma	LIHC
Lung adenocarcinoma	LUAD
Lung squamous cell carcinoma	LUSC
Prostate adenocarcinoma	PRAD
Stomach adenocarcinoma	STAD
Thyroid carcinoma	THCA
Uterine Corpus Endometrial Carcinoma	UCEC

b)

Cancer code	Sample type	count
BRCA	NAT	114
	Tumour	1097
COAD	NAT	41
	Tumour	452
HNSC	NAT	44
	Tumour	520
LIHC	NAT	50
	Tumour	371
LUAD	NAT	59
	Tumour	515
LUSC	NAT	51
	Tumour	503
PRAD	NAT	52
	Tumour	497
STAD	NAT	35
	Tumour	415
THCA	NAT	59
	Tumour	505
UCEC	NAT	34
	Tumour	532

c)

Sample type	count
NAT	539
Tumour	5407

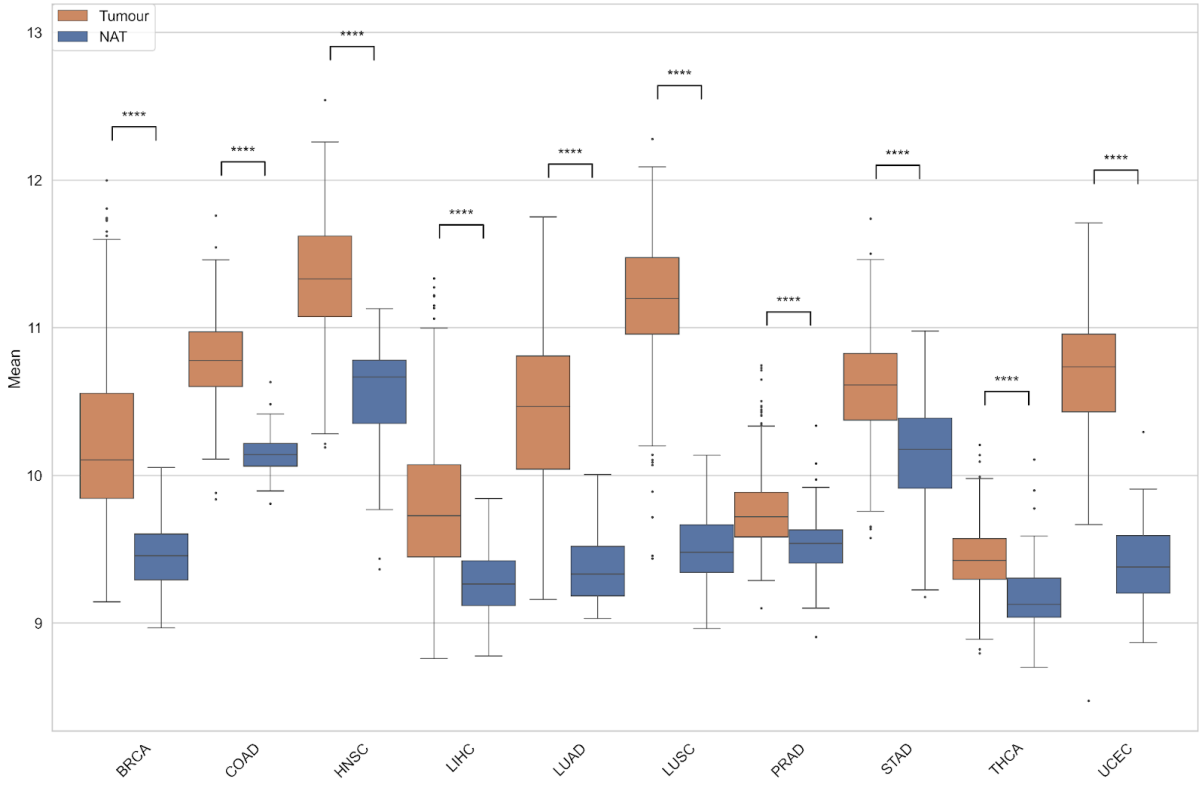
Figure 6.1: *TCGA cancer types considered in the hypoxia signature evaluation*

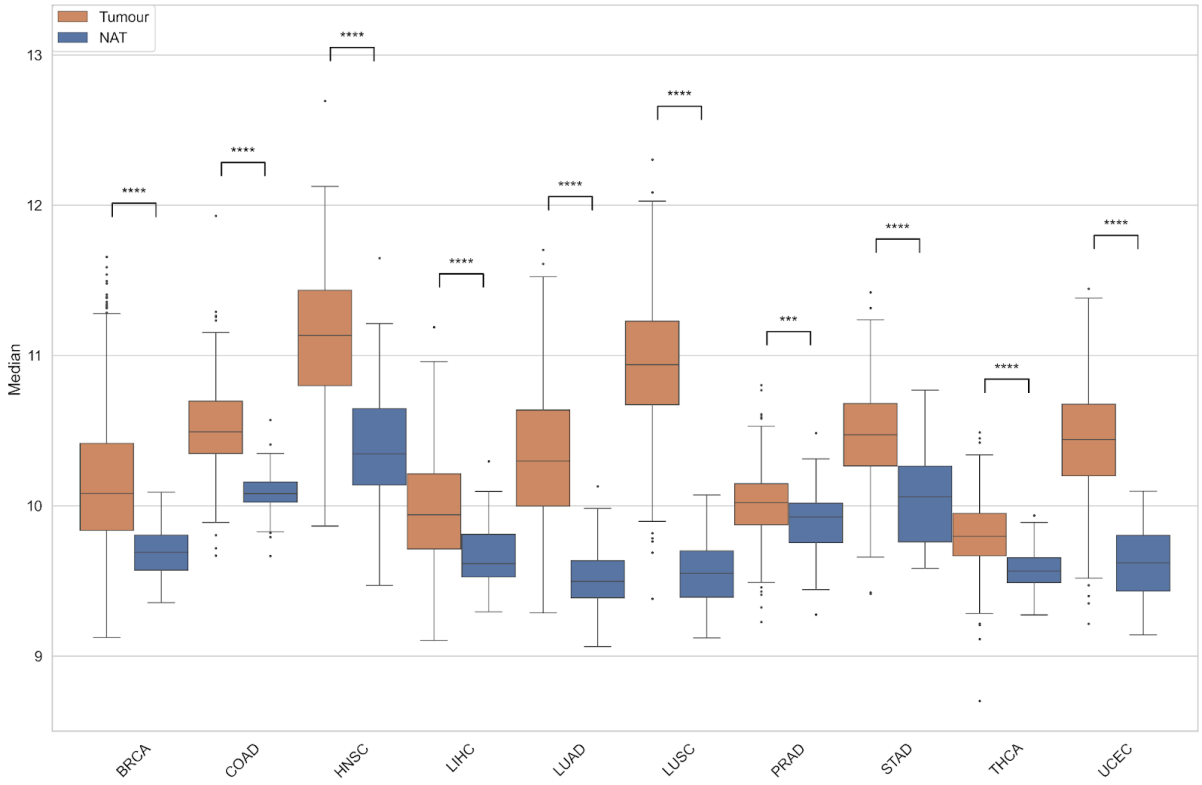
List of the 10 cancer types that met the criteria for analysis from the TGCA (a) including the count for both NAT and tumour tissues (b). The total number of NAT and tumour samples available in the study is reported in (c).

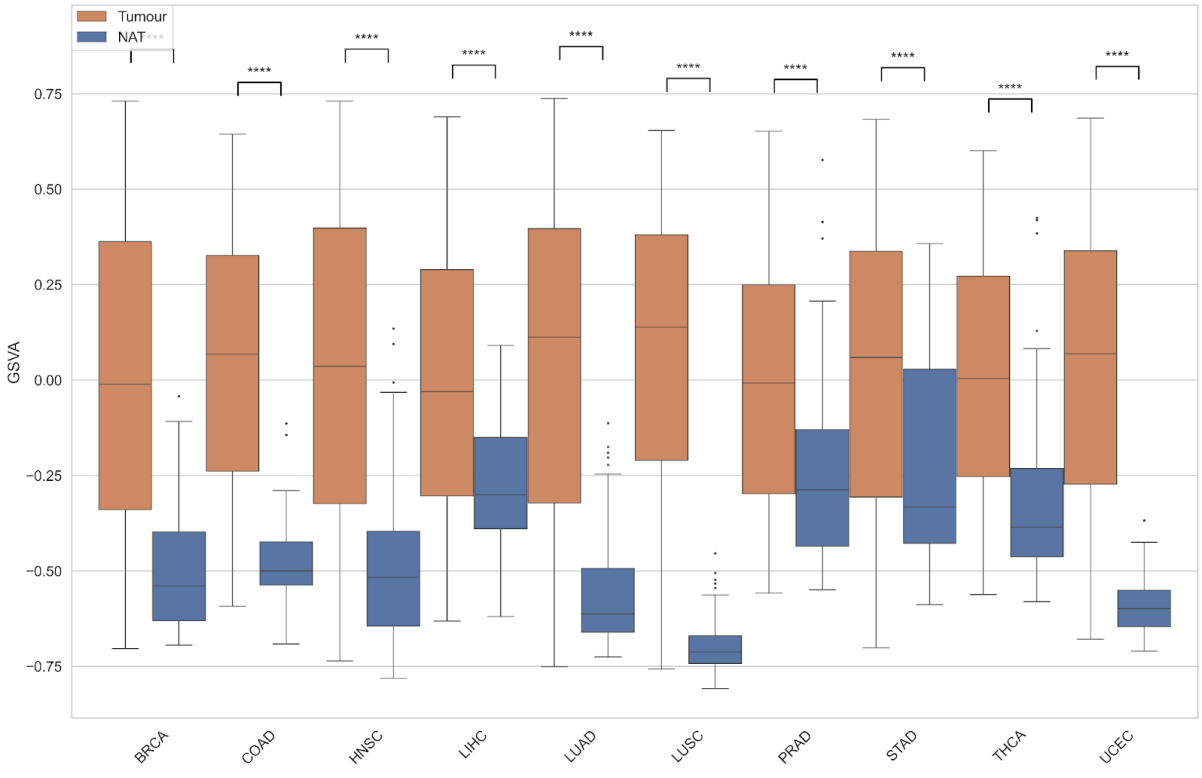
6.3 Evaluation Step 1: Do hypoxia signatures scores tend to be higher in tumours?

Below in *Fig. 6.2* one can see boxplots for the Buffa signature using the four scores (represented on the y-axis) in 10 cancer types (orange box) and NAT (blue box). One can see that using the mean, median, GSVA and ssGSEA there is a significant difference (using the Mann-Whitney-Wilcoxon test) in scores seen between the NAT and the tumours across all the cancer types.

This analysis was then expanded to all 53 signatures comparing NAT/tumour tissues across the 10 cancer types, considering the four scoring types (*Table 6.1a-d*). Using the mean, median, GSVA and ssGSEA scores, three signatures (Buffa, Ghazoui and Ragnum) showed significant differences in tumour vs NAT in the expected direction across the 10 tumour types (tumour higher hypoxia scores than NAT) using a threshold of $p<0.005$ (Bonferroni corrected, calculated as 0.05 divided by the number of cancer types). Therefore, when looking for a generalisable hypoxia signature across the 10 cancer types, Buffa, Ghazoui and Ragnum using all four scores (mean, median, GSVA and ssGSEA) appear the front runners.


Tumour
NAT
Mean
13
12
11
10
9
****
BRCA
COAD
HNSC
LIHC
LUAD
LUSC
PRAD
STAD
THCA
UCEC


Tumour
NAT
Median
13
12
11
10
9
****
***
BRCA
COAD
HNSC
LIHC
LUAD
LUSC
PRAD
STAD
THCA
UCEC


Tumour
NAT
GSVA
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
****
BRCA
COAD
HNSC
LIHC
LUAD
LUSC
PRAD
STAD
THCA
UCEC

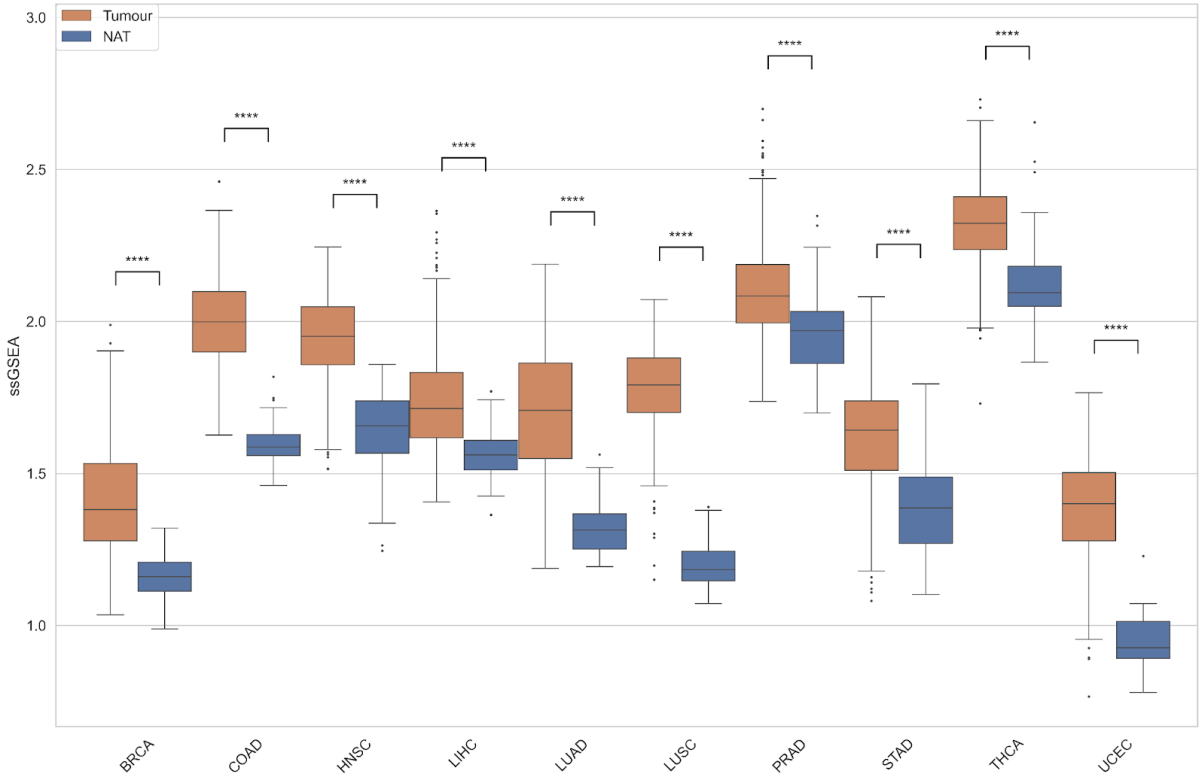

Tumour
NAT
ssGSEA
3.0
2.5
2.0
1.5
1.0
****
BRCA
COAD
HNSC
LIHC
LUAD
LUSC
PRAD
STAD
THCA
UCEC

Figure 6.2: *Comparison of Buffa signature hypoxia scores across ten TCGA cancer types*

Each box plot represents the distribution of the four hypoxia scores derived using Buffa signatures across the 10 cancer types under examination in this study. Orange and blue boxes represent hypoxia scores of tumour and NAT samples respectively. The p-values were calculated using the Mann-Whitney-Wilcoxon test (as per *Methods*, Chapter 2.6.1). TCGA cancer abbreviations found on the x-axis, as described in *Fig. 6.1(a)*. P-value annotation legend: ns: p >5.00e-02, *****: 1.00e-02 < p <= 5.00e-02, ******: 1.00e-03 < p <= 1.00e-02, *******: 1.00e-04 < p <= 1.00e-03, ********: p <= 1.00e-04

a) Median

	COAD	LUSC	HNSC	STAD	THCA	BRCA	PRAD	UCEC	LIHC	LUAD
Aprelikova	1.00E+00	2.46E-07	9.57E-03	7.49E-01	3.35E-04	1.00E+00	1.00E+00	1.67E-02	6.51E-01	2.13E-04
Benita	2.44E-05	9.43E-19	2.33E-06	9.75E-01	1.00E+00	1.00E+00	1.00E+00	9.55E-04	1.00E+00	2.50E-10
Beyer	1.00E+00	9.80E-01	1.60E-01	9.99E-01	8.74E-01	1.00E+00	1.00E+00	9.86E-01	1.00E+00	1.58E-02
Boidot Cont.	1.08E-05	5.52E-28	3.20E-07	1.95E-01	9.41E-01	2.84E-14	4.06E-07	6.13E-16	4.68E-07	4.89E-23
Boidot Cyc.	1.00E+00	1.28E-11	1.45E-02	8.08E-01	8.25E-07	4.20E-15	1.74E-03	5.26E-06	8.10E-05	1.65E-13
Bosco	6.02E-01	1.00E+00	7.40E-07	1.90E-03	8.52E-04	1.00E+00	1.00E+00	1.85E-01	1.00E+00	1.00E+00
Buffa	4.12E-24	5.73E-32	2.81E-22	2.06E-08	1.64E-13	7.77E-54	1.09E-09	2.90E-22	1.71E-14	1.94E-31
Chen	4.11E-10	5.49E-32	2.84E-18	7.27E-01	3.04E-07	2.82E-02	1.00E+00	2.67E-13	9.83E-10	4.33E-31
Chi	6.42E-01	3.24E-11	6.82E-10	3.24E-02	1.91E-03	1.00E+00	1.00E+00	3.13E-03	9.69E-01	2.93E-13
Denko	1.00E+00	6.45E-01	1.96E-05	8.11E-01	1.29E-03	3.59E-02	1.00E+00	1.14E-03	1.00E+00	3.80E-03
Detwiller	5.18E-06	7.76E-05	9.10E-17	7.50E-05	9.66E-09	1.00E+00	1.00E+00	1.00E+00	1.00E+00	6.26E-02
Elvidge	7.18E-01	8.02E-03	7.05E-04	7.39E-01	3.14E-03	1.00E+00	1.00E+00	9.28E-01	1.00E+00	1.03E-01
Eustace	9.75E-25	1.14E-31	1.95E-20	1.36E-05	1.46E-05	6.82E-05	1.00E+00	6.89E-21	2.94E-02	1.91E-29
Fardin 2009	2.48E-08	2.43E-31	1.03E-07	1.00E+00	9.08E-01	1.00E+00	9.99E-01	1.01E-10	1.00E+00	1.13E-29
Fardin 2010	9.99E-01	5.81E-28	9.79E-01	1.00E+00	9.54E-01	1.85E-07	9.53E-01	5.63E-19	2.17E-01	8.44E-24
Fjeldbo	4.76E-23	1.77E-23	2.84E-03	6.69E-05	5.93E-23	5.64E-03	1.00E+00	6.82E-17	4.77E-21	6.11E-12
Ghazoui	1.68E-24	5.08E-32	8.89E-23	2.52E-09	1.03E-04	7.18E-52	1.24E-07	3.62E-22	5.66E-09	7.27E-32
Ghorbel	1.00E+00	1.00E+00	1.00E+00	2.62E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Halle	3.68E-18	7.13E-31	6.94E-11	1.01E-02	6.81E-12	9.49E-13	9.86E-01	3.85E-22	3.11E-12	5.53E-23
Hu	1.00E+00	2.55E-24	9.79E-09	9.52E-01	2.30E-05	1.00E+00	3.66E-01	1.71E-13	8.18E-05	4.70E-07
Jogi	9.99E-01	9.77E-01	7.01E-07	4.34E-01	2.28E-03	1.00E+00	1.00E+00	9.46E-01	1.00E+00	9.90E-01
Koong	3.82E-13	1.05E-23	4.39E-18	6.20E-12	3.20E-07	1.83E-46	8.48E-01	8.22E-09	1.00E+00	7.98E-25
Lendhal	1.03E-07	1.00E-29	2.37E-10	6.09E-02	3.17E-02	9.44E-01	1.00E+00	1.22E-17	7.95E-01	4.18E-22
Lin	4.41E-05	2.98E-20	6.71E-04	1.17E-03	9.99E-01	1.98E-01	1.57E-01	7.13E-17	1.00E+00	1.61E-06
Manalo	1.00E+00	1.00E+00	1.71E-04	2.48E-01	7.66E-02	1.00E+00	1.00E+00	1.00E+00	9.96E-01	1.00E+00
Mense	3.46E-02	4.73E-04	6.30E-03	1.00E+00	4.31E-01	1.00E+00	1.00E+00	6.59E-01	1.00E+00	9.10E-01
Mo	5.57E-04	1.55E-29	4.05E-08	4.78E-01	1.83E-13	1.00E+00	1.00E+00	2.34E-11	2.60E-01	3.85E-27
Ning	1.27E-01	3.41E-10	2.32E-12	7.88E-04	3.47E-07	9.93E-01	1.00E+00	4.75E-01	3.26E-03	4.82E-10
Peters	1.82E-07	3.10E-18	1.07E-07	4.77E-03	3.41E-01	8.97E-06	8.86E-01	1.22E-04	3.85E-08	9.79E-17
Ragnum	1.80E-22	6.81E-32	2.18E-25	1.85E-15	1.89E-12	4.33E-59	5.02E-15	2.45E-22	4.56E-26	1.65E-33
Seigneuric C.	6.03E-03	1.20E-21	5.84E-24	6.52E-20	2.48E-05	2.29E-37	8.49E-05	8.43E-09	6.88E-02	6.69E-09
Seigneuric E0	1.00E+00	4.26E-04	6.09E-08	3.19E-02	5.01E-01	1.00E+00	9.98E-01	1.00E+00	5.07E-01	5.73E-01
Seigneuric E2	4.64E-12	4.67E-30	8.95E-26	1.80E-18	2.64E-05	7.63E-46	8.05E-03	2.41E-17	1.03E-07	7.81E-22
Shi	1.00E+00	2.65E-01	2.69E-08	4.02E-02	9.84E-01	1.27E-07	1.00E+00	2.41E-03	2.18E-02	1.59E-09
Shou	1.28E-14	1.22E-28	9.67E-01	5.73E-02	2.91E-01	4.26E-01	1.00E+00	7.12E-16	1.50E-04	2.85E-30
Suh	1.88E-01	3.49E-23	1.11E-04	3.01E-04	8.27E-01	1.00E+00	9.99E-01	5.60E-07	1.00E+00	1.53E-06
Sorensen	9.90E-01	3.63E-30	8.08E-04	1.00E+00	1.20E-09	1.00E+00	1.00E+00	2.63E-11	1.00E+00	7.14E-25
Starmans	6.49E-01	1.74E-02	7.31E-02	1.10E-01	9.98E-01	1.00E+00	1.00E+00	5.54E-03	1.00E+00	1.92E-10
Sun	2.46E-14	4.56E-05	4.71E-02	4.01E-02	5.11E-01	7.64E-08	9.98E-01	1.86E-19	3.75E-01	1.33E-13
Sung	4.36E-02	2.57E-06	7.04E-09	2.72E-01	1.01E-04	1.00E+00	1.00E+00	3.70E-04	1.00E+00	1.65E-03
Tardon	2.36E-13	1.15E-22	9.25E-13	8.87E-01	1.46E-01	1.00E+00	1.00E+00	1.84E-12	1.00E+00	4.79E-13
Toustrup	8.83E-01	5.73E-27	1.31E-08	1.00E+00	1.56E-09	1.00E+00	1.00E+00	5.81E-11	9.73E-01	2.96E-18
Trong	1.00E+00	6.52E-01	4.54E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Van Malenstein	9.72E-01	4.23E-29	6.27E-01	3.57E-01	2.70E-02	1.87E-15	9.94E-01	1.45E-12	1.35E-02	1.58E-30
Wang 2005	1.00E+00	2.12E-09	1.80E-01	9.94E-01	1.00E+00	1.00E+00	9.95E-01	1.86E-03	1.00E+00	3.24E-03
Wang 2020	1.00E+00	2.66E-01	2.01E-05	1.00E+00	4.16E-01	9.96E-01	9.28E-01	6.02E-01	1.00E+00	8.84E-01
Winter	2.95E-23	2.02E-31	8.81E-21	2.09E-09	5.57E-13	2.79E-22	1.00E+00	3.98E-20	5.79E-05	1.73E-29
Yang 2017	1.00E+00	1.00E+00	1.13E-06	8.58E-01	8.35E-06	1.00E+00	9.99E-01	1.00E+00	9.54E-01	1.00E+00
Yang Prostate	5.39E-01	1.00E+00	1.00E+00	9.98E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Yang Sarcoma	1.25E-02	2.23E-21	2.25E-06	1.00E+00	9.95E-01	1.00E+00	1.00E+00	1.09E-04	1.00E+00	6.69E-11
Ye	4.72E-07	6.49E-28	1.88E-09	9.91E-01	1.00E+00	1.00E+00	1.00E+00	1.06E-13	9.98E-01	2.82E-23
Zhang	1.63E-23	1.48E-31	5.93E-08	1.88E-09	3.83E-01	8.46E-56	3.64E-21	1.57E-20	3.44E-27	8.42E-32
Zou	5.48E-01	1.00E+00	4.15E-02	1.62E-01	6.89E-03	9.14E-01	1.00E+00	1.27E-01	1.00E+00	1.00E+00

b) Mean

	COAD	LUSC	HNSC	STAD	THCA	BRCA	PRAD	UCEC	LIHC	LUAD
Aprelikova	9.14E-01	6.35E-18	6.89E-03	1.17E-01	4.46E-06	1.00E+00	9.97E-01	3.60E-02	6.09E-01	1.07E-12
Benita	6.18E-04	5.21E-14	8.34E-06	4.08E-01	9.62E-01	1.00E+00	1.00E+00	3.16E-01	1.00E+00	2.78E-08
Beyer	1.00E+00	4.83E-01	3.62E-02	1.00E+00	8.32E-13	1.00E+00	1.00E+00	9.63E-01	1.00E+00	3.89E-02
Boidot Cont.	6.13E-07	1.46E-23	6.09E-08	3.98E-02	2.72E-03	2.40E-36	1.19E-04	3.57E-11	6.14E-04	8.51E-20
Boidot Cyc.	1.06E-08	2.82E-09	1.12E-07	3.82E-03	6.07E-05	9.16E-19	2.01E-07	1.20E-09	1.32E-02	5.48E-06
Bosco	9.94E-02	1.00E+00	5.42E-07	6.58E-05	1.81E-02	1.00E+00	1.00E+00	7.89E-01	1.00E+00	1.00E+00
Buffa	3.62E-19	7.27E-32	1.93E-17	6.66E-10	3.62E-15	2.42E-34	5.30E-04	1.73E-20	1.27E-09	1.84E-30
Chen	4.14E-11	5.52E-29	1.29E-16	5.39E-01	6.90E-23	8.22E-01	1.00E+00	5.13E-04	5.97E-15	8.90E-26
Chi	1.26E-01	2.57E-11	1.60E-07	1.16E-04	3.36E-08	8.65E-01	1.00E+00	2.81E-02	4.31E-01	1.37E-08
Denko	1.00E+00	9.02E-02	3.20E-07	9.55E-01	1.16E-08	9.71E-01	1.00E+00	1.89E-03	1.00E+00	7.56E-01
Detwiller	2.44E-02	2.70E-02	3.27E-18	1.44E-03	7.42E-10	1.00E+00	1.00E+00	1.00E+00	1.00E+00	4.93E-01
Elvidge	8.84E-01	1.89E-01	2.47E-02	7.68E-01	6.58E-10	1.00E+00	1.00E+00	9.90E-01	9.98E-01	1.50E-01
Eustace	1.23E-24	7.93E-31	9.71E-13	2.36E-09	2.29E-04	9.83E-01	1.00E+00	2.67E-12	7.78E-05	2.73E-22
Fardin 2009	1.17E-06	7.99E-29	1.68E-06	1.00E+00	5.92E-01	1.00E+00	9.81E-01	5.05E-10	1.00E+00	1.52E-21
Fardin 2010	1.00E+00	1.13E-25	1.00E+00	1.00E+00	4.88E-01	2.05E-18	1.15E-01	1.24E-16	8.53E-01	3.19E-18
Fjeldbo	8.68E-13	3.93E-17	6.51E-11	9.68E-01	1.74E-20	7.12E-08	1.00E+00	2.63E-11	6.36E-07	5.31E-05
Ghazoui	2.94E-25	4.92E-32	5.91E-23	9.44E-12	1.99E-04	1.04E-47	6.46E-14	1.99E-21	1.46E-06	6.87E-31
Ghorbel	9.15E-01	1.00E+00	9.97E-01	8.50E-02	1.00E+00	1.00E+00	9.93E-01	1.00E+00	1.00E+00	9.82E-01
Halle	9.23E-01	1.58E-19	1.24E-10	9.57E-01	6.85E-15	1.61E-06	1.00E+00	9.90E-13	3.79E-01	3.49E-09
Hu	1.00E+00	7.37E-15	5.80E-10	9.62E-01	2.02E-13	1.00E+00	9.99E-01	1.79E-14	4.34E-02	5.51E-02
Jogi	9.99E-01	1.00E+00	7.03E-04	6.74E-02	1.69E-10	1.00E+00	9.99E-01	9.98E-01	1.00E+00	1.00E+00
Koong	1.51E-07	4.58E-13	1.14E-12	3.15E-07	9.72E-06	6.60E-12	7.76E-01	7.51E-01	1.00E+00	5.74E-08
Lendhal	4.91E-06	1.64E-25	1.76E-05	6.18E-02	1.29E-01	2.88E-01	1.00E+00	2.42E-16	1.00E+00	4.68E-14
Lin	1.00E+00	8.76E-01	8.68E-01	9.99E-01	1.00E+00	1.00E+00	8.61E-01	6.03E-01	1.00E+00	3.12E-01
Manalo	1.00E+00	1.00E+00	3.98E-05	4.39E-01	3.06E-03	1.00E+00	1.00E+00	1.00E+00	5.14E-01	1.00E+00
Mense	9.36E-01	3.00E-06	4.08E-01	9.99E-01	4.08E-03	1.00E+00	1.00E+00	6.80E-01	1.00E+00	7.17E-01
Mo	1.10E-02	1.08E-30	5.46E-04	2.24E-03	4.63E-14	6.20E-01	1.00E+00	3.93E-12	5.76E-04	1.51E-29
Ning	2.78E-10	3.58E-15	1.92E-18	1.27E-07	1.56E-06	4.48E-02	7.63E-01	5.62E-02	2.21E-10	1.29E-13
Peters	2.48E-12	3.77E-29	7.32E-09	1.46E-05	1.08E-05	9.10E-13	6.61E-01	5.31E-14	5.50E-13	1.98E-27
Ragnum	2.15E-17	1.33E-31	2.47E-23	1.40E-16	6.08E-15	2.18E-53	4.75E-12	5.22E-22	1.21E-23	3.53E-22
Seigneuric C.	2.04E-05	1.57E-05	3.27E-09	1.24E-05	9.37E-01	8.32E-13	1.00E+00	9.64E-03	1.00E+00	1.93E-09
Seigneuric E0	1.00E+00	1.00E+00	7.36E-01	8.06E-02	9.43E-01	1.00E+00	1.00E+00	1.00E+00	6.09E-01	1.00E+00
Seigneuric E2	4.89E-06	9.32E-09	9.87E-17	1.04E-05	4.67E-10	1.05E-29	1.00E+00	1.12E-05	1.34E-01	6.23E-18
Shi	9.95E-01	1.00E+00	1.38E-09	1.91E-04	1.00E+00	1.65E-05	1.00E+00	3.25E-01	2.23E-03	5.62E-02
Shou	3.29E-02	3.77E-24	4.54E-01	9.88E-01	2.10E-04	9.54E-01	9.95E-01	4.33E-12	2.94E-01	4.26E-23
Suh	3.33E-08	1.94E-04	6.44E-02	4.94E-03	3.70E-01	9.99E-01	1.00E+00	4.08E-04	1.00E+00	7.43E-01
Sorensen	9.65E-01	1.18E-29	3.60E-03	2.56E-01	5.45E-19	2.16E-08	1.00E+00	2.42E-07	9.99E-01	5.81E-27
Starmans	6.64E-01	2.33E-02	1.28E-02	7.17E-01	1.00E+00	1.00E+00	9.96E-01	6.51E-01	1.00E+00	2.67E-10
Sun	2.08E-10	6.76E-10	4.54E-01	2.64E-03	9.98E-01	1.00E+00	1.00E+00	2.71E-11	3.52E-02	9.91E-04
Sung	9.94E-01	1.70E-06	2.46E-04	3.61E-01	8.17E-13	1.00E+00	1.00E+00	3.61E-02	1.00E+00	1.10E-03
Tardon	3.93E-06	2.00E-23	1.61E-07	8.24E-01	4.00E-13	1.00E+00	9.97E-01	1.55E-12	1.00E+00	1.16E-16
Toustrup	7.81E-02	2.39E-29	6.59E-07	3.20E-01	2.31E-21	2.84E-15	9.80E-01	8.06E-11	1.00E+00	1.53E-24
Trong	1.00E+00	1.00E+00	9.99E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Van Malenstein	8.47E-01	2.09E-25	7.98E-01	3.33E-02	1.85E-02	1.44E-10	9.92E-01	1.25E-10	8.50E-02	5.13E-30
Wang 2005	1.00E+00	1.04E-02	1.93E-02	1.00E+00	8.82E-01	4.29E-01	7.01E-01	2.93E-08	1.00E+00	1.60E-02
Wang 2020	9.69E-01	9.87E-01	6.96E-01	1.00E+00	1.53E-02	1.00E+00	9.96E-01	9.99E-01	1.00E+00	9.99E-01
Winter	3.11E-22	1.80E-31	4.61E-19	3.43E-11	4.50E-17	2.34E-22	9.98E-01	8.17E-17	1.41E-07	1.41E-31
Yang 2017	1.00E+00	9.40E-03	1.88E-13	3.47E-01	4.11E-15	1.00E+00	8.44E-01	1.00E+00	3.43E-01	9.38E-01
Yang Prostate	7.52E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Yang Sarcoma	1.07E-01	1.88E-20	1.39E-02	1.00E+00	4.43E-03	1.00E+00	1.00E+00	5.37E-03	1.00E+00	5.68E-07
Ye	1.60E-05	3.20E-26	3.68E-04	9.63E-01	9.80E-01	1.00E+00	9.94E-01	4.21E-10	1.00E+00	1.41E-14
Zhang	1.26E-10	3.34E-31	5.56E-07	1.85E-07	4.14E-01	7.20E-47	1.42E-17	1.03E-21	7.56E-27	6.43E-30
Zou	7.83E-09	1.00E+00	1.00E+00	2.08E-02	8.54E-12	3.23E-08	1.27E-06	9.64E-03	1.00E+00	1.00E+00

c) GSVA

	COAD	LUSC	HNSC	STAD	THCA	BRCA	PRAD	UCEC	LIHC	LUAD
Aprelikova	5.04E-02	2.77E-06	9.00E-03	9.96E-01	3.55E-20	1.00E+00	9.91E-01	3.23E-03	9.91E-01	6.31E-04
Benita	8.87E-06	1.88E-11	1.19E-03	9.95E-01	4.04E-01	8.45E-01	1.00E+00	3.19E-03	1.00E+00	2.22E-06
Beyer	1.00E+00	9.35E-01	2.50E-01	1.00E+00	1.79E-03	1.00E+00	1.00E+00	8.47E-01	1.00E+00	9.44E-01
Boidot Cont.	2.64E-14	2.31E-23	4.38E-06	6.55E-01	9.92E-03	1.30E-26	2.06E-08	6.96E-14	2.70E-07	1.30E-18
Boidot Cyc.	4.67E-02	2.79E-16	1.02E-02	9.91E-01	1.75E-10	1.03E-32	2.36E-07	6.76E-08	1.56E-05	1.49E-12
Bosco	6.71E-02	9.95E-01	1.97E-06	5.82E-02	3.03E-10	1.00E+00	1.00E+00	4.81E-02	1.00E+00	9.98E-01
Buffa	5.45E-20	1.02E-30	9.69E-16	7.10E-05	1.35E-14	4.74E-41	2.03E-07	7.01E-20	4.48E-08	1.18E-24
Chen	2.76E-14	2.54E-30	3.21E-14	8.70E-01	2.65E-08	7.07E-17	1.00E+00	8.59E-14	1.14E-11	2.04E-30
Chi	8.56E-05	9.95E-13	3.40E-11	9.15E-02	2.05E-10	1.00E+00	1.00E+00	1.29E-05	1.00E+00	2.85E-10
Denko	9.98E-01	9.48E-02	1.29E-02	1.00E+00	2.68E-04	1.29E-08	1.00E+00	7.77E-04	1.00E+00	1.77E-02
Detwiller	4.52E-09	6.05E-05	1.92E-14	3.40E-04	1.67E-15	9.99E-01	1.00E+00	9.76E-01	1.00E+00	2.25E-01
Elvidge	1.82E-01	5.25E-01	3.40E-03	9.92E-01	4.91E-08	1.00E+00	1.00E+00	8.82E-01	1.00E+00	9.65E-01
Eustace	1.93E-16	1.60E-29	4.77E-14	4.38E-03	2.67E-13	4.74E-26	9.99E-01	1.42E-18	6.44E-01	3.05E-27
Fardin 2009	1.03E-08	1.05E-26	2.24E-07	1.00E+00	2.76E-02	9.89E-01	5.85E-01	1.06E-08	1.00E+00	6.71E-27
Fardin 2010	9.97E-01	1.85E-19	9.99E-01	1.00E+00	7.95E-01	2.15E-16	3.51E-01	7.20E-17	9.51E-01	4.12E-15
Fjeldbo	2.45E-17	2.37E-18	2.11E-07	1.19E-02	1.17E-23	1.34E-14	1.00E+00	1.64E-11	2.71E-15	2.84E-09
Ghazoui	2.61E-21	1.89E-30	1.46E-15	4.64E-05	5.17E-06	1.06E-40	3.97E-09	1.40E-20	1.66E-05	3.16E-26
Ghorbel	4.05E-01	9.98E-01	1.00E+00	4.31E-01	1.00E+00	1.00E+00	9.57E-01	9.96E-01	1.00E+00	1.00E+00
Halle	2.84E-10	1.60E-21	6.55E-08	6.42E-01	1.19E-15	9.79E-20	3.77E-01	4.32E-17	2.01E-04	6.84E-18
Hu	5.46E-01	7.29E-17	2.78E-08	8.69E-01	3.79E-11	8.60E-01	6.75E-01	1.97E-13	2.54E-04	9.61E-04
Jogi	9.72E-01	9.98E-01	2.75E-04	9.56E-01	7.71E-19	1.00E+00	1.00E+00	4.75E-02	1.00E+00	1.00E+00
Koong	5.25E-14	1.05E-17	4.03E-14	2.09E-05	2.79E-07	3.94E-29	4.07E-01	1.41E-04	1.00E+00	3.31E-19
Lendhal	1.59E-07	1.57E-21	4.72E-06	7.26E-01	5.18E-02	1.41E-01	1.00E+00	2.09E-12	1.00E+00	9.65E-14
Lin	7.02E-06	3.64E-11	1.36E-02	2.29E-01	1.02E-05	3.43E-07	1.83E-03	1.07E-11	1.00E+00	1.99E-08
Manalo	9.97E-01	1.00E+00	2.43E-05	3.83E-01	4.00E-10	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Mense	1.07E-03	2.37E-03	5.93E-02	1.00E+00	8.26E-02	1.00E+00	1.00E+00	3.78E-01	1.00E+00	8.09E-01
Mo	2.33E-04	3.27E-25	2.20E-06	2.58E-01	1.05E-13	5.80E-01	1.00E+00	3.37E-11	2.79E-01	2.20E-24
Ning	9.67E-10	1.25E-16	2.14E-05	3.04E-01	2.64E-15	6.88E-09	1.40E-02	3.59E-05	1.30E-03	8.51E-12
Peters	4.09E-16	6.16E-15	2.56E-03	5.53E-01	2.46E-10	2.11E-18	3.18E-03	1.24E-06	2.67E-07	4.39E-11
Ragnum	6.84E-21	4.76E-31	4.85E-24	8.46E-11	2.33E-19	1.44E-54	9.03E-13	3.73E-22	4.93E-24	3.05E-28
Seigneuric C.	2.85E-04	7.73E-11	1.83E-18	1.24E-08	3.16E-13	9.57E-31	1.02E-02	1.53E-05	5.81E-01	1.12E-01
Seigneuric E0	9.96E-01	7.79E-01	8.43E-05	7.91E-01	1.51E-03	9.98E-01	7.26E-01	9.57E-01	9.95E-01	1.00E+00
Seigneuric E2	8.45E-16	8.52E-14	6.50E-23	1.50E-05	1.16E-10	1.14E-44	2.79E-03	9.50E-10	3.60E-02	1.41E-08
Shi	8.48E-01	5.48E-03	9.42E-09	6.98E-01	8.09E-11	3.62E-54	3.18E-01	1.07E-12	4.33E-08	2.86E-22
Shou	1.21E-13	6.75E-26	8.88E-01	1.72E-01	7.74E-05	1.79E-03	1.00E+00	5.27E-15	7.33E-04	1.13E-20
Suh	2.04E-11	3.09E-12	6.49E-04	6.64E-06	3.98E-05	9.03E-02	9.42E-01	1.90E-07	1.00E+00	2.32E-02
Sorensen	9.36E-01	4.85E-22	6.71E-02	1.00E+00	1.24E-11	9.97E-01	1.00E+00	8.96E-08	1.00E+00	7.85E-17
Starmans	1.80E-02	1.09E-02	1.31E-01	9.32E-01	9.97E-01	5.56E-01	8.66E-01	7.93E-02	1.00E+00	2.07E-06
Sun	3.53E-01	9.28E-04	9.55E-01	4.22E-01	1.00E+00	1.16E-12	7.72E-01	3.64E-16	8.15E-01	3.05E-14
Sung	1.44E-01	6.99E-02	1.43E-03	9.96E-01	1.78E-12	1.00E+00	1.00E+00	3.14E-02	1.00E+00	5.18E-01
Tardon	1.67E-09	1.97E-13	4.68E-06	9.80E-01	3.05E-03	9.17E-01	1.00E+00	1.08E-08	1.00E+00	7.06E-07
Toustrup	8.18E-01	6.91E-17	1.43E-05	1.00E+00	4.02E-12	8.02E-03	9.99E-01	8.69E-09	1.00E+00	9.65E-12
Trong	1.00E+00	3.62E-07	1.14E-02	1.00E+00	1.37E-01	1.00E+00	9.99E-01	1.84E-01	1.00E+00	1.16E-03
Van Malenstein	2.08E-01	1.48E-24	9.82E-01	4.10E-01	4.07E-02	1.51E-18	5.54E-02	3.55E-10	9.31E-03	9.64E-30
Wang 2005	9.75E-01	2.12E-05	1.43E-01	1.00E+00	1.10E-01	5.23E-06	4.02E-01	8.80E-08	1.00E+00	5.21E-03
Wang 2020	1.00E+00	1.12E-02	3.28E-03	1.00E+00	5.09E-02	1.84E-01	4.50E-01	2.85E-01	1.00E+00	4.46E-01
Winter	9.04E-21	4.47E-28	5.65E-15	9.95E-06	5.37E-19	1.95E-42	2.14E-01	6.33E-19	1.25E-07	4.79E-22
Yang 2017	9.99E-01	1.00E+00	6.48E-08	7.74E-01	7.24E-08	1.00E+00	9.96E-01	1.00E+00	1.00E+00	1.00E+00
Yang Prostate	3.88E-01	1.00E+00	1.00E+00	1.00E+00	9.94E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Yang Sarcoma	5.58E-03	1.20E-09	2.37E-03	1.00E+00	7.24E-02	1.00E+00	1.00E+00	3.10E-04	1.00E+00	3.66E-04
Ye	2.30E-05	4.77E-18	4.81E-04	9.98E-01	9.49E-01	5.80E-01	1.00E+00	1.02E-08	1.00E+00	4.17E-13
Zhang	3.23E-19	5.75E-28	1.43E-07	3.99E-07	2.95E-02	1.32E-50	1.71E-20	3.09E-19	2.73E-27	1.42E-25
Zou	2.33E-02	8.12E-01	2.62E-02	7.42E-02	1.56E-11	3.09E-01	8.88E-01	9.27E-07	1.00E+00	1.00E+00

d) ssGSEA

	COAD	LUSC	HNSC	STAD	THCA	BRCA	PRAD	UCEC	LIHC	LUAD
Aprelikova	1.70E-01	7.85E-13	4.23E-04	9.70E-01	2.04E-21	1.00E+00	1.00E+00	1.54E-04	9.87E-01	3.08E-06
Benita	2.79E-07	2.14E-14	1.98E-05	9.99E-01	8.84E-01	1.00E+00	1.00E+00	3.12E-01	1.00E+00	5.88E-05
Beyer	1.00E+00	9.72E-01	2.23E-02	1.00E+00	6.85E-07	1.00E+00	1.00E+00	9.91E-01	1.00E+00	2.99E-01
Boidot Cont.	8.34E-12	4.20E-28	7.43E-08	7.60E-01	7.64E-03	1.92E-31	2.25E-08	1.35E-13	5.64E-06	1.08E-23
Boidot Cyc.	6.95E-03	1.38E-17	1.50E-04	8.06E-01	8.54E-12	5.62E-41	3.18E-08	3.57E-11	7.14E-04	2.88E-12
Bosco	2.82E-02	1.00E+00	3.95E-09	8.36E-03	5.63E-07	1.00E+00	1.00E+00	5.87E-01	1.00E+00	1.00E+00
Buffa	5.89E-26	6.25E-32	8.24E-23	1.33E-09	2.60E-20	1.73E-53	5.60E-10	9.27E-22	1.92E-13	1.51E-29
Chen	2.97E-15	1.16E-31	5.96E-23	4.43E-01	1.63E-25	1.35E-01	1.00E+00	4.56E-11	2.06E-07	1.46E-30
Chi	1.55E-02	9.69E-13	2.51E-13	5.93E-02	3.01E-11	1.00E+00	1.00E+00	1.89E-03	9.99E-01	1.63E-12
Denko	1.00E+00	6.01E-02	8.97E-05	9.89E-01	3.44E-06	2.39E-01	1.00E+00	1.84E-04	1.00E+00	5.09E-02
Detwiller	3.48E-05	4.11E-07	4.74E-18	9.27E-05	7.86E-10	1.00E+00	1.00E+00	1.00E+00	1.00E+00	7.86E-02
Elvidge	1.53E-01	5.16E-02	2.33E-04	9.68E-01	8.38E-09	1.00E+00	1.00E+00	9.46E-01	1.00E+00	3.28E-01
Eustace	3.17E-26	1.84E-31	3.63E-20	1.62E-06	1.98E-16	7.35E-04	1.00E+00	5.80E-19	9.02E-04	2.92E-27
Fardin 2009	1.73E-13	5.24E-31	2.44E-07	1.00E+00	9.29E-05	1.00E+00	9.52E-01	3.00E-13	1.00E+00	2.31E-29
Fardin 2010	9.99E-01	4.60E-28	9.79E-01	1.00E+00	5.94E-01	1.43E-15	3.88E-01	2.83E-19	9.24E-01	2.04E-22
Fjeldbo	1.74E-20	3.88E-22	4.45E-08	4.95E-02	1.82E-28	4.15E-12	1.00E+00	5.74E-15	7.65E-20	1.94E-16
Ghazoui	7.78E-26	5.49E-32	1.09E-22	6.06E-09	2.07E-05	2.17E-54	2.10E-10	7.52E-22	2.48E-09	2.40E-30
Ghorbel	9.87E-01	1.00E+00	1.00E+00	3.71E-01	1.00E+00	1.00E+00	9.98E-01	1.00E+00	1.00E+00	1.00E+00
Halle	7.05E-14	6.12E-29	1.69E-13	8.80E-01	7.78E-17	1.21E-18	1.35E-01	1.52E-20	6.67E-05	5.11E-20
Hu	4.53E-01	4.00E-16	1.06E-06	8.93E-01	1.94E-16	1.00E+00	9.91E-01	8.78E-16	3.64E-02	1.13E-02
Jogi	9.93E-01	1.00E+00	2.66E-08	3.32E-01	8.29E-17	1.00E+00	1.00E+00	9.98E-01	1.00E+00	1.00E+00
Koong	1.08E-12	3.52E-23	9.31E-18	4.10E-09	1.79E-06	4.38E-39	8.88E-01	1.17E-01	1.00E+00	3.12E-19
Lendhal	1.90E-13	1.87E-28	5.26E-10	5.18E-01	1.32E-05	7.73E-01	1.00E+00	9.44E-17	1.00E+00	7.77E-19
Lin	2.42E-10	1.27E-02	9.43E-05	2.49E-06	4.80E-04	6.76E-01	3.13E-03	2.69E-16	9.95E-01	1.00E+00
Manalo	1.00E+00	1.00E+00	5.42E-07	6.68E-01	8.91E-06	1.00E+00	1.00E+00	1.00E+00	9.99E-01	1.00E+00
Mense	2.16E-04	4.88E-05	1.38E-02	1.00E+00	9.65E-03	1.00E+00	1.00E+00	2.86E-01	1.00E+00	8.94E-01
Mo	7.47E-06	1.40E-28	4.31E-03	1.13E-01	4.00E-15	1.00E+00	1.00E+00	1.57E-11	1.60E-01	3.82E-29
Ning	6.05E-02	8.33E-05	6.97E-19	8.05E-03	5.30E-14	5.40E-01	1.00E+00	1.75E-01	3.39E-03	8.78E-06
Peters	1.14E-18	1.00E-28	3.12E-15	5.10E-04	5.39E-03	5.92E-18	9.49E-01	2.34E-09	4.83E-16	1.01E-25
Ragnum	6.40E-24	9.94E-32	2.85E-25	3.79E-15	4.77E-23	1.62E-61	8.53E-16	3.98E-22	4.05E-26	4.54E-31
Seigneuric C.	1.75E-10	1.07E-17	1.54E-19	2.54E-13	4.62E-10	2.23E-49	3.56E-02	2.26E-11	1.48E-01	4.89E-10
Seigneuric E0	4.49E-01	1.00E+00	1.86E-02	6.27E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.49E-01	1.00E+00
Seigneuric E2	1.35E-14	3.77E-29	1.71E-24	1.09E-12	2.97E-08	2.90E-55	8.50E-01	3.81E-16	2.03E-07	2.62E-24
Shi	1.00E+00	9.57E-01	1.80E-15	7.01E-02	9.82E-01	1.44E-32	1.00E+00	1.42E-09	1.39E-03	1.03E-08
Shou	8.97E-01	5.35E-29	4.94E-01	7.74E-01	9.84E-01	1.01E-11	1.00E+00	4.05E-12	6.12E-05	1.44E-29
Suh	2.09E-02	1.81E-21	5.50E-07	2.86E-01	2.80E-01	9.88E-01	9.99E-01	1.28E-05	1.00E+00	1.06E-04
Sorensen	9.36E-01	1.66E-30	5.92E-04	1.00E+00	8.60E-16	1.00E+00	1.00E+00	7.94E-11	1.00E+00	9.75E-27
Starmans	2.45E-02	4.05E-04	3.06E-02	9.73E-01	9.99E-01	1.00E+00	9.90E-01	1.69E-01	1.00E+00	1.97E-09
Sun	5.75E-05	7.75E-29	4.84E-05	7.61E-01	8.92E-04	4.31E-26	1.70E-01	6.80E-20	7.12E-01	8.69E-25
Sung	9.98E-03	5.49E-09	4.20E-09	7.27E-01	7.38E-12	1.00E+00	1.00E+00	5.03E-04	1.00E+00	1.40E-03
Tardon	3.90E-17	1.54E-22	2.59E-09	9.15E-01	1.52E-11	1.00E+00	1.00E+00	1.95E-13	1.00E+00	2.54E-11
Toustrup	3.42E-01	8.10E-28	4.85E-09	9.99E-01	6.44E-21	4.02E-01	9.99E-01	5.33E-11	1.00E+00	1.74E-21
Trong	1.00E+00	1.00E+00	1.21E-01	1.00E+00	1.00E+00	1.00E+00	9.99E-01	4.43E-02	1.00E+00	1.00E+00
Van Malenstein	9.75E-01	3.07E-29	6.36E-01	5.71E-01	1.54E-03	5.82E-18	9.97E-01	1.05E-12	8.48E-03	7.06E-32
Wang 2005	1.00E+00	3.60E-07	3.65E-02	1.00E+00	5.79E-01	4.88E-01	6.84E-01	1.25E-10	1.00E+00	1.02E-04
Wang 2020	1.00E+00	9.52E-01	1.96E-03	1.00E+00	2.01E-03	9.97E-01	9.99E-01	7.55E-01	1.00E+00	9.93E-01
Winter	7.32E-26	1.82E-31	5.56E-20	1.61E-08	1.36E-22	1.42E-40	9.91E-01	1.56E-19	1.83E-09	1.82E-30
Yang 2017	1.00E+00	1.00E+00	2.19E-08	8.84E-01	1.62E-13	1.00E+00	9.98E-01	1.00E+00	9.91E-01	1.00E+00
Yang Prostate	3.23E-01	1.00E+00	9.99E-01	9.98E-01	9.99E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Yang Sarcoma	9.33E-05	7.85E-21	5.24E-05	1.00E+00	5.68E-01	1.00E+00	1.00E+00	2.81E-04	1.00E+00	9.56E-09
Ye	2.12E-10	4.25E-28	6.10E-07	9.97E-01	3.43E-01	1.00E+00	1.00E+00	3.64E-12	1.00E+00	6.88E-20
Zhang	2.28E-22	1.27E-31	7.96E-10	8.25E-09	9.22E-02	2.03E-63	2.14E-21	7.74E-17	1.91E-28	9.51E-31
Zou	1.18E-03	1.00E+00	5.00E-01	2.99E-01	1.76E-13	1.23E-09	1.43E-03	3.24E-07	1.00E+00	1.00E+00

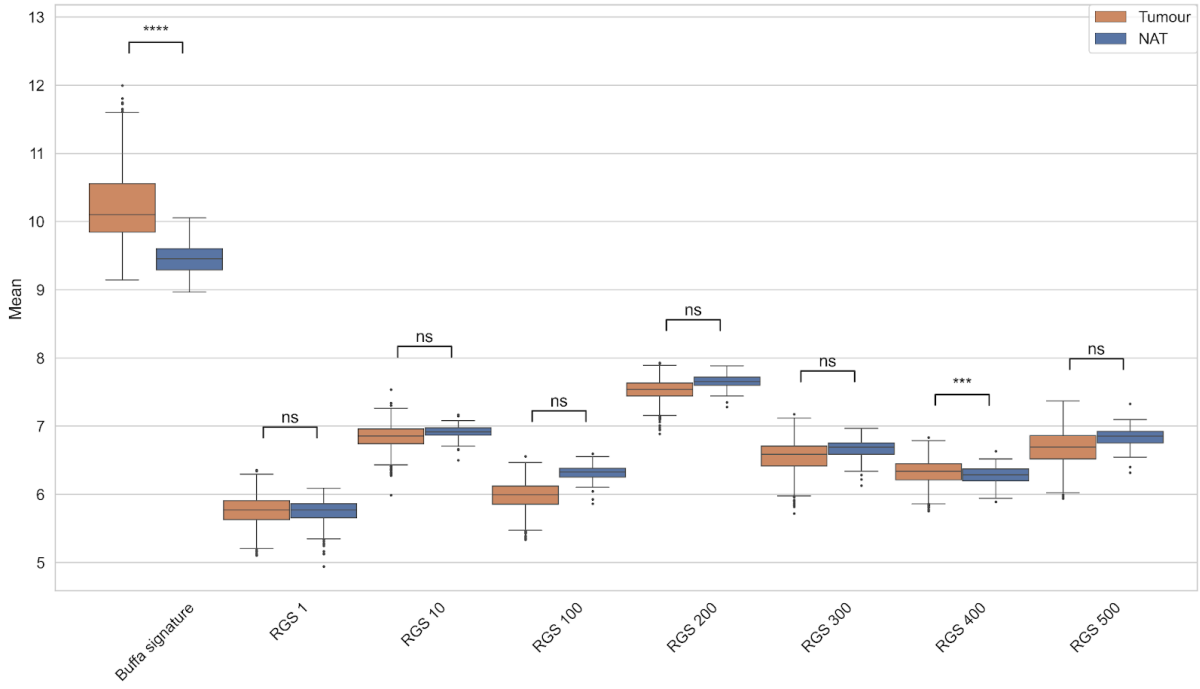
Table 6.1: *Summary of the p-values calculated using the Mann-Whitney-Wilcoxon test across ten cancer types*

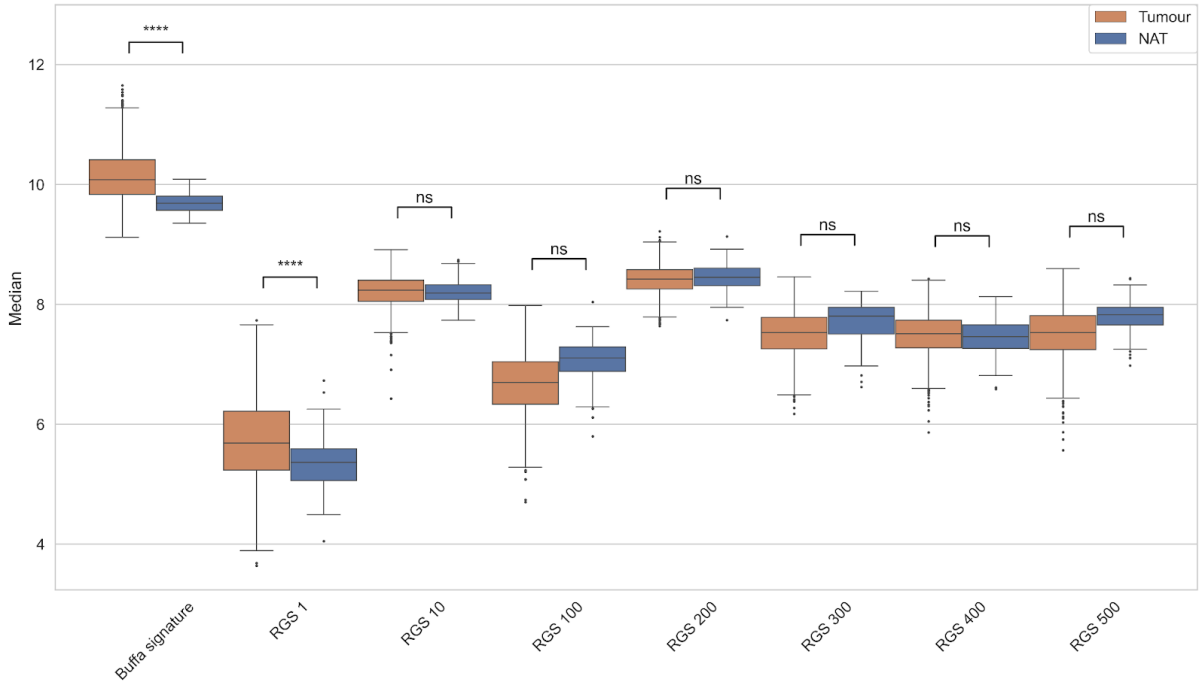
Comparison of the p-values calculated using the Mann-Whitney-Wilcoxon test for each cancer type in this study. Values highlighted in orange have $p < 0.005$. Values highlighted in red represent the signature with the lowest p-value on each cancer type. Panel A = median, panel B = mean, panel C = GSVA and panel D = ssGSEA.

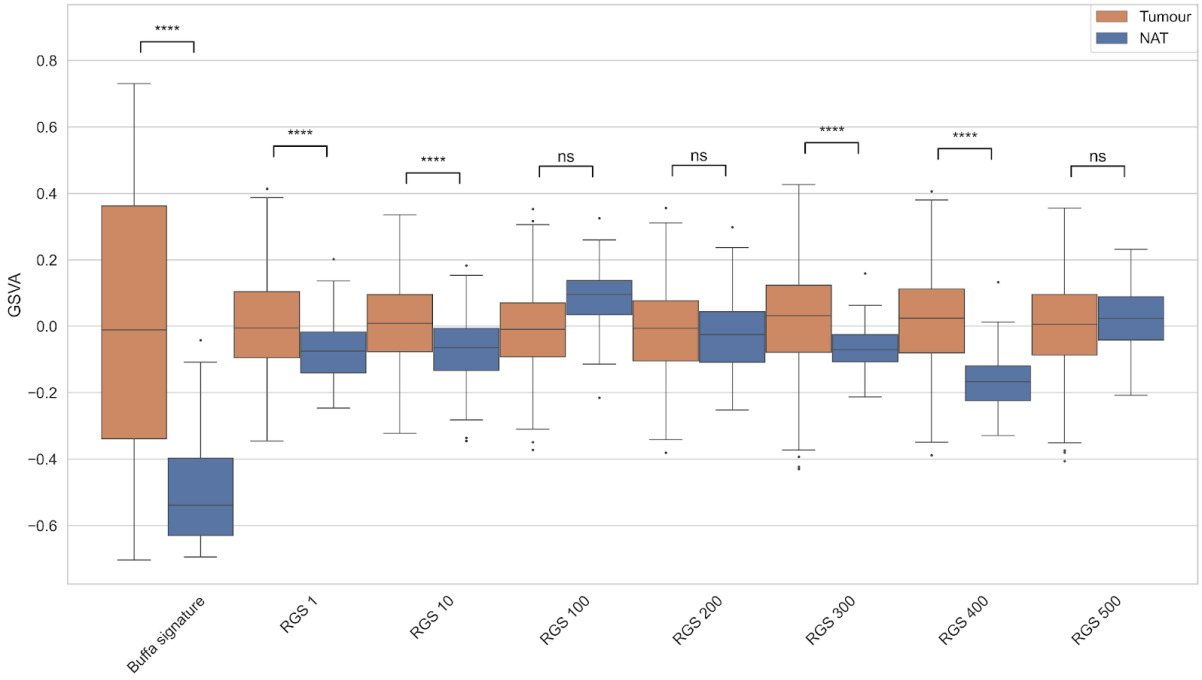
6.4 Evaluation Step 2: Do hypoxia signatures scores tend to outperform random gene signatures of the same length in simulations when comparing matched normal tissues and tumour tissues?

As in the cell lines, one may expect if a signature is actually defining a feature, such as hypoxia, that the tumour samples may have significantly different scores between tumour and normal tissues when compared to random gene sets (RGS) of the same length. An example of this point can be illustrated using the Buffa signature in breast cancer (BRCA) from the TCGA.

In *Fig. 6.3*, Buffa signature outperforms seven RGS in both median and mean scores as opposed to GSVA and ssGSEA when comparing their p-values (threshold $p<0.005$). Finally, a Signature Performance Index (SPI) was defined by looking at the percentage of times each signature and score in tumour vs normal differed in performance from a RGS of the same length, across all tissue types. Using the mean score, the top performers were Ragnum and Buffa by our SPI. These two signatures were different from RGS in 99.66% and 99.22% of the time respectively. Ragnum, Buffa and Ghazoui as the top signatures using GSVA (99.98%, 94.48% and 92.78%) and ssGSEA (88.80%, 75.98% and 73.38%). Zhang signature also reached 87.84% using ssGSEA, however, it failed the Evaluation Step 1. Finally, Ragnum (99.18%), Winter (88.40%) and Buffa (88.26%) are the top performers on this criterion using the median score. This work suggests Ragnum (mean, median & GSVA) or Buffa (mean & GSVA) and Ghazoui (GSVA) might be the most promising signatures for clinical use. To give extra confidence to using these signatures, sigQC metrics (as discussed in Methods and Chapter 3.3) appeared consistent with high performing signatures (*Supplementary figures S1-3* detail performance in BRCA cohort as an example, *Appendix 3*).


Tumour
NAT
Mean
13
12
11
10
9
8
7
6
5
****
ns
ns
ns
ns
ns
***
ns
Buffa signature
RGS 1
RGS 10
RGS 100
RGS 200
RGS 300
RGS 400
RGS 500


Tumour
NAT
Median
12
10
8
6
4
****
****
ns
ns
ns
ns
ns
ns
ns
Buffa signature
RGS 1
RGS 10
RGS 100
RGS 200
RGS 300
RGS 400
RGS 500


Tumour
NAT
GSVA
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
****
****
****
ns
ns
****
****
ns
Buffa signature
RGS 1
RGS 10
RGS 100
RGS 200
RGS 300
RGS 400
RGS 500

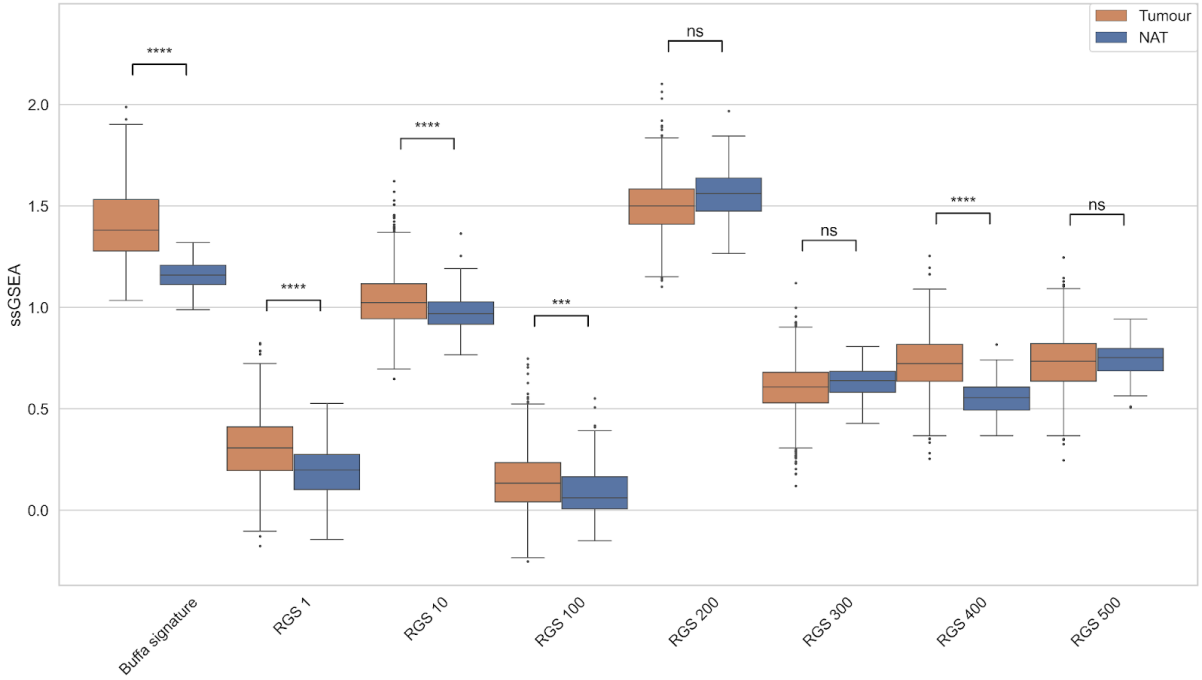

Tumour
NAT
ssGSEA
2.0
1.5
1.0
0.5
0.0
****
****
****
***
ns
ns
****
ns
Buffa signature
RGS 1
RGS 10
RGS 100
RGS 200
RGS 300
RGS 400
RGS 500

Figure 6.3: *Buffa signature and seven RGS scores for the BRCA dataset*

Each box plot represents the distribution of the four hypoxia scores (y axis) derived using Buffa signatures on the BRCA dataset. Each pair of orange and blue boxes represent hypoxia scores from tumour and NAT respectively for the original signatures and seven other RGS. The p-values were calculated using the Mann-Whitney-Wilcoxon test (as per *Methods*, Chapter 2.6.2). P-value annotation legend: ns: p >5.00e-02, *: 1.00e-02 < p <= 5.00e-02, **: 1.00e-03 < p <= 1.00e-02, ***: 1.00e-04 < p <= 1.00e-03, ****: p <= 1.00e-04.

a) Mean

	COAD	LUSC	HNSC	STAD	THCA	BRCA	PRAD	UCEC	LIHC	LUAD	Mean percentage
Aprelikova	0.00	97.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.76
Benita	97.60	98.60	97.20	0.00	0.00	0.00	0.00	0.00	0.00	93.20	38.66
Beyer	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Boidot Cont.	99.60	100.00	99.60	0.00	0.00	99.60	100.00	100.00	92.40	99.20	79.04
Boidot Cyc.	0.00	95.80	0.00	0.00	98.20	100.00	0.00	98.00	86.60	96.40	57.50
Bosco	0.00	0.00	99.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.98
Buffa	100.00	100.00	100.00	95.40	99.00	100.00	99.20	100.00	98.60	100.00	99.22
Chen	93.80	100.00	99.60	0.00	91.40	0.00	0.00	97.80	90.80	99.20	67.26
Chi	0.00	96.00	99.80	0.00	0.00	0.00	0.00	0.00	0.00	98.20	29.40
Denko	0.00	0.00	97.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.72
Detwiller	94.80	81.20	99.60	84.80	95.80	0.00	0.00	0.00	0.00	0.00	45.62
Elvidge	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Eustace	100.00	100.00	100.00	86.00	87.80	87.00	0.00	100.00	0.00	99.60	76.04
Fardin 2009	87.40	99.80	91.00	0.00	0.00	0.00	0.00	92.20	0.00	97.80	46.82
Fardin 2010	0.00	99.40	0.00	0.00	0.00	94.80	0.00	99.80	0.00	98.80	39.28
Fjeldbo	99.00	94.00	0.00	78.40	98.60	0.00	0.00	98.80	99.00	83.60	65.14
Ghazoui	100.00	100.00	100.00	98.40	0.00	100.00	99.00	100.00	95.60	100.00	89.30
Ghorbel	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Halle	100.00	100.00	98.40	0.00	97.80	95.40	0.00	100.00	94.60	98.40	78.46
Hu	0.00	94.80	91.40	0.00	82.80	0.00	0.00	97.80	76.80	78.60	52.22
Jogi	0.00	0.00	99.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.94
Koong	95.00	94.20	98.20	97.00	87.60	98.00	0.00	92.60	0.00	96.80	75.94
Lendhal	94.60	99.80	97.60	0.00	0.00	0.00	0.00	100.00	0.00	95.20	48.72
Lin	80.00	91.60	0.00	0.00	0.00	0.00	0.00	97.40	0.00	72.60	34.16
Manalo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mense	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mo	0.00	97.80	89.80	0.00	92.20	0.00	0.00	90.60	0.00	97.00	46.74
Ning	0.00	95.80	99.80	0.00	99.00	0.00	0.00	0.00	0.00	94.60	38.92
Peters	100.00	99.40	100.00	0.00	0.00	99.80	0.00	0.00	99.20	99.60	59.80
Ragnum	100.00	100.00	100.00	99.20	98.00	100.00	99.40	100.00	100.00	100.00	99.66
Seigneuric C.	0.00	92.80	100.00	99.80	85.80	98.40	90.80	95.40	0.00	79.20	74.22
Seigneuric E0	0.00	0.00	98.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.86
Seigneuric E2	98.40	99.80	100.00	99.80	92.40	100.00	0.00	100.00	88.20	97.40	87.60
Shi	0.00	0.00	96.80	0.00	0.00	93.20	0.00	0.00	0.00	86.20	27.62
Shou	93.60	97.80	0.00	0.00	0.00	0.00	0.00	95.80	0.00	97.80	38.50
Suh	0.00	96.40	0.00	0.00	0.00	0.00	0.00	90.00	0.00	80.20	26.66
Sorensen	0.00	99.80	0.00	0.00	94.00	0.00	0.00	97.60	0.00	98.60	39.00
Starmans	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	10.00
Sun	98.00	71.60	0.00	0.00	0.00	88.40	0.00	100.00	0.00	88.60	44.66
Sung	0.00	89.20	99.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	18.86
Tardon	97.80	94.80	96.80	0.00	0.00	0.00	0.00	96.60	0.00	88.20	47.42
Toustrup	0.00	98.00	95.40	0.00	92.00	0.00	0.00	95.00	0.00	93.00	47.34
Trong	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Van Malenstein	0.00	97.20	0.00	0.00	0.00	85.60	0.00	94.80	0.00	98.20	37.58
Wang 2005	0.00	89.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.92
Wang 2020	0.00	0.00	86.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.60
Winter	100.00	100.00	100.00	98.40	100.00	100.00	0.00	100.00	89.40	100.00	88.78
Yang 2017	0.00	0.00	95.40	0.00	90.80	0.00	0.00	0.00	0.00	0.00	18.62
Yang Prostate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yang Sarcoma	0.00	95.80	95.20	0.00	0.00	0.00	0.00	0.00	0.00	86.80	27.78
Ye	97.40	99.00	98.60	0.00	0.00	0.00	0.00	98.80	0.00	98.00	49.18
Zhang	99.60	99.40	87.20	90.60	0.00	99.80	99.20	99.00	100.00	99.20	87.40
Zou	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

b) Median

	COAD	LUSC	HNSC	STAD	THCA	BRCA	PRAD	UCEC	LIHC	LUAD	Mean percentage
Aprelikova	0.00	99.80	0.00	0.00	98.80	0.00	0.00	0.00	0.00	97.80	29.64
Benita	0.00	94.20	94.20	0.00	0.00	0.00	0.00	0.00	0.00	88.60	27.70
Beyer	0.00	0.00	0.00	0.00	99.40	0.00	0.00	0.00	0.00	0.00	9.94
Boidot Cont.	88.80	99.00	98.80	0.00	0.00	99.80	0.00	99.60	0.00	98.00	58.40
Boidot Cyc.	94.80	91.20	98.00	0.00	90.20	98.80	99.80	99.40	0.00	84.00	75.62
Bosco	0.00	0.00	99.60	70.00	0.00	0.00	0.00	0.00	0.00	0.00	16.96
Buffa	99.60	100.00	100.00	94.60	99.00	99.80	0.00	100.00	89.80	99.80	88.26
Chen	90.60	100.00	99.40	0.00	99.60	0.00	0.00	0.00	96.00	99.40	58.50
Chi	0.00	93.00	97.80	0.00	97.40	0.00	0.00	0.00	0.00	91.20	37.94
Denko	0.00	0.00	97.40	0.00	96.20	0.00	0.00	0.00	0.00	0.00	19.36
Detwiller	0.00	0.00	99.80	0.00	93.60	0.00	0.00	0.00	0.00	0.00	19.34
Elvidge	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	10.00
Eustace	99.80	99.80	98.60	93.20	0.00	0.00	0.00	99.00	73.00	96.60	66.00
Fardin 2009	78.60	98.20	88.60	0.00	0.00	0.00	0.00	92.40	0.00	95.20	45.30
Fardin 2010	0.00	98.80	0.00	0.00	0.00	95.00	0.00	99.80	0.00	94.80	38.84
Fjeldbo	86.80	88.00	93.40	0.00	98.00	81.60	0.00	91.80	78.20	68.80	68.66
Ghazoui	100.00	100.00	100.00	97.80	0.00	100.00	99.80	100.00	80.80	100.00	87.84
Ghorbel	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Halle	0.00	95.20	97.40	0.00	97.80	88.80	0.00	99.20	0.00	87.80	56.62
Hu	0.00	88.80	95.80	0.00	94.80	0.00	0.00	98.60	0.00	0.00	37.80
Jogi	0.00	0.00	0.00	0.00	99.40	0.00	0.00	0.00	0.00	0.00	9.94
Koong	81.00	84.80	96.20	86.60	84.00	84.00	0.00	0.00	0.00	80.20	59.68
Lendhal	84.60	98.60	89.40	0.00	0.00	0.00	0.00	100.00	0.00	90.60	46.32
Lin	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Manalo	0.00	0.00	96.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.60
Mense	0.00	90.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.00
Mo	0.00	98.80	0.00	0.00	92.00	0.00	0.00	94.20	0.00	98.60	38.36
Ning	96.00	96.20	100.00	85.80	95.60	0.00	0.00	0.00	95.00	96.60	66.52
Peters	98.60	100.00	99.40	66.00	96.60	99.20	0.00	100.00	98.20	100.00	85.80
Ragnum	98.00	100.00	99.80	99.40	98.20	100.00	99.40	100.00	99.80	97.20	99.18
Seigneuric C.	76.40	77.00	94.20	83.00	0.00	89.40	0.00	0.00	0.00	82.60	50.26
Seigneuric E0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Seigneuric E2	83.20	85.20	99.80	81.20	95.60	99.00	0.00	88.00	0.00	96.20	72.82
Shi	0.00	0.00	97.40	0.00	0.00	82.60	0.00	0.00	0.00	0.00	18.00
Shou	0.00	96.00	0.00	0.00	0.00	0.00	0.00	96.40	0.00	96.00	28.84
Suh	87.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.76
Sorensen	0.00	99.80	0.00	0.00	99.00	88.00	0.00	92.40	0.00	99.00	47.82
Starmans	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	99.60	9.96
Sun	88.00	79.80	0.00	0.00	0.00	0.00	0.00	95.00	0.00	0.00	26.28
Sung	0.00	85.80	0.00	0.00	99.00	0.00	0.00	0.00	0.00	0.00	18.48
Tardon	80.60	93.80	92.60	0.00	93.80	0.00	0.00	97.60	0.00	94.00	55.24
Toustrup	0.00	99.20	91.00	0.00	99.60	90.40	0.00	96.40	0.00	98.40	57.50
Trong	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Van Malenstein	0.00	94.60	0.00	0.00	0.00	83.40	0.00	94.40	0.00	99.00	37.14
Wang 2005	0.00	0.00	0.00	0.00	0.00	0.00	0.00	96.40	0.00	0.00	9.64
Wang 2020	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Winter	100.00	100.00	100.00	97.40	100.00	99.20	0.00	100.00	87.40	100.00	88.40
Yang 2017	0.00	0.00	99.40	0.00	97.60	0.00	0.00	0.00	0.00	0.00	19.70
Yang Prostate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yang Sarcoma	0.00	93.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	81.00	17.42
Ye	82.80	99.20	0.00	0.00	0.00	0.00	0.00	96.80	0.00	93.40	37.22
Zhang	85.00	99.60	88.60	87.60	0.00	99.00	98.20	100.00	99.80	98.60	85.64
Zou	84.40	0.00	0.00	0.00	93.00	82.40	91.80	0.00	0.00	0.00	35.16

c) GSVA

	COAD	LUSC	HNSC	STAD	THCA	BRCA	PRAD	UCEC	LIHC	LUAD	Mean percentage
Aprelikova	0.00	73.20	0.00	0.00	97.00	0.00	0.00	0.00	0.00	0.00	17.02
Benita	74.40	85.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	76.20	23.58
Beyer	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Boidot Cont.	93.80	97.80	87.40	0.00	0.00	93.20	92.40	97.20	81.80	92.40	73.60
Boidot Cyc.	0.00	90.20	0.00	0.00	85.20	93.00	88.20	85.80	74.40	89.20	60.60
Bosco	0.00	0.00	87.60	0.00	87.20	0.00	0.00	0.00	0.00	0.00	17.48
Buffa	97.80	100.00	99.20	88.40	93.00	97.40	88.40	99.80	83.40	97.40	94.48
Chen	92.00	100.00	97.60	0.00	87.80	82.20	0.00	96.00	92.20	99.40	74.72
Chi	71.00	85.40	93.80	0.00	84.80	0.00	0.00	82.00	0.00	85.20	50.22
Denko	0.00	0.00	0.00	0.00	0.00	71.60	0.00	0.00	0.00	0.00	7.16
Detwiller	83.40	73.80	98.40	0.00	93.80	0.00	0.00	0.00	0.00	0.00	34.94
Elvidge	0.00	0.00	0.00	0.00	83.80	0.00	0.00	0.00	0.00	0.00	8.38
Eustace	94.80	100.00	98.00	0.00	91.00	89.00	0.00	100.00	0.00	98.80	67.16
Fardin 2009	80.20	99.00	89.20	0.00	0.00	0.00	0.00	91.00	0.00	99.00	45.84
Fardin 2010	0.00	94.00	0.00	0.00	0.00	82.00	0.00	99.20	0.00	91.00	36.62
Fjeldbo	95.80	91.00	86.40	0.00	99.00	82.60	0.00	94.00	93.40	81.00	72.32
Ghazoui	99.00	100.00	98.40	89.20	79.40	96.60	92.20	100.00	74.60	98.40	92.78
Ghorbel	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Halle	85.60	94.20	91.60	0.00	95.60	84.60	0.00	99.20	0.00	94.00	64.48
Hu	0.00	91.00	89.20	0.00	87.40	0.00	0.00	95.00	0.00	0.00	36.26
Jogi	0.00	0.00	0.00	0.00	97.20	0.00	0.00	0.00	0.00	0.00	9.72
Koong	92.00	91.00	97.00	89.80	82.00	90.60	0.00	0.00	0.00	93.20	63.56
Lendhal	81.00	95.00	87.20	0.00	0.00	0.00	0.00	96.00	0.00	86.40	44.56
Lin	74.60	85.00	0.00	0.00	74.60	70.00	0.00	93.20	0.00	79.20	47.66
Manalo	0.00	0.00	86.20	0.00	86.20	0.00	0.00	0.00	0.00	0.00	17.24
Mense	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mo	0.00	98.00	87.20	0.00	92.80	0.00	0.00	92.40	0.00	97.60	46.80
Ning	83.20	92.60	82.60	0.00	92.60	72.60	0.00	80.00	0.00	84.60	58.82
Peters	96.20	88.80	0.00	0.00	86.60	84.00	0.00	85.40	84.00	84.40	60.94
Ragnum	99.20	100.00	100.00	97.40	98.00	99.60	97.00	100.00	99.60	99.00	98.98
Seigneuric C.	0.00	81.80	99.80	95.80	92.80	94.20	0.00	83.20	0.00	0.00	54.76
Seigneuric E0	0.00	0.00	80.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.08
Seigneuric E2	94.60	88.80	99.80	87.60	89.20	98.80	0.00	91.60	0.00	82.00	73.24
Shi	0.00	0.00	93.60	0.00	90.60	99.40	0.00	96.40	81.80	96.40	55.82
Shou	90.60	98.40	0.00	0.00	78.60	0.00	0.00	97.60	0.00	94.00	45.92
Suh	89.60	86.00	0.00	89.20	74.80	0.00	0.00	86.00	0.00	0.00	42.56
Sorensen	0.00	94.20	0.00	0.00	90.80	0.00	0.00	85.20	0.00	91.80	36.20
Starmans	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	74.60	7.46
Sun	0.00	0.00	0.00	0.00	0.00	80.80	0.00	98.20	0.00	90.20	26.92
Sung	0.00	0.00	0.00	0.00	91.80	0.00	0.00	0.00	0.00	0.00	9.18
Tardon	85.00	84.20	84.40	0.00	0.00	0.00	0.00	87.80	0.00	76.20	41.76
Toustrup	0.00	92.80	84.60	0.00	90.40	0.00	0.00	88.20	0.00	84.60	44.06
Trong	0.00	76.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.60
Van Malenstein	0.00	97.60	0.00	0.00	0.00	82.20	0.00	92.80	0.00	99.20	37.18
Wang 2005	0.00	69.60	0.00	0.00	0.00	68.00	0.00	87.00	0.00	0.00	22.46
Wang 2020	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Winter	97.80	99.80	98.20	88.80	96.80	98.00	0.00	99.40	85.80	96.00	86.06
Yang 2017	0.00	0.00	90.00	0.00	83.40	0.00	0.00	0.00	0.00	0.00	17.34
Yang Prostate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yang Sarcoma	0.00	78.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.86
Ye	74.80	94.00	0.00	0.00	0.00	0.00	0.00	85.80	0.00	85.80	34.04
Zhang	97.40	99.20	88.80	91.00	0.00	99.60	100.00	99.20	100.00	98.40	87.36
Zou	0.00	0.00	0.00	0.00	90.00	0.00	0.00	82.60	0.00	0.00	17.26

d) ssGSEA

	COAD	LUSC	HNSC	STAD	THCA	BRCA	PRAD	UCEC	LIHC	LUAD	Mean percentage
Aprelikova	0.00	75.00	0.00	0.00	95.60	0.00	0.00	0.00	0.00	67.40	23.80
Benita	0.00	80.60	79.40	0.00	0.00	0.00	0.00	0.00	0.00	67.40	22.74
Beyer	0.00	0.00	0.00	0.00	74.40	0.00	0.00	0.00	0.00	0.00	7.44
Boidot Cont.	0.00	96.60	89.20	0.00	0.00	90.00	87.00	0.00	71.40	90.80	52.50
Boidot Cyc.	0.00	81.20	0.00	0.00	83.20	93.20	82.20	0.00	0.00	81.00	42.08
Bosco	0.00	0.00	90.60	0.00	78.20	0.00	0.00	0.00	0.00	0.00	16.88
Buffa	0.00	99.80	100.00	97.20	92.00	97.20	88.80	0.00	88.60	96.20	75.98
Chen	0.00	99.60	99.60	0.00	98.20	0.00	0.00	0.00	78.80	98.40	47.46
Chi	0.00	77.40	93.40	0.00	82.60	0.00	0.00	0.00	0.00	81.60	33.50
Denko	0.00	0.00	76.20	0.00	74.80	0.00	0.00	0.00	0.00	0.00	15.10
Detwiller	0.00	69.20	99.00	0.00	82.40	0.00	0.00	0.00	0.00	0.00	25.06
Elvidge	0.00	0.00	0.00	0.00	79.40	0.00	0.00	0.00	0.00	0.00	7.94
Eustace	0.00	99.60	99.00	93.80	88.00	0.00	0.00	0.00	0.00	95.20	47.56
Fardin 2009	0.00	98.60	84.00	0.00	72.60	0.00	0.00	0.00	0.00	96.00	35.12
Fardin 2010	0.00	95.20	0.00	0.00	0.00	74.80	0.00	0.00	0.00	92.40	26.24
Fjeldbo	0.00	85.80	82.60	0.00	99.20	75.00	0.00	0.00	96.80	83.20	52.26
Ghazoui	0.00	100.00	99.40	98.00	70.80	98.40	89.40	0.00	80.60	97.20	73.38
Ghorbel	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Halle	0.00	95.80	94.60	0.00	90.80	79.40	0.00	0.00	71.40	89.00	52.10
Hu	0.00	76.80	81.20	0.00	89.00	0.00	0.00	0.00	0.00	0.00	24.70
Jogi	0.00	0.00	85.40	0.00	90.20	0.00	0.00	0.00	0.00	0.00	17.56
Koong	0.00	88.40	96.00	98.60	74.60	90.40	0.00	0.00	0.00	88.80	53.68
Lendhal	0.00	96.40	90.20	0.00	72.60	0.00	0.00	0.00	0.00	86.60	34.58
Lin	0.00	0.00	78.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.88
Manalo	0.00	0.00	86.00	0.00	75.20	0.00	0.00	0.00	0.00	0.00	16.12
Mense	0.00	67.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.72
Mo	0.00	96.60	0.00	0.00	88.20	0.00	0.00	0.00	0.00	97.20	28.20
Ning	0.00	67.00	99.00	0.00	86.20	0.00	0.00	0.00	0.00	68.40	32.06
Peters	0.00	95.60	97.60	0.00	0.00	78.80	0.00	0.00	92.60	95.40	46.00
Ragnum	0.00	100.00	100.00	100.00	95.00	99.60	97.40	100.00	99.60	96.40	88.80
Seigneuric C.	0.00	80.80	98.80	99.80	83.00	97.20	0.00	0.00	0.00	75.20	53.48
Seigneuric E0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Seigneuric E2	0.00	96.40	99.60	99.80	80.40	98.80	0.00	0.00	76.60	94.00	64.56
Shi	0.00	0.00	96.80	0.00	0.00	89.60	0.00	0.00	0.00	74.40	26.08
Shou	0.00	96.00	0.00	0.00	0.00	68.80	0.00	0.00	72.40	96.00	33.32
Suh	0.00	88.40	85.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	17.42
Sorensen	0.00	96.60	0.00	0.00	87.80	0.00	0.00	0.00	0.00	95.40	27.98
Starmans	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	76.00	7.60
Sun	0.00	97.20	79.40	0.00	0.00	86.40	0.00	0.00	0.00	94.80	35.78
Sung	0.00	70.00	86.40	0.00	84.80	0.00	0.00	0.00	0.00	0.00	24.12
Tardon	0.00	85.60	85.60	0.00	79.60	0.00	0.00	0.00	0.00	78.20	32.90
Toustrup	0.00	94.00	87.60	0.00	94.20	0.00	0.00	0.00	0.00	91.00	36.68
Trong	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Van Malenstein	0.00	96.20	0.00	0.00	0.00	79.20	0.00	0.00	0.00	98.60	27.40
Wang 2005	0.00	66.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.66
Wang 2020	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Winter	0.00	99.60	99.40	96.60	95.40	92.40	0.00	0.00	82.60	97.60	66.36
Yang 2017	0.00	0.00	87.40	0.00	86.60	0.00	0.00	0.00	0.00	0.00	17.40
Yang Prostate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yang Sarcoma	0.00	86.00	81.80	0.00	0.00	0.00	0.00	0.00	0.00	75.60	24.34
Ye	0.00	94.40	81.60	0.00	0.00	0.00	0.00	0.00	0.00	90.20	26.62
Zhang	98.80	99.80	89.20	93.20	0.00	100.00	99.60	99.40	100.00	98.40	87.84
Zou	0.00	0.00	0.00	0.00	85.20	67.20	0.00	0.00	0.00	0.00	15.24

Table 6.2: *The percentage of the time the p-value of the original signature is lower than the p-values of any of the 500 RGS across 10 cancer types from the TCGA for all 53 published hypoxia signatures*

Comparison of the SPI calculated as the number of times the p-value using the Mann-Whitney-Wilcoxon test on the original signatures is significant and lower than any other of the 500 p-values derived from the RGS. The percentage is shown in different shades of blue from lowest (light blue) to highest (dark blue). Red cells highlight the best performing signature(s) for each cancer type. The last column represents the mean percentage across all cancer types and is shown in different shades of green from lowest (light green) to highest (dark green). Panel A = mean, panel B = median, panel C = GSVA and panel D = ssGSEA.

6.5 Evaluation Step 3: Do hypoxia signature scores from tumours confer prognostic information?

Hypoxia is associated with poor prognosis as it causes phenotypic changes in cells as well as hampering core therapies such as radiotherapy. The work above identified three signatures and associated scores that tend to be higher in tumours compared to normal tissue and tend to outperform RGS (Ragnum, Buffa and Ghazoui). Therefore, the next step is to investigate the survival of individuals using these three signatures and the scores in which they performed well in the previous two steps (Ragnum: mean, median & GSVA, Buffa: mean & GSVA and Ghazoui: GSVA). Here the goal is to compare the “traditional” method, comparing survival for those who have hypoxia scores above and below the median in cancer samples, to new methods making use of the NAT scores, namely:

1) Those samples above and below the **maximum** NAT sample hypoxia score (max normal)
2) Those samples above and below the **mean** NAT sample hypoxia score (mean normal)
3) Those samples above and below the **median** NAT sample hypoxia score (median normal)

Perhaps when trialling hypoxic modifiers to maximise potential therapeutic benefit, one of the above three methods might be more appropriate. If hypoxia modifiers are allocated based on raw hypoxia score, and the incorrect score is chosen without reference to normal tissue or NAT samples, you may end up in a position where high hypoxia scores in the tumour tissue are still less than in the normal tissue. This may lead to mis-stratification of patients and assumed failure of hypoxia modifying therapies. Therefore, as with comparing a normoxic reference in the cell line work, it is possible a comparison to NAT using the right signatures and scores might be an exciting avenue to investigate.

I investigated survival adopting two commonly used methods, Kaplan-Meier (KM) survival curves compared with the log-rank test and Cox survival on the three promising

signatures identified. It is worth noting that when using KM, whichever of the grouping approaches is chosen from the list previously outlined can change the hypoxia score group of samples (low or high), illustrated in *Fig. 6.4*. Here, taking Buffa as an example, one can see that selecting the median value of the hypoxia score on all the tumour samples alone places the threshold for the two groups *higher* than using the mean NAT sample score as a threshold (*Fig. 6.4*).

I repeated the KM survival curves for Ragnum, Buffa and Ghazoui using the mean and GSVA scores and the four different approaches (traditional, max NAT, mean NAT and median NAT). Looking at the 10 tissue types, Ragnum, Buffa and Ghazoui using the mean score showed significant differences in survival over five years using the aforementioned four different approaches. The most significant p was Buffa signature using mean split on the median tumour score as a threshold for the two groups (p = 3.69E-67, *Fig. 6.5*, other results found in *Appendix 4, Supplementary Fig. S4-S13*), and notably the median tumour threshold was lower than the max NAT hypoxia score (*Fig. 6.6*). However, looking at the mean NAT and median NAT cut offs, although not as significant, average survival at 5 years appeared higher than in the “lower hypoxia group” than using the traditional method (blue line): >85% Buffa (*Fig. 6.7, 6.8*), >90% with Ragnum (*Appendix 4, Supplementary Fig. S9 c-d*) and >85% Ghazoui (*Appendix 4, Supplementary Fig. S10 c-d*) compared to ~78% with Buffa (*Fig. 6.5*), Ragnum (*Appendix 4, Supplementary Fig. S9 a*), or Ghazoui (*Appendix 4, Supplementary Fig. S10 a*) using the traditional method. This perhaps suggests the traditional method may place individuals who have some degree of hypoxia within their tumours in the “lower hypoxia group”. The most significant result with GSVA was achieved by Ghazoui using the median tumour score as a threshold for the two groups (p=3.18E-13, *Appendix 4, Supplementary Fig. S13 a*). Ragnum using the median score also showed the lowest p (p=1.55E-33) using the median tumour score as a threshold for the two groups.

Since survival can be affected by various covariates which are not considered in the KM model, I moved to investigate prognostic significance using Cox-proportional hazards

modelling, using age, stage and histological types as covariates, sex as strata and the relevant scores as continuous variables. Here, Buffa signature using the mean score appeared the best (*Fig. 6.9*), yielding an hazard ratio greater than age, stage or histological type (1.82 [1.66-2.01], p = 4.62E-35 vs age: 1.38 [1.27-1.52], p = 1.07E-12, stage: 1.011 [1.008-1.014], p = 9.9E-12, histological subtype 1.003 [0.997-1.009], p= 0.294). It is worth noting that the Buffa signature was not derived from the TCGA data. Other results can be found in *Appendix 5, Supplementary Fig. S14-S18*.

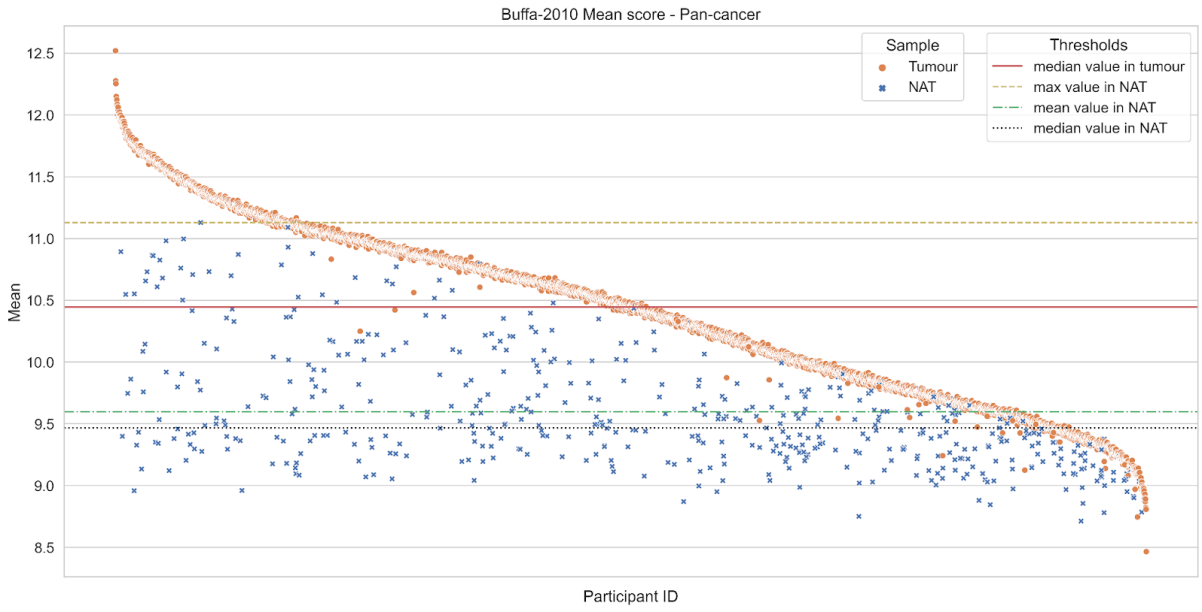


Figure 6.4: *The distribution of Buffa hypoxia scores using the mean in 10 cancer types in the TCGA cohort*

Each data point represents a sample in the TCGA cohort. All NAT samples are shown with a blue cross and all tumour samples are shown with an orange circle. The four coloured horizontal lines show the position of the thresholds: red non-broken line (median value of the tumour samples), yellow broken line (max NAT population), green broken line (mean NAT population) and black broken line (median NAT population). Different thresholds can result in different sample classification.

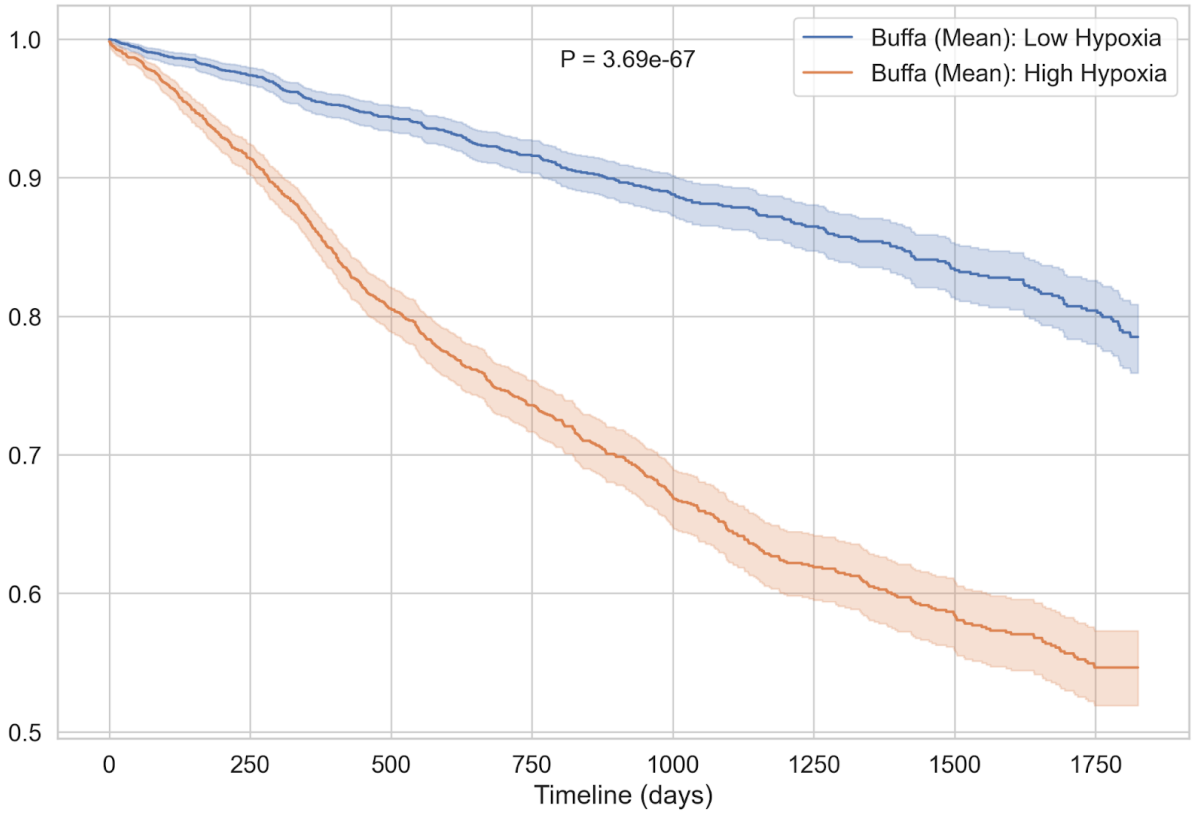


Figure 6.5: *Survival of patients in the TCGA dataset according to Buffa signature (mean score, traditional threshold)*

KM survival curve of patients in the TCGA cohort. The median value of the hypoxia scores on tumour samples was used to denote the high/low hypoxia groups (the “traditional approach”). Hypoxia scores were calculated using the Buffa signature and the mean score. The timeline (x-axis) is expressed in days. Patients with low hypoxia scoring samples are presented in blue, whereas those with high hypoxia scores are shown in orange. The log-rank test showed a p-value = 3.69E-67.

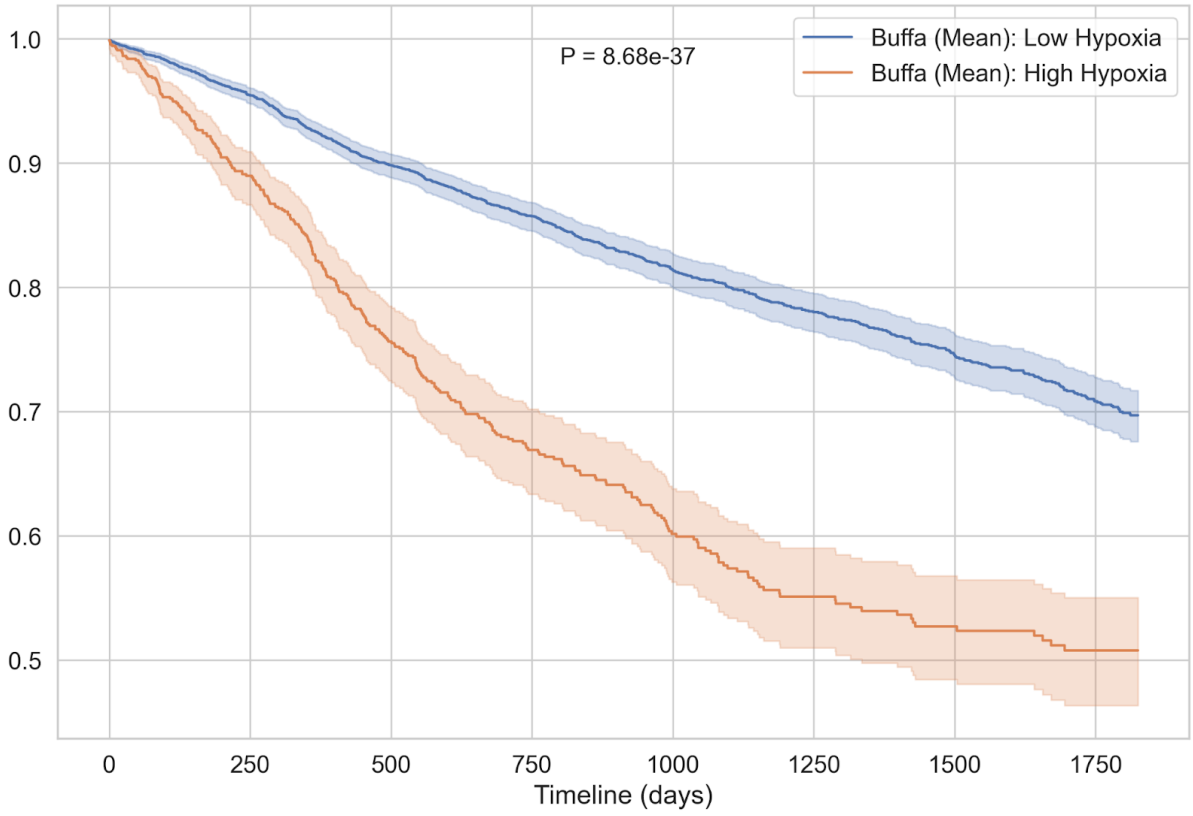


Figure 6.6: *Survival of patients in the TCGA dataset according to Buffa signature (mean score, maximum NAT threshold)*

KM survival curve of patients in the TCGA cohort. The maximum value of NAT hypoxia scores was used to denote the high/low hypoxia groups. Hypoxia scores were calculated using the Buffa signature and the mean score. Conventions as in *Fig. 6.5*. The log-rank test showed a p-value = 8.68E-37.

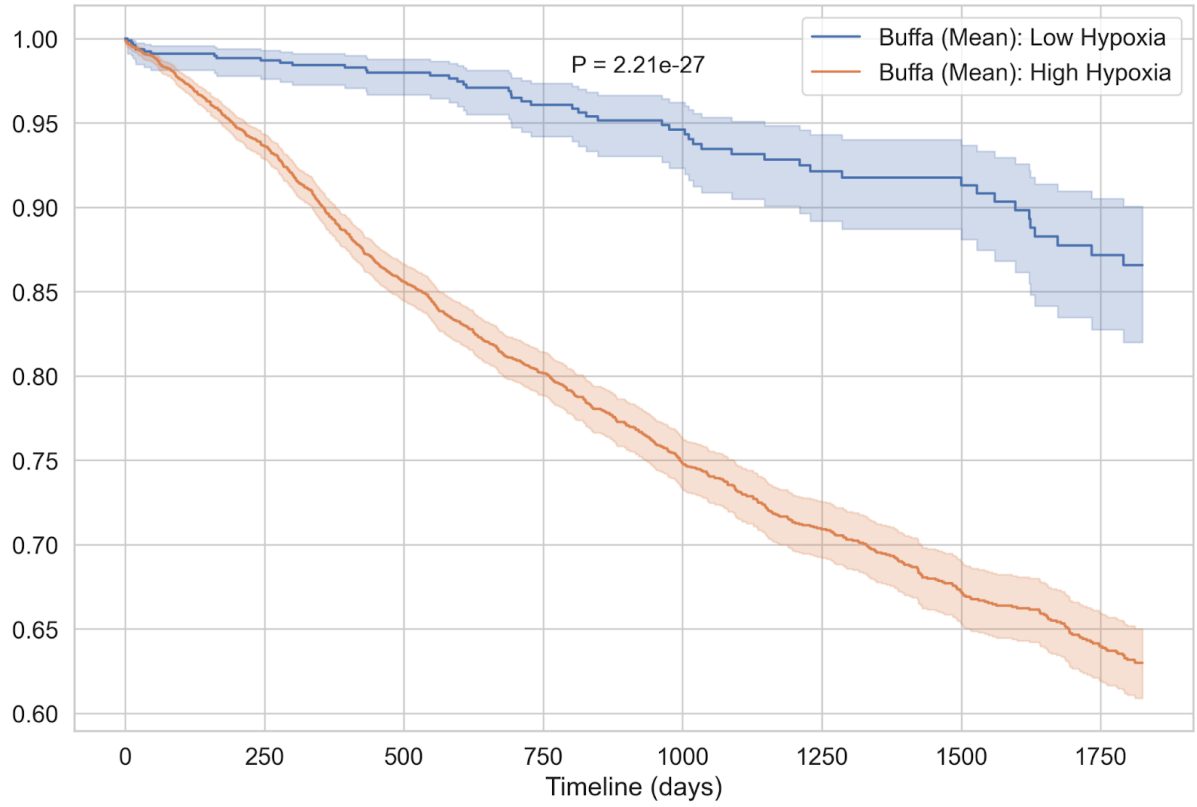


Figure 6.7: *Survival of patients in the TCGA dataset according to Buffa signature (mean score, mean NAT threshold)*

KM survival curve of patients in the TCGA cohort. The mean value of NAT hypoxia scores was used to denote the high/low hypoxia groups. Hypoxia scores were calculated using the Buffa signature and the mean score. Conventions as in *Fig. 6.5*. The log-rank test showed a p-value = 2.21E-27.

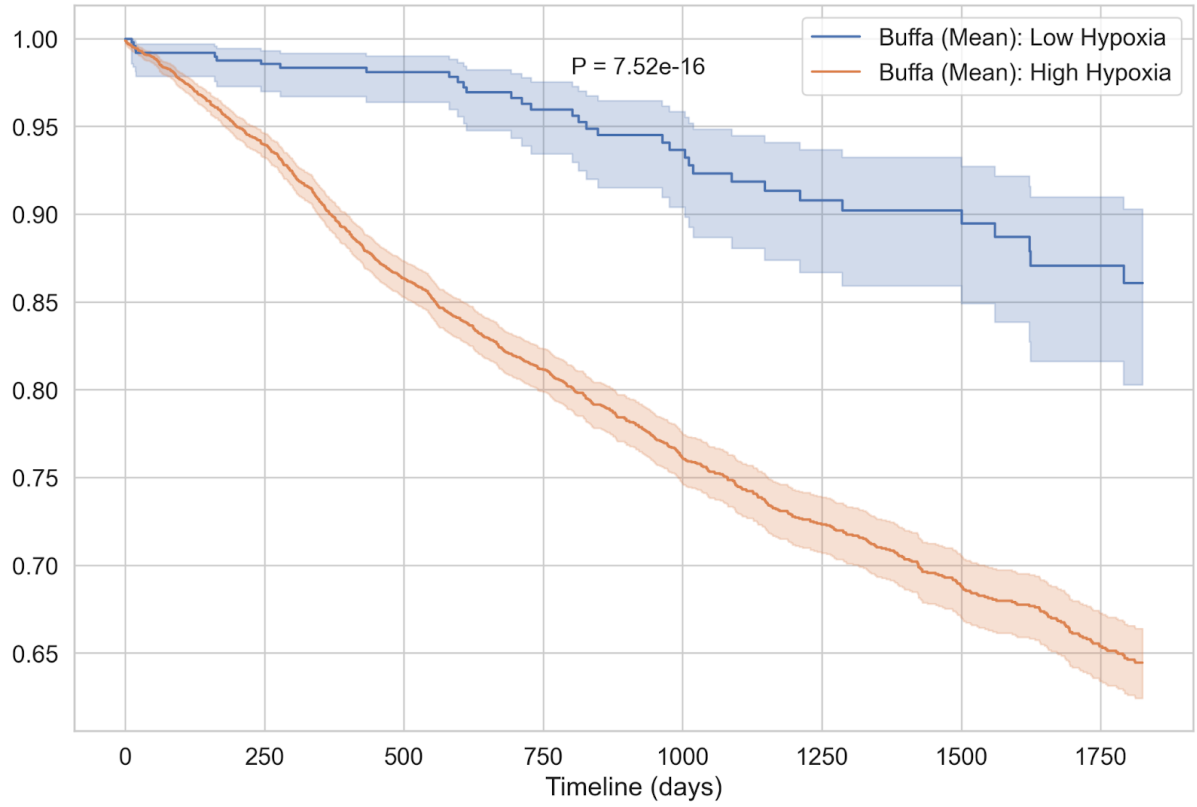
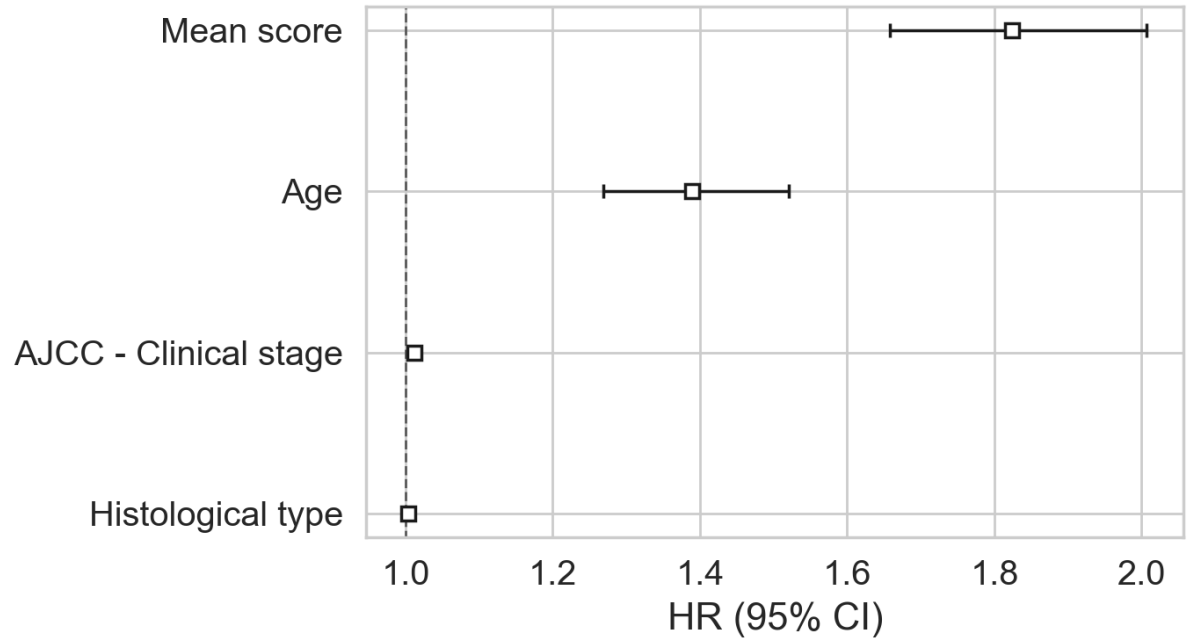


Figure 6.8: *Survival of patients in the TCGA dataset according to Buffa signature (mean score, median NAT threshold)*

KM survival curve of patients in the TCGA cohort. The median value of NAT hypoxia scores was used to denote the high/low hypoxia groups. Hypoxia scores were calculated using the Buffa signature and the mean score. Conventions as in *Fig. 6.5*. The log-rank test showed a p-value = 7.52E-16.

a)



b)

covariate	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
AJCC - Clinical stage	0.011380	1.011445	0.001672	0.008104	0.014656	1.008137	1.014764	0.0	6.807890	9.904030e-12	36.555121
Age	0.328744	1.389222	0.046169	0.238254	0.419234	1.269031	1.520796	0.0	7.120410	1.076061e-12	39.757377
Histological type	0.003138	1.003142	0.002992	-0.002727	0.009002	0.997277	1.009043	0.0	1.048590	2.943666e-01	1.764314
Mean score	0.601171	1.824254	0.048662	0.505796	0.696546	1.658305	2.006809	0.0	12.354124	4.627694e-35	114.057190

Figure 6.9: *Survival analysis using a cox-proportional hazards model on the TCGA dataset using the Buffa signature (mean score)*

Results for common covariates and the Buffa signature mean score from a cox-proportional hazards model on 10 tumour types in the TCGA. (a) shows Hazard Ratios (HR) whilst (b) is a summary table detailing coefficients, standard error, z and p-values.

6.6 Final recommendation for clinical samples

Through this comprehensive three stage evaluation in the search for the most promising hypoxia signature for clinical use, the Buffa signature using the mean score appears most appropriate. While this study represents the most comprehensive work in this field to date, it is recognised that it does not include in vivo human comparisons to oxygen levels using oxygen electrodes or other methods. However, this data-driven approach in the landmark TCGA cohort of over 1000 cancer samples highlights the Buffa signature using the mean score as a) being on average higher in tumour samples than in normal tissue samples, b) differing in performance compared to random gene signatures in 99.22% of the time c) serving as a strong prognostic marker (using a number of different thresholds) across 10 important tumour types. Therefore, until large prospective studies involving various hypoxia markers are conducted, it is recommended to use the Buffa signature with the mean score for determining the hypoxic status of clinical samples.

Chapter 7: Final thoughts

This work is the largest and most comprehensive analysis and validation of hypoxia signatures to date. The study includes *a)* a systematic review of all published hypoxia signatures, *b)* an unbiased analysis of hypoxia signature performance in all open access hypoxia cell line data spanning 90 different cell line/tissue experiments and, finally *c)* a data-driven approach to find the most promising signature for use in clinical samples using the landmark cancer genomics study, The Cancer Genome Atlas (TCGA). This study also applied a novel approach for signature evaluation using random gene sets (RGS) which addresses a major query surrounding a use of signatures in this field. This work not only summarises current hypoxia signatures in the literature but also highlights which signatures and scores should be considered at the benchside and for validation in the clinic, and importantly highlights signatures which should be used with extreme caution.

After uncovering 53 published hypoxia signatures using a systematic approach, marked differences were discovered in their performance depending on the signature, score and even cell line used. Overall, consistently ssGSEA appeared the worse summary score whereas median, mean and GVSA in general appeared more effective. The likely reason for scoring methods working better in certain conditions is likely the method used to derive the scores (discussed in Chapter 3.3).

Looking at the individual scores, they have different characteristics. One of them is their reliance on assuming underlying distributions. The median, mean and ssGSEA scores do not assume anything about the distribution of the data. However, the first step in calculating the GSVA score consists of fitting the data into a model, in which the intensity of a gene is determined following a Gaussian or Poisson distribution.

Another obvious point of difference is how the scores are calculated. How to calculate the median and mean scores are well known and relatively simple, whereas GSVA and ssGSEA are more complex. The GSVA score tends to perform very well in hypoxia signatures that include a lot of coexpressed genes. The GSVA score is based on the concept that a gene set's enrichment in a sample will depend on the position of such genes

in the ranked list of all the genes available in the dataset. Thus, if most of the genes in the signature are all ranked in higher positions compared to the other genes in the dataset, then the GSVA score will be high. On the other hand, ssGSEA is a gene set enrichment score that represents the activity level of the biological process in which the genes belonging to the signature are coordinately up-regulated or down-regulated. As a consequence, this score becomes very helpful when researchers need to determine the cell state in terms of the activity levels of biological pathways rather than through the expression levels of individual genes. This strong reliance on particular biological pathways and co-ordinated expression might be the reason why ssGSEA performed consistently worse than any other score in the study.

The mean and median too have pros and cons: both mean and median perform well on normal distributions, however, the mean score is more sensitive to outliers (not necessarily a negative attribute in the evaluation of hypoxia scores). On the other hand, median score does not take into account all the precise values in a distribution and is highly affected by sampling fluctuations. Notably, both scores would be identical in a completely normal distribution, but their values would differ according to the skewness of distributions.

On top of this, there are several variables that come into play during the evaluation of gene expression scores, which can be related to the score itself and/or the experimental data. This work attempted to overcome these problems by increasing the statistical power of the analysis. An unbiased approach was implemented, looking at all published signatures, a variety of scores and carrying out the largest analysis to date, across a variety of cell line datasets from different sources and sequencing technologies as well as the large and well-curated TCGA dataset.

The most effective signature and score combination on any cell line tested was Sorensen using the median score. Across 1090 pairwise combinations, Sorensen hypoxia signature using the median score yields an impressive 92.84% accuracy in identifying hypoxic samples in experiments (i.e cells in hypoxic chambers vs normoxic conditions) in

non-VHL mutated lines. However, individual tissue analyses of signatures (subsections of Chapter 4) did yield slightly higher percentage accuracies in determining hypoxic samples, e.g. Lendhal using the median score was most effective in breast cancer cell lines (Chapter 4.1, 96.90% accuracy). Further, as signatures have a strong HIF-1a reliance, VHL-mutated cell lines were analysed separately, highlighting Sung using the GSVA score as the most effective (Chapter 5.3, 81.25% accuracy). Interestingly, all signatures tended to perform badly in recognising hypoxia on *ex vivo* lung cancer samples from the GSE30979 dataset (Chapter 4.2), suggesting, perhaps, there might be a difference in signature performance *in vivo*. And, indeed, this difference turned out to be likely the case.

Looking at clinical samples, one major limitation of this study, and others in the literature, was the lack of hypoxia measurements. However, this work proposes a novel three stage approach to identify the most promising signature and score for use in clinical samples. Buffa signature using the mean score appeared the most promising, being consistently on average higher in tumour samples than in normal tissue samples, differing in performance compared to random gene signatures in 99.22% of the time and serving as a strong prognostic marker (using a number of different thresholds) across 10 important tumour types. Large prospective studies are urgently needed with multiple measures of hypoxia, however these three criteria give us some confidence that using Buffa signature and mean score in already existing genomic datasets and looking to validate it in future studies would be wise. Also, I highlight the potential value of building a bank of normal tissue samples, which should be normoxic, which could be used as a reference when deploying hypoxia signatures in clinical studies and allow patients perhaps to be stratified more appropriately.

Before this work, the choice of hypoxia signatures and scoring systems used in studies was done almost as a matter of personal choice (exemplar publication[5]), or at best on minimal analytic grounds (exemplar publication[151]). This large analysis provides an important reference to laboratory and clinician scientists who seek validation of hypoxic

status and/or are considering prospective trials. This approach has shown the potential advantage of including normal samples in assessing hypoxic status in tumour clinical samples. The strategy of using matched normal samples in the same population might be transformative in assessing tumour microenvironmental conditions, such as hypoxia, however it must be noted extra biopsies are not always available/possible and are not without risk, thus conferring a limitation of this approach. Large cohort studies of informed patients are needed.

Following on from this work, I hope my postdoctoral studies will build on some excellent validation work that has already started in Oxford, currently only focussed on a small sample of lung cancer patients[32]. In this study, a number of hypoxia markers were measured including gene expression, immunohistochemistry and [18F]-fluoromisonidazole PET-CT as part of a clinical trial of 15 patients which helps gain clinical validation of hypoxia signatures *in vivo*. I hope with the help of my supervisor I will be able to secure funding to co-lead a collaborative initiative to collect a large number of tumour and matched normal tissue samples with a variety of hypoxic markers across a range of cancer types. This will ultimately help address the key question of whether hypoxia signatures reflect the true state of hypoxia *in vivo* and which hypoxic measurement method, or even combination of approaches, might be the most appropriate for clinical use.

Acknowledgements

I would like to take this opportunity to thank all those who have contributed in any way, directly or indirectly, to the completion of this work.

- Professor Francesca Buffa for being my primary supervisor and her invaluable suggestions and support.
- Professor Benjamin Harris for his invaluable support and professionalism as a supervisor.
- Professor Gareth Bond for his enthusiasm and direction.
- Professor Adrian Harris for academic critique and advice.
- Professor David Harris for proofreading this thesis.
- Mr Samuel Harris for life advice and direction.

Badran Elshenawy, Alessandro Barberis, Ryan Carter, Philip Charlton, Wei-Chen Cheng, Yueachan Chi, Joachim Hagel, Fiona Hartley, Sakura Maezono, Ana Miar Cuervo, Dawn O'Reilly, Jayathilake Pahala Gedara, Clara Pavillet, Karen Sayal, Helen Sheldon, Charalampos Triantafyllidis, Ava Van Ess, Pedro Victori and others I may have accidentally omitted.

My family and friends for their considerable help and support, particularly my wife and beautiful baby Elizabeth.

I am also truly indebted to the European Research Council for their financial support.

Appendices

Appendix 1

Koong 2000

- ***HGNC***: BIK, EDN2, FGF3, GADD45A, IGFBP3, LRP1, MIF, MMP13, SERPINE1, VEGFA
- ***ENSG ID***: ENSG00000276701, ENSG00000100290, ENSG00000186895, ENSG00000106366, ENSG00000127129, ENSG00000240972, ENSG00000116717, ENSG00000123384, ENSG00000112715, ENSG00000146674, ENSG00000137745

Denko 2003:

- ***HGNC***: ACADSB, ALDOA, ALDOC, ARHGAP5, ARSL, BCL2L2, TGFBI, BPI, C9orf153, CALD1, NDRG1, CCNG2, CDKN1A, CLK1, CNOT8, BHLHE40, EFNA1, EIF4A3, ELL2, ENO3, ELF3, FABP5, FN1, GLRX, GPI, GPRC5A, HBP1, HERC3, HERPUD1, HILPDA, HK1, HLA-DQB1, HSPA5, IGFBP3, INSIG1, IRF6, ITGA5, LDHA, LONP1, LOX, MCC, MPI, MT1L, MUC1, MXI1, PDK3, PIM2, PLIN2, PLOD2, PNN, POLM, PPARD, PTGS1, QSOX1, RBPJ, RIOK3, RNASEL, SGSM2, SIAH2, SIRPA, SLC16A3, SLC2A3, SLC6A6, SLC6A8, SYT7, TNFAIP3, TPBG, TPD52, TPI1, UCA1, TXNIP, VEGFA
- ***ENSG ID***: ENSG00000282019, ENSG00000206237, ENSG00000231939, ENSG00000231286, ENSG00000283085, ENSG00000225824, ENSG00000206302, ENSG00000283847, ENSG00000233209, ENSG00000101782, ENSG00000288299, ENSG00000101425, ENSG00000102096, ENSG00000141258, ENSG00000100852, ENSG00000108515, ENSG00000059804, ENSG00000196365, ENSG00000179344, ENSG00000011347, ENSG00000198053, ENSG00000095303, ENSG00000178802, ENSG00000161638, ENSG00000147872, ENSG00000173221, ENSG00000196177, ENSG00000164687, ENSG00000130821, ENSG00000105220, ENSG00000067992, ENSG00000134333, ENSG00000120708, ENSG00000260549, ENSG00000181788, ENSG00000186480, ENSG00000118503, ENSG00000138641, ENSG00000171444, ENSG00000105856, ENSG00000138764, ENSG00000152952, ENSG00000124762, ENSG00000146242, ENSG00000118985, ENSG00000122786, ENSG00000129473, ENSG00000157399, ENSG00000214049, ENSG00000149925, ENSG00000104419, ENSG00000044574, ENSG00000119950, ENSG00000156515, ENSG00000135245, ENSG00000100941, ENSG00000116260, ENSG00000051108, ENSG00000141543, ENSG00000112715, ENSG00000111669, ENSG00000265972, ENSG00000122678, ENSG00000117595, ENSG00000163435, ENSG00000187753, ENSG00000134107, ENSG00000076554, ENSG00000135828, ENSG00000113083, ENSG00000168214, ENSG00000146674, ENSG00000155508, ENSG00000112033, ENSG00000109107, ENSG00000169242, ENSG00000185499, ENSG00000141526, ENSG00000013588, ENSG00000131389, ENSG00000115414, ENSG00000013441

Jogi 2004

- ***HGNC***: ACHE, ADM, AGPAT2, ALDOA, ALDOB, ALDOC, AOC3, AQP1, BCAT1, BCL6, BMP5, BNIP3L, C1orf21, CA12, CA9, CARTPT, CD9, CDH13, COL5A1, COX4I2, CPO, VCAN, SLC44A2, CXCR4, CYB5A, DDC, DLK1, EIF5, ENO1,

EPB41L4B, F7, VPS13C, RFK, UFSP2, DEPTOR, FOXF1, FUT1, GAL, GBE1, GNAL, GNAS, NEDD9, UBE2K, HK2, ID2, IGF2, IGFBP3, ITGAM, JAK1, PPIP5K2, UFL1, SIK2, SYNPO, ARRDC3, VPS13C, LMO1, LOXL2, GDE1, MT1H, MT2A, MXI1, NDRG1, NDUFB7, SMIM3, NMB, NRN1, NRP1, NUTF2, OLFM1, P4HA1, P4HA2, PAM, ARHGAP29, PDK1, PFKFB4, PFKL, PFKP, PGK1, PLOD1, CYTH3, SIRPA, PTPRN, RAB40C, RGS5, SCD, SLC12A3, SLC2A1, SLC2A3, SLC30A1, SNAPC1, ATL1, STC1, STRBP, TH, TNFRSF10B, RAD51D, MED16, TRIM29, KDM3A, VEGFA, WSB1, WT1, ZNF177

- ***ENSG ID***: ENSG00000282092, ENSG00000197562, ENSG00000057593, ENSG00000167244, ENSG00000175221, ENSG00000166407, ENSG00000141959, ENSG00000099250, ENSG00000059804, ENSG00000008256, ENSG00000184937, ENSG00000109046, ENSG00000159167, ENSG00000060982, ENSG00000074800, ENSG00000111859, ENSG00000134013, ENSG00000104765, ENSG00000141404, ENSG00000099194, ENSG00000198053, ENSG00000132437, ENSG00000120889, ENSG00000131055, ENSG00000115738, ENSG00000135002, ENSG00000166347, ENSG00000114480, ENSG00000074410, ENSG00000038427, ENSG00000006007, ENSG00000102898, ENSG00000069482, ENSG00000131471, ENSG00000130635, ENSG00000140945, ENSG00000078140, ENSG00000170145, ENSG00000137699, ENSG00000112175, ENSG00000113369, ENSG00000103241, ENSG00000107159, ENSG00000162434, ENSG00000100664, ENSG00000155792, ENSG00000136872, ENSG00000159399, ENSG00000087460, ENSG00000125148, ENSG00000185559, ENSG00000197696, ENSG00000137962, ENSG00000165209, ENSG00000169692, ENSG00000205358, ENSG00000130558, ENSG00000180176, ENSG00000102144, ENSG00000169896, ENSG00000149925, ENSG00000129003, ENSG00000109775, ENSG00000171992, ENSG00000104419, ENSG00000121966, ENSG00000148926, ENSG00000067057, ENSG00000188629, ENSG00000113916, ENSG00000124785, ENSG00000232995, ENSG00000119950, ENSG00000070915, ENSG00000112715, ENSG00000144410, ENSG00000095203, ENSG00000198513, ENSG00000099795, ENSG00000143248, ENSG00000145725, ENSG00000122884, ENSG00000087085, ENSG00000129353, ENSG00000023608, ENSG00000240583, ENSG00000010278, ENSG00000146674, ENSG00000072682, ENSG00000115548, ENSG00000174951, ENSG00000256235, ENSG00000109107, ENSG00000164326, ENSG00000152256, ENSG00000145730, ENSG00000054356, ENSG00000114268, ENSG00000014123, ENSG00000083444, ENSG00000185379, ENSG00000117394, ENSG00000116667, ENSG00000170385

Ning 2004

- ***HGNC***: ACTB, ACTR3, AIFM2, ANGPTL4, ATP5PO, BCAR3, BNIP3L, C1QTNF5, CASP1, CAV1, CBX3, CCT6A, CLASP1, CLTC, CSNK2B, DAPK3, CTTNBP2NL, MAGT1, DPP3, EEF1A1, EEF1G, EEF2, EFEMP1, ELOVL1, EMP3, ENO1, EWSR1, ACSL4, RUSF1, ORAI1, TOR4A, TTC19, EXOSC4, GAPDH, GMPPA, GNB2, GPI, LPAR5, CDV3, HDLBP, HNRNPH3, HSPH1, HSPA1A, HSPA5, HSPD1, HSPE1, IFI16, SEL1L3, SEC31A, RNF213, LAPTM4A, LOXL2, LXN, CAPRIN1, ZNF276, MMP2, MPG, MRPL34, NACA, BEX3, YBX1, PABPC1, PAK3, PARVB, SUB1, PDCD2, PLOD1, PLOD2, PML, PPP1CB, PTMS, RAC1, RRAGA, ARID4A, RBMX,

RNPEPL1, RPL23, RPL27A, RPS17, RPS27, RPS29, SELENOW, SERPINE1, SH3BGRL3, SPTB, TARDBP, TGFB1, TIMP2, TPI1, TPM3, TPT1, TRRAP, EFTUD2, UMPS, UQCR11, UQCRH, VEGFC, WASF2, YY1

- ***ENSG ID***: ENSG00000278229, ENSG00000237724, ENSG00000235941, ENSG00000282019, ENSG00000234475, ENSG00000230700, ENSG00000232960, ENSG00000206406, ENSG00000215328, ENSG00000120694, ENSG00000224774, ENSG00000224398, ENSG00000241837, ENSG00000228875, ENSG00000133112, ENSG00000167657, ENSG00000135387, ENSG00000127540, ENSG00000167658, ENSG00000166681, ENSG00000146731, ENSG00000188677, ENSG00000074800, ENSG00000011295, ENSG00000134013, ENSG00000104765, ENSG00000254772, ENSG00000100811, ENSG00000091490, ENSG00000158195, ENSG00000125691, ENSG00000087245, ENSG00000106366, ENSG00000182944, ENSG00000140464, ENSG00000113387, ENSG00000213639, ENSG00000173660, ENSG00000115380, ENSG00000147274, ENSG00000105220, ENSG00000213741, ENSG00000141367, ENSG00000158805, ENSG00000182774, ENSG00000223953, ENSG00000138674, ENSG00000068697, ENSG00000142227, ENSG00000137936, ENSG00000077264, ENSG00000198113, ENSG00000079257, ENSG00000143079, ENSG00000071994, ENSG00000032219, ENSG00000115091, ENSG00000152952, ENSG00000196367, ENSG00000068366, ENSG00000103152, ENSG00000178980, ENSG00000150630, ENSG00000276045, ENSG00000140688, ENSG00000102158, ENSG00000155876, ENSG00000042286, ENSG00000075624, ENSG00000035862, ENSG00000143549, ENSG00000044574, ENSG00000178896, ENSG00000070756, ENSG00000105974, ENSG00000173821, ENSG00000122565, ENSG00000167772, ENSG00000156508, ENSG00000204389, ENSG00000204435, ENSG00000111669, ENSG00000070182, ENSG00000096746, ENSG00000184574, ENSG00000159335, ENSG00000172354, ENSG00000136238, ENSG00000091527, ENSG00000105329, ENSG00000254986, ENSG00000130312, ENSG00000108883, ENSG00000196531, ENSG00000111640, ENSG00000142327, ENSG00000137752, ENSG00000083444, ENSG00000114491, ENSG00000177954, ENSG00000074054, ENSG00000115541, ENSG00000166441, ENSG00000144381, ENSG00000142669, ENSG00000065978, ENSG00000163565, ENSG00000066322, ENSG00000120948, ENSG00000144591, ENSG00000115677

Manalo 2005

- ***HGNC***: ACVR1B, ADM, ADORA2A, AK3, ANGPTL4, NUAK1, AXL, BCL6, BHLHE40, CASK, CCNG2, CDH2, CDK19, CDKN1C, PLEKHO1, ACKR3, COL18A1, COL1A2, COL4A1, COL4A2, COL5A1, COL9A1, CREB3L2, CUBN, CX3CL1, CXCR4, DUSP6, EDN1, EGLN1, EGLN3, ENPP1, EPOR, ERO1A, FNDC3B, FKBP9, GABRP, GADD45B, GDF10, GRK5, HES1, HIF3A, HIVEP2, IGFBP3, INHBA, INHBE, INSR, ITM2A, ITPR2, KDM3A, LEPR, LIMS1, LOX, LOXL2, SMAD6, MAFF, MEF2A, MXI1, NDRG1, NFATC4, NOTCH4, NPAS2, NPR1, OPN3, P4HA2, PAM, PDGFB, PDZD2, PELI2, PGF, PLCG2, PLOD1, PLOD2, PPARG, PTGIS, PTGS1, PTPRB, PTPRF, PTPRR, PPP1R13L, RASSF2, RGS3, RHOBTB1, RLN1, RPS6KA2, RRAS, SHOX2, SIRT3, SOX4, NABP2, STC1, STC2, TCF7L1, ZEB1, TLE1, TNFRSF10B, TNFRSF14, TRIO, TXNIP, VEGFA, VEGFC, VLDLR, WASF2, ZBTB1, ZHX2, ZNF292

- ***ENSG ID***: ENSG00000273936, ENSG00000273707, ENSG00000238196, ENSG00000223355, ENSG00000232339, ENSG00000235396, ENSG00000234876, ENSG00000206312, ENSG00000185022, ENSG00000134871, ENSG00000187498, ENSG00000107611, ENSG00000182871, ENSG00000100311, ENSG00000124212, ENSG00000147853, ENSG00000107018, ENSG00000170558, ENSG00000266524, ENSG00000285485, ENSG00000128271, ENSG00000078401, ENSG00000171105, ENSG00000124766, ENSG00000078596, ENSG00000129521, ENSG00000159167, ENSG00000197930, ENSG00000139946, ENSG00000134013, ENSG00000038382, ENSG00000120889, ENSG00000072422, ENSG00000164692, ENSG00000196781, ENSG00000095303, ENSG00000137834, ENSG00000158195, ENSG00000101265, ENSG00000133401, ENSG00000188994, ENSG00000138835, ENSG00000197943, ENSG00000130635, ENSG00000153233, ENSG00000198873, ENSG00000139318, ENSG00000119630, ENSG00000006210, ENSG00000155111, ENSG00000182158, ENSG00000071242, ENSG00000152284, ENSG00000010818, ENSG00000142949, ENSG00000068305, ENSG00000075420, ENSG00000147044, ENSG00000100968, ENSG00000178764, ENSG00000138764, ENSG00000152952, ENSG00000123104, ENSG00000169756, ENSG00000170485, ENSG00000150630, ENSG00000023902, ENSG00000094755, ENSG00000104419, ENSG00000114315, ENSG00000121966, ENSG00000169418, ENSG00000148926, ENSG00000139269, ENSG00000113916, ENSG00000167601, ENSG00000119950, ENSG00000147852, ENSG00000112715, ENSG00000167772, ENSG00000187266, ENSG00000126458, ENSG00000265972, ENSG00000099860, ENSG00000122641, ENSG00000148516, ENSG00000134107, ENSG00000144476, ENSG00000135766, ENSG00000126804, ENSG00000135503, ENSG00000124440, ENSG00000113083, ENSG00000204301, ENSG00000146674, ENSG00000122642, ENSG00000072682, ENSG00000142082, ENSG00000115548, ENSG00000127329, ENSG00000113739, ENSG00000197594, ENSG00000112280, ENSG00000104881, ENSG00000145730, ENSG00000116678, ENSG00000139579, ENSG00000083444, ENSG00000074590, ENSG00000168779, ENSG00000132170, ENSG00000129757, ENSG00000157873, ENSG00000054277

Wang 2005

- ***HGNC***: ALDOC, MANF, BHLHE40, BNIP3, BNIP3L, CACNA1A, CITED2, CXCR4, DECR2, DUSP1, FAM162A, FMO5, FOS, GADD45B, GBE1, GCHFR, HILPDA, H1-2, H2AC6, H2AC17, H2BC21, H2AW, HK2, HMGCL, HOXA13, HSPA5, ITGA2B, LDHA, LOX, MAFF, MARS1, OMA1, MXI1, MYOM2, NDRG1, NEFL, NPPB, NRN1, P4HA1, NAMPT, PCDH10, PFKFB4, PGF, PGK1, PKIA, PKIB, PLOD2, PPFIA4, PTPRZ1, RAB20, RASSF1, DDIT4, SAT1, TRIM21, VEGFA
- ***ENSG ID***: ENSG00000274296, ENSG00000274137, ENSG00000185022, ENSG00000284841, ENSG00000288299, ENSG00000132109, ENSG00000139832, ENSG00000170345, ENSG00000104765, ENSG00000117305, ENSG00000114480, ENSG00000137880, ENSG00000171033, ENSG00000176171, ENSG00000119630, ENSG00000145050, ENSG00000134333, ENSG00000162600, ENSG00000135549, ENSG00000114023, ENSG00000159399, ENSG00000106278, ENSG00000138650, ENSG00000120129, ENSG00000164442, ENSG00000152952, ENSG00000102144, ENSG00000277586, ENSG00000130066, ENSG00000104419, ENSG00000121966, ENSG00000131781, ENSG00000166986, ENSG00000044574, ENSG00000124785,

ENSG00000187837, ENSG00000119950, ENSG00000135245, ENSG00000168209, ENSG00000112715, ENSG00000099860, ENSG00000134107, ENSG00000143847, ENSG00000122884, ENSG00000036448, ENSG00000105835, ENSG00000278677, ENSG00000113083, ENSG00000242612, ENSG00000141837, ENSG00000180573, ENSG00000106031, ENSG00000109107, ENSG00000114268, ENSG00000120937, ENSG00000068028, ENSG00000005961, ENSG00000181218, ENSG00000184678

Detwiller 2005

- ***HGNC***: ADM, ANGPT2, ANXA1, ANXA5, BNIP3L, CD99, COL5A1, ENO1, HIF1A, IGF2, IGFBP3, ITGA5, LDHB, LOX, MIR483, PGF, PGK1, PLOD2, SERPINE1, SERPINE2, SLC2A3, TGFB3, TGFBI, THBS1, TXN, VEGFC, VIM
- ***ENSG ID***: ENSG00000026025, ENSG00000167244, ENSG00000059804, ENSG00000074800, ENSG00000137801, ENSG00000104765, ENSG00000136810, ENSG00000135046, ENSG00000106366, ENSG00000161638, ENSG00000130635, ENSG00000119630, ENSG00000120708, ENSG00000119699, ENSG00000207805, ENSG00000152952, ENSG00000102144, ENSG00000150630, ENSG00000002586, ENSG00000148926, ENSG00000164111, ENSG00000111716, ENSG00000100644, ENSG00000091879, ENSG00000113083, ENSG00000146674, ENSG00000135919

Chi 2006

- ***HGNC***: PLIN2, ADM, ALKBH1, ANGPTL4, ARL6IP1, ARRDC3, BHLHE40, BNIP3L, BRCA2, DEPP1, C14orf132, VTA1, CA9, CAV1, CCL28, CCNG2, CEP250, CKB, CLEC2B, CNTNAP1, CRABP2, CSRP2, CTSF, CXCR4, DDIT4, DPYSL2, DPYSL4, DST, DUSP10, UBE2O, FAM162A, EGLN3, ENPP3, ERO1A, EXOSC4, FAM13A, HAUS6, ANKZF1, GAL3ST1, GNRH1, GPI, GYS1, HEMGN, HIPK1, HK2, HLA-DRB5, IGFBP3, INSIG2, ITPR1, JAG2, KDM3A, UFL1, ARFGEF3, ZNF777, LNPK, LDHA, LDHC, LOX, LPIN2, MEG3, SLC25A42, MKNK2, MPP2, MSH2, NAB2, NARF, NAV2, NDRG1, NF1, OCIAD1, OLR1, P4HA1, P4HA2, PDK1, PFKFB3, PFKP, PGK1, PIAS2, PIK3R4, PLEKHA2, PLOD1, PLOD2, PLXNB3, PMP22, POU5F1, PPFIA4, PPL, PPP5C, PRODH, RAB20, GPRC5A, RNASE4, RTTN, SEMA4B, SERPINE1, SLC16A3, SLC1A6, SLC2A1, SLC2A3, SLC6A3, SLC6A8, SOX4, STC1, STC2, STRBP, SULT4A1, THOP1, WIPF1, WSB1, ZEB2, ZNF292
- ***ENSG ID***: ENSG00000282019, ENSG00000276996, ENSG00000206454, ENSG00000233911, ENSG00000237582, ENSG00000229094, ENSG00000230336, ENSG00000235068, ENSG00000170525, ENSG00000288299, ENSG00000078043, ENSG00000130540, ENSG00000139832, ENSG00000147874, ENSG00000059804, ENSG00000139618, ENSG00000109046, ENSG00000173391, ENSG00000124766, ENSG00000129521, ENSG00000159167, ENSG00000166796, ENSG00000126001, ENSG00000101577, ENSG00000092964, ENSG00000196712, ENSG00000197930, ENSG00000142319, ENSG00000136929, ENSG00000198502, ENSG00000104765, ENSG00000176225, ENSG00000165507, ENSG00000170540, ENSG00000147437, ENSG00000184916, ENSG00000106366, ENSG00000147872, ENSG00000188994, ENSG00000095002, ENSG00000128242, ENSG00000198753, ENSG00000009844, ENSG00000130821, ENSG00000151640, ENSG00000100601, ENSG00000105220, ENSG00000100033, ENSG00000011485, ENSG00000134333, ENSG00000258818,

ENSG00000113369, ENSG00000166165, ENSG00000107159, ENSG00000114023, ENSG00000159399, ENSG00000151882, ENSG00000175183, ENSG00000214548, ENSG00000196453, ENSG00000138640, ENSG00000112379, ENSG00000104812, ENSG00000165209, ENSG00000181035, ENSG00000163349, ENSG00000138764, ENSG00000154269, ENSG00000152952, ENSG00000102144, ENSG00000125629, ENSG00000227051, ENSG00000104419, ENSG00000204531, ENSG00000121966, ENSG00000148926, ENSG00000067057, ENSG00000178896, ENSG00000144320, ENSG00000169499, ENSG00000168209, ENSG00000105974, ENSG00000167772, ENSG00000166833, ENSG00000143507, ENSG00000166886, ENSG00000134107, ENSG00000143847, ENSG00000196455, ENSG00000122884, ENSG00000172009, ENSG00000109180, ENSG00000174080, ENSG00000175931, ENSG00000113083, ENSG00000185033, ENSG00000141562, ENSG00000109099, ENSG00000146674, ENSG00000118898, ENSG00000072682, ENSG00000115548, ENSG00000113739, ENSG00000099875, ENSG00000152256, ENSG00000163516, ENSG00000105143, ENSG00000151914, ENSG00000169554, ENSG00000014123, ENSG00000083444, ENSG00000110852, ENSG00000141526, ENSG00000108852, ENSG00000108797, ENSG00000013588, ENSG00000117394, ENSG00000150995, ENSG00000115935, ENSG00000143320

Mense 2006

- ***HGNC***: ABCB6, PLIN2, ADM, ALDOC, ANG, ANGPTL4, ATF3, BHLHE40, BHLHE41, BNIP3, BNIP3L, CCNI, CDK19, CEBPB, CEBPD, CLK3, DIDO1, SERGEF, EEF1AKMT3, DPCD, KLHL24, FAM162A, EGLN1, ELL2, CTTN, ENO2, ERO1A, FAM13A, LARP6, WDR54, FOXD1, FUT11, GADD45A, GBE1, GNA13, GOSR2, GPI, HCFC1R1, HILPDA, HK2, INSIG1, INSIG2, KDM3A, KDM4B, KDM4C, KLF4, KLF7, LOX, LRP2BP, MAFF, MED6, KIAA2013, LCORL, TC2N, MXI1, NADSYN1, NDRG1, NFIL3, NOL3, OSMR, P4HA1, P4HA2, NAMPT, PDK1, PEX13, PFKFB3, PFKFB4, PGK1, PLOD2, PPFIA4, PPP1R15A, PPP1R3C, PRPSAP1, JMJD6, RBPJ, RIOK3, RNASE4, RORA, SCD, CLPB, SLC2A14, SOD2, SPAG4, STARD4, STC2, STK4, THAP8, TIPARP, TSLP, UFM1, VEGFA, WDR5B, WSB1, ZBTB25
- ***ENSG ID***: ENSG00000276776, ENSG00000282019, ENSG00000120686, ENSG00000185022, ENSG00000101782, ENSG00000172216, ENSG00000170525, ENSG00000288401, ENSG00000101109, ENSG00000129158, ENSG00000109046, ENSG00000127663, ENSG00000119938, ENSG00000197930, ENSG00000165030, ENSG00000104765, ENSG00000099194, ENSG00000166171, ENSG00000114480, ENSG00000136826, ENSG00000147872, ENSG00000145623, ENSG00000101191, ENSG00000176171, ENSG00000145777, ENSG00000162928, ENSG00000164211, ENSG00000105220, ENSG00000221869, ENSG00000061656, ENSG00000196981, ENSG00000258818, ENSG00000155111, ENSG00000214274, ENSG00000108433, ENSG00000116717, ENSG00000114023, ENSG00000159399, ENSG00000166173, ENSG00000069667, ENSG00000186480, ENSG00000138640, ENSG00000103145, ENSG00000107077, ENSG00000152952, ENSG00000118985, ENSG00000102144, ENSG00000163659, ENSG00000112096, ENSG00000114796, ENSG00000125629, ENSG00000087074, ENSG00000123095, ENSG00000196968, ENSG00000104419, ENSG00000089775, ENSG00000165929, ENSG00000148926, ENSG00000123427,

ENSG00000120063, ENSG00000119950, ENSG00000178177, ENSG00000133997, ENSG00000135245, ENSG00000112715, ENSG00000167772, ENSG00000118816, ENSG00000140939, ENSG00000111674, ENSG00000118263, ENSG00000134107, ENSG00000143847, ENSG00000122884, ENSG00000135766, ENSG00000162772, ENSG00000105835, ENSG00000161542, ENSG00000251493, ENSG00000113083, ENSG00000168214, ENSG00000179335, ENSG00000109771, ENSG00000072682, ENSG00000085733, ENSG00000005448, ENSG00000070495, ENSG00000115548, ENSG00000161277, ENSG00000109107, ENSG00000113739, ENSG00000173262, ENSG00000152256, ENSG00000115657, ENSG00000114268, ENSG00000116685, ENSG00000172890, ENSG00000162129

Elvidge 2006

- ***HGNC***: PLIN2, ADM, ADORA2B, AK3, AK4, AKAP12, ALDOC, ANGPTL4, SH2B2, ARTN, ASPH, ATF3, B3GNT4, BBX, BCOR, BHLHE40, BNIP3, BNIP3L, DST, AHNAK2, RBCK1, NREP, CA9, CAV1, CCNG2, CD59, CSGALNACT1, CITED2, KLF6, CSRP2, CXCR4, CYB5A, CYP1A1, CYP1B1, CCN1, PXDN, DAAM1, DDIT4, DDR1, DHRS3, ZNF395, DPYSL2, DPYSL4, KLHL24, DSC2, DTNA, FAM162A, EFNA1, EFNA3, EGFR, EGR1, ELF3, ADGRE2, ENO2, ERO1A, FAM13A, TMEM45A, YEATS2, ANKZF1, ZNF654, SRD5A3, FLNB, FOS, FYN, GADD45B, GBE1, GDAP1L1, GDF15, STBD1, GJA1, GLRX, GPR87, GYS1, HCFC1R1, HEY1, HILPDA, HK2, HLA-DRB3, HOXA7, FAM216A, IGFBP3, IGFBP5, CADM1, ILVBL, INHA, INSIG2, ISG20, ITGB4, ITPR1, KDM3A, KDM4B, JUN, OBSL1, ATG14, CEMIP, KLF7, KRT15, KRT7, LOX, LOXL1, LOXL2, MAGED4, MET, ORAI3, NDRG1, NFIL3, NOL3, NR3C1, OPN3, P4HA1, P4HA2, PAM, PTTG1IP, PDGFB, PDGFRL, PDK1, PFKFB3, PFKP, PGK1, PGM1, EGLN1, EGLN3, PHLDA1, PIM1, PLAC8, PLAUR, POLR2J2, PPFIA4, PRKCA, ZMYND8, PRRX1, PTPRO, QSOX1, GPRC5A, RBPJ, RLF, RNASE4, RRAGD, S100A2, S100A4, S100A6, SAMD4A, SAT1, ATXN1, SCARB1, SCNN1B, SERPINE1, SH3GL3, SLC2A1, SLCO4A1, SORL1, SOX9, SPAG4, SPOCK1, SRCAP, SRPX, STC1, STC2, TBC1D3, TGFBI, TIPARP, TMEFF1, TNFAIP8, TRA2A, TXNIP, UPK1A, VEGFA, VEGFC, VLDLR, CCN5, WSB1, YPEL1, ZNF292
- ***ENSG ID***: ENSG00000230463, ENSG00000223680, ENSG00000229767, ENSG00000196101, ENSG00000234078, ENSG00000215522, ENSG00000231679, ENSG00000137332, ENSG00000230456, ENSG00000101955, ENSG00000067082, ENSG00000101040, ENSG00000170525, ENSG00000100311, ENSG00000147853, ENSG00000183255, ENSG00000104213, ENSG00000147408, ENSG00000109046, ENSG00000127663, ENSG00000164548, ENSG00000151490, ENSG00000129521, ENSG00000170345, ENSG00000159167, ENSG00000092964, ENSG00000168447, ENSG00000127507, ENSG00000197930, ENSG00000162496, ENSG00000105135, ENSG00000085063, ENSG00000134013, ENSG00000186918, ENSG00000165030, ENSG00000104765, ENSG00000080603, ENSG00000124194, ENSG00000130508, ENSG00000137193, ENSG00000166347, ENSG00000134769, ENSG00000114480, ENSG00000185567, ENSG00000106366, ENSG00000198363, ENSG00000100027, ENSG00000134755, ENSG00000138061, ENSG00000241697, ENSG00000147872, ENSG00000171346, ENSG00000188994, ENSG00000164683, ENSG00000173221, ENSG00000152661, ENSG00000101187, ENSG00000011422, ENSG00000140465,

ENSG00000176171, ENSG00000117000, ENSG00000151640, ENSG00000114439, ENSG00000125826, ENSG00000025039, ENSG00000129038, ENSG00000020577, ENSG00000105668, ENSG00000061656, ENSG00000064205, ENSG00000137642, ENSG00000258818, ENSG00000162433, ENSG00000134986, ENSG00000107159, ENSG00000120708, ENSG00000103888, ENSG00000114023, ENSG00000159399, ENSG00000154229, ENSG00000131016, ENSG00000175938, ENSG00000175183, ENSG00000138640, ENSG00000140600, ENSG00000104812, ENSG00000079739, ENSG00000138764, ENSG00000183337, ENSG00000103145, ENSG00000145779, ENSG00000164442, ENSG00000010810, ENSG00000100592, ENSG00000102144, ENSG00000113580, ENSG00000163659, ENSG00000150630, ENSG00000114796, ENSG00000125629, ENSG00000130066, ENSG00000135480, ENSG00000104419, ENSG00000121966, ENSG00000115461, ENSG00000148926, ENSG00000067057, ENSG00000154545, ENSG00000152377, ENSG00000163872, ENSG00000176383, ENSG00000130513, ENSG00000135245, ENSG00000228049, ENSG00000116260, ENSG00000126775, ENSG00000168209, ENSG00000105974, ENSG00000105976, ENSG00000147852, ENSG00000112715, ENSG00000167772, ENSG00000140939, ENSG00000111674, ENSG00000181458, ENSG00000265972, ENSG00000099860, ENSG00000125398, ENSG00000118263, ENSG00000163435, ENSG00000116132, ENSG00000134107, ENSG00000143847, ENSG00000172183, ENSG00000122884, ENSG00000135766, ENSG00000162772, ENSG00000160999, ENSG00000145287, ENSG00000182985, ENSG00000113083, ENSG00000168214, ENSG00000128039, ENSG00000146674, ENSG00000072682, ENSG00000132470, ENSG00000146648, ENSG00000122592, ENSG00000118804, ENSG00000120738, ENSG00000115548, ENSG00000109107, ENSG00000113739, ENSG00000143590, ENSG00000169242, ENSG00000175105, ENSG00000152256, ENSG00000073060, ENSG00000145730, ENSG00000124788, ENSG00000204580, ENSG00000163516, ENSG00000204856, ENSG00000151914, ENSG00000139289, ENSG00000142871, ENSG00000013588, ENSG00000136068, ENSG00000117394, ENSG00000150995, ENSG00000124006, ENSG00000123999, ENSG00000138271, ENSG00000117407, ENSG00000177606, ENSG00000170425, ENSG00000274611, ENSG00000054277, ENSG00000197956, ENSG00000196154, ENSG00000196754

Peters 2006

- ***HGNC***: ADSS2, AK1, AKR1C3, AIFM2, ANGPTL4, ATG4A, APLP2, APP, ARHGDIB, ARID4A, MANF, ASXL1, ATP2A2, ATP5MC1, BACE2, BCL7B, BHLHE40, BNIP3L, BPI, MYG1, TIGAR, HSPB11, GCFC2, CNRIP1, CALR, CAMLG, CANX, CCT6A, CDH5, CERK, CFHR1, RCC1, CLTA, CORIN, CTDSP2, CCN2, CTHRC1, CXCR4, PCBD2, DCTN2, DDIT4, DDX3X, DDX47, DEGS1, DNAJB1, UBE2O, EFEMP1, EIF4A2, EWSR1, EXTL3, IFI27L2, FEN1, FKBP4, FLI1, PRPF38B, PRPF38A, PIH1D1, QSER1, GAPDH, GGH, GMPPB, GSTP1, GTF3C1, HBG2, HLA-A, HM13, HMG20B, HMGA1, HMMR, HSPA5, HSPA8, HSPE1, HSPH1, IARS1, IGFBP7, ITGA10, JUN, TUBA1B, ERGIC1, PPP4R3B, KRT10, LDB1, LOXL2, LSM6, LTBR, MALAT1, MEF2B, MKI67, MORF4L1, MRPL32, MVP, NAP1L1, NDFIP1, NDST3, NFKBIE, NOP10, NQO1, NSFL1C, OAZ1, OPTN, PAFAH2, PCYOX1, PDCD4, PITPNA, PLOD2, POLR2F, PPIF, PPOX, PRDX1, PRDX6, PRKAR1A, PROCR, PSMB2, WDR83OS, PTGIS, PXN, RAB11A, RAB5A, RAC1, IPO5, RCL1, RHOD,

RPL13A, RPL14, RPL3, RPL39, RPL8, RPLP0, RPS11, RPS3, RPS8, RTN4, SAP18, SIGLEC8, SNRNP70, SPAG4, SPTAN1, SRPX, SSR3, STAT3, AURKA, STX10, SUMO3, TCF19, ALYREF, TOR3A, TPD52L2, TPI1, TPT1, TRAP1, TUBA1C, TUBB2A, TWIST2, TXNDC5, PDIA6, UQCRH, WDR43, WWTR1

- ***ENSG ID***: ENSG00000224472, ENSG00000206455, ENSG00000227715, ENSG00000206505, ENSG00000224941, ENSG00000223980, ENSG00000229215, ENSG00000224320, ENSG00000276879, ENSG00000233890, ENSG00000224379, ENSG00000231834, ENSG00000150459, ENSG00000235657, ENSG00000120694, ENSG00000142192, ENSG00000234674, ENSG00000065150, ENSG00000283777, ENSG00000171456, ENSG00000196139, ENSG00000101955, ENSG00000100316, ENSG00000123240, ENSG00000184900, ENSG00000288335, ENSG00000100422, ENSG00000101425, ENSG00000182240, ENSG00000124212, ENSG00000133112, ENSG00000120158, ENSG00000004478, ENSG00000239264, ENSG00000101844, ENSG00000104904, ENSG00000206503, ENSG00000188846, ENSG00000146731, ENSG00000106635, ENSG00000101294, ENSG00000134013, ENSG00000104765, ENSG00000144566, ENSG00000119632, ENSG00000137309, ENSG00000101000, ENSG00000013364, ENSG00000251562, ENSG00000101150, ENSG00000150593, ENSG00000213999, ENSG00000158006, ENSG00000100142, ENSG00000175215, ENSG00000077235, ENSG00000182944, ENSG00000145244, ENSG00000173156, ENSG00000148773, ENSG00000126067, ENSG00000146232, ENSG00000173660, ENSG00000115380, ENSG00000115310, ENSG00000116005, ENSG00000087586, ENSG00000145050, ENSG00000275052, ENSG00000061656, ENSG00000215301, ENSG00000088833, ENSG00000181019, ENSG00000109971, ENSG00000126602, ENSG00000134748, ENSG00000164932, ENSG00000164100, ENSG00000005436, ENSG00000149273, ENSG00000185787, ENSG00000132570, ENSG00000198918, ENSG00000196565, ENSG00000131507, ENSG00000032219, ENSG00000089157, ENSG00000151702, ENSG00000152952, ENSG00000103769, ENSG00000106992, ENSG00000113719, ENSG00000111348, ENSG00000164167, ENSG00000137310, ENSG00000121966, ENSG00000119865, ENSG00000042286, ENSG00000182117, ENSG00000198728, ENSG00000143127, ENSG00000122705, ENSG00000159199, ENSG00000044574, ENSG00000105366, ENSG00000168496, ENSG00000072571, ENSG00000197694, ENSG00000137563, ENSG00000168209, ENSG00000163453, ENSG00000106591, ENSG00000167772, ENSG00000108179, ENSG00000123416, ENSG00000233608, ENSG00000104872, ENSG00000142534, ENSG00000111669, ENSG00000187109, ENSG00000060749, ENSG00000104852, ENSG00000078237, ENSG00000174238, ENSG00000012232, ENSG00000117592, ENSG00000179776, ENSG00000132002, ENSG00000186283, ENSG00000136238, ENSG00000142541, ENSG00000105583, ENSG00000134107, ENSG00000161016, ENSG00000196305, ENSG00000114850, ENSG00000018408, ENSG00000163811, ENSG00000173540, ENSG00000175931, ENSG00000156976, ENSG00000164615, ENSG00000111321, ENSG00000137267, ENSG00000139637, ENSG00000035687, ENSG00000104915, ENSG00000111640, ENSG00000179218, ENSG00000127022, ENSG00000175203, ENSG00000118523, ENSG00000084234, ENSG00000143870, ENSG00000064961, ENSG00000134186, ENSG00000167553, ENSG00000168610, ENSG00000084207, ENSG00000186395, ENSG00000143224, ENSG00000089159, ENSG00000183684, ENSG00000142937, ENSG00000213782, ENSG00000174437, ENSG00000108946,

ENSG00000115541, ENSG00000143753, ENSG00000177606, ENSG00000117450, ENSG00000081870, ENSG00000244414, ENSG00000180198

Aprelikova 2006

- ***HGNC***: SYDE1, ABCB6, PLIN2, ADM, ADORA2B, NCKIPSD, AK2, AK3, ALDOC, ANGPTL4, RIPK4, KLK3, ASPH, B3GNT4, BBX, BCKDK, MALL, BHLHE40, BLCAP, BNIP3, BNIP3L, FOXN3, MRGBP, PDXK, NREP, CA4, CABIN1, CAV1, CCNG2, CFDP1, FOXN3, CHSY1, CITED2, CKB, CLK1, CLK3, CNNM1, CNOT8, CRKL, TLCD3A, RTL8C, PXDN, DAAM1, DARS1, DDX41, DHX40, C15orf39, ZNF395, EEF1AKMT3, RNF208, DNAJB2, DNM2, CDK2AP2, DPYSL4, DSCAM, DSP, DUSP3, FAM162A, EFNA1, EFNA3, EGLN1, EGLN3, EHD2, EI24, EIF4EBP1, ELF3, CTTN, ENDOG, ENO2, ERO1A, F8A1, FAM13A, FER1L4, TMEM45A, YEATS2, ANKZF1, TSR1, NAT10, COA1, ZNF654, RSBN1, MAP6D1, EFHD1, OXSM, ATG9A, FOSL2, FOS, FOSL2, FOXD1, FZD1, FZD7, GADD45B, KDM4C, GBE1, STBD1, TNFAIP8, GGA2, GLG1, GLRX, GPI, SPSB2, GRPEL1, GYS1, HEY1, HILPDA, H4C8, HK1, HK2, HLA-B, HLA-C, HLA-DRB3, HLA-E, HLA-G, HMG20B, HOXA4, FAM216A, IGFBP3, ILVBL, INHBB, INSIG2, JAG2, KDM3A, JUND, VGLL4, OBSL1, KDM4B, LDHA, LDLR, LGALS8, LOXL1, LOXL2, AGPAT5, MAPK7, MAPT, PRR7, ORAI3, ZBTB17, MLH3, MPI, MT1X, MXI1, N4BP3, NARF, NDRG1, NDUFB8, NFE2L1, NOL3, NR2F2, P4HA1, P4HA2, PAM, PAQR6, NAMPT, ZNF395, PDK1, PDLIM2, ECI2, PES1, PEX3, PFKFB3, PFKFB4, PFKP, PGK1, PGM1, PGRMC2, PIGA, PITPNC1, PKD1, PKM, GDF15, PLAC8, PLOD1, PLOD2, KDM5B, PPME1, PRRX1, PNRC1, PPFIA4, PPP1R3C, PPP2R5B, ZMYND8, LONP1, CAVIN1, PYGL, QSOX1, RAB20, RAB3A, RNH1, GPRC5A, RASSF1, RBPJ, RDH11, RHBDL2, RLF, RRAGA, RRAGD, DDIT4, SAMD4A, SAP30, SAV1, SCARB1, SCD, NPTN, SLCO4A1, SLC35E1, SMYD2, SNAPC1, SOX12, SOX9, SPAG4, STC1, STC2, TBC1D3, THBS4, TNFAIP3, TNIP1, TRAM2, MED24, TRIB2, RNF216, TRIM52, VEGFA, VEGFB, VLDLR, WASF2, CCN5, WSB1, ZFP36L1, RNF113A, ZNF292
- ***ENSG ID***: ENSG00000276230, ENSG00000276051, ENSG00000281670, ENSG00000233095, ENSG00000206450, ENSG00000230463, ENSG00000276155, ENSG00000277956, ENSG00000206493, ENSG00000282019, ENSG00000235680, ENSG00000206435, ENSG00000223532, ENSG00000228964, ENSG00000225201, ENSG00000233841, ENSG00000196101, ENSG00000236632, ENSG00000230413, ENSG00000230254, ENSG00000224608, ENSG00000235346, ENSG00000231679, ENSG00000229252, ENSG00000232126, ENSG00000237022, ENSG00000206452, ENSG00000233904, ENSG00000237216, ENSG00000228299, ENSG00000225691, ENSG00000206506, ENSG00000284980, ENSG00000160209, ENSG00000101040, ENSG00000170525, ENSG00000288401, ENSG00000288299, ENSG00000171587, ENSG00000183421, ENSG00000147853, ENSG00000008710, ENSG00000139832, ENSG00000166619, ENSG00000196365, ENSG00000144560, ENSG00000120913, ENSG00000109046, ENSG00000127663, ENSG00000129521, ENSG00000170345, ENSG00000159167, ENSG00000119938, ENSG00000088340, ENSG00000135372, ENSG00000105137, ENSG00000197930, ENSG00000187840, ENSG00000105135, ENSG00000134013, ENSG00000109519, ENSG00000186918, ENSG00000104765, ENSG00000065308, ENSG00000099194, ENSG00000184916, ENSG00000127526,

ENSG00000130508, ENSG00000158195, ENSG00000114480, ENSG00000198363, ENSG00000099991, ENSG00000156642, ENSG00000178802, ENSG00000147872, ENSG00000188994, ENSG00000008838, ENSG00000165195, ENSG00000164683, ENSG00000173221, ENSG00000075426, ENSG00000100029, ENSG00000108861, ENSG00000158315, ENSG00000101187, ENSG00000101189, ENSG00000176171, ENSG00000117000, ENSG00000119946, ENSG00000151640, ENSG00000114439, ENSG00000151748, ENSG00000100504, ENSG00000167173, ENSG00000119684, ENSG00000105220, ENSG00000025039, ENSG00000129038, ENSG00000099942, ENSG00000020577, ENSG00000034693, ENSG00000061656, ENSG00000134333, ENSG00000064205, ENSG00000153774, ENSG00000177732, ENSG00000166165, ENSG00000134986, ENSG00000113296, ENSG00000149547, ENSG00000114023, ENSG00000108406, ENSG00000159399, ENSG00000154217, ENSG00000175938, ENSG00000103507, ENSG00000164040, ENSG00000187193, ENSG00000212864, ENSG00000146278, ENSG00000138640, ENSG00000082641, ENSG00000145911, ENSG00000068971, ENSG00000118503, ENSG00000104812, ENSG00000079739, ENSG00000125352, ENSG00000090863, ENSG00000138764, ENSG00000185551, ENSG00000145779, ENSG00000164442, ENSG00000107077, ENSG00000152952, ENSG00000100592, ENSG00000134590, ENSG00000288722, ENSG00000024422, ENSG00000102144, ENSG00000142515, ENSG00000130164, ENSG00000072042, ENSG00000125629, ENSG00000103365, ENSG00000104419, ENSG00000053254, ENSG00000155876, ENSG00000148926, ENSG00000067057, ENSG00000167434, ENSG00000163872, ENSG00000123427, ENSG00000176383, ENSG00000011275, ENSG00000115866, ENSG00000163083, ENSG00000105649, ENSG00000130522, ENSG00000130513, ENSG00000164105, ENSG00000119950, ENSG00000156515, ENSG00000135245, ENSG00000131188, ENSG00000183258, ENSG00000116260, ENSG00000168209, ENSG00000157240, ENSG00000105974, ENSG00000147852, ENSG00000112715, ENSG00000167772, ENSG00000167136, ENSG00000140939, ENSG00000180834, ENSG00000111674, ENSG00000181458, ENSG00000166136, ENSG00000185650, ENSG00000204592, ENSG00000151093, ENSG00000204525, ENSG00000099860, ENSG00000125398, ENSG00000143499, ENSG00000131873, ENSG00000117139, ENSG00000163435, ENSG00000116132, ENSG00000214517, ENSG00000134107, ENSG00000143847, ENSG00000122884, ENSG00000135766, ENSG00000067225, ENSG00000145901, ENSG00000023608, ENSG00000105835, ENSG00000155189, ENSG00000234745, ENSG00000251493, ENSG00000145287, ENSG00000141562, ENSG00000168214, ENSG00000173511, ENSG00000146674, ENSG00000155508, ENSG00000166484, ENSG00000179335, ENSG00000072682, ENSG00000085733, ENSG00000197576, ENSG00000118804, ENSG00000115548, ENSG00000204632, ENSG00000198721, ENSG00000109107, ENSG00000106603, ENSG00000113739, ENSG00000143590, ENSG00000169242, ENSG00000175105, ENSG00000167721, ENSG00000152256, ENSG00000111671, ENSG00000064961, ENSG00000073060, ENSG00000145730, ENSG00000183718, ENSG00000079805, ENSG00000096696, ENSG00000158406, ENSG00000115657, ENSG00000198925, ENSG00000163516, ENSG00000204856, ENSG00000177469, ENSG00000114268, ENSG00000167797, ENSG00000213672, ENSG00000083444, ENSG00000144063, ENSG00000186868, ENSG00000023191, ENSG00000167695, ENSG00000013588, ENSG00000068028, ENSG00000124006, ENSG00000160781, ENSG00000115468,

ENSG00000116977, ENSG00000170425, ENSG00000274611, ENSG00000004455, ENSG00000155760, ENSG00000013441, ENSG00000135924, ENSG00000071575, ENSG00000081019, ENSG00000116809

Bosco 2006

- ***HGNC***: ACVR1B, ADAM8, ADM, ADORA2B, AGPAT5, AK3, AK4, ALDOA, ALDOC, APOBR, APOC4, ATF2, ATF5, BACH1, BCAT1, BCKDHA, BHLHE40, BLZF1, BNIP3, BNIP3L, CD93, CA12, CALCRL, CCL2, CCL20, CCNG2, CCR5, CD28, CD55, CD55, CD59, CD69, CD79A, CD86, CELSR3, CKLF, CD300C, CNR1, COL7A1, COLEC12, CSF1, CSF3, CTSC, CTSD, CX3CR1, CXCL2, CXCL3, CXCL5, CXCR4, DLG1, DSC3, DSP, EGLN1, EGR1, EGR3, ENO1, ENO2, ERO1A, F3, FA2H, FABP4, FCAR, FCGR2A, FCGR2B, FCGR3A, FGF1, FLNB, FLT1, FN1, FOSB, FOSL2, GAPDH, GBE1, GPI, HAS1, HILPDA, HK2, HRH4, HSF2, ICAM5, ID1, ID2, IL1A, IL36G, IL1RAP, IL1RN, IL23A, IL4, IL6, IL6ST, INHBA, IRAK3, IRF4, KDM3A, JUN, KIR3DL2, KLF10, KLRF1, KRTAP4-7, LAMB3, LDHA, LGALS8, MARCO, MCL1, MET, MIF, MMP1, MMP16, MMP19, MSR1, MXI1, NDRG1, NFYA, NPM1, NR2C1, NR4A2, OSBP, OSBPL10, P4HA1, P4HA2, PAM, PANX1, PARVB, PCDHGC3, PDK1, PFKFB4, PFKP, PGAM1, PGK1, PIK3CB, PIK3R1, PIP5K1A, PKD2, PLCL2, PTGIS, PTGS2, RAB7A, ACKR3, RGS1, RUNX1, RUNX2, SCARF1, SEMA4C, SEMA4D, SEMA4F, SLC2A1, SLC2A3, SMTN, SNAPC1, SNTA1, SP4, SPP1, TAF9B, TCF7L2, MLX, TFPI, TFPI2, TGFBR1, TRIM24, TNF, TNFRSF10B, TNFRSF11B, TNFRSF21, TNFRSF9, TNFSF14, TNFSF15, TNS1, TPI1, TREM1, VEGFA, VLDLR, WNT5A, WSB1
- ***ENSG ID***: ENSG00000273735, ENSG00000276858, ENSG00000276424, ENSG00000273738, ENSG00000276985, ENSG00000277982, ENSG00000275511, ENSG00000274722, ENSG00000278850, ENSG00000276004, ENSG00000278474, ENSG00000278710, ENSG00000278442, ENSG00000278758, ENSG00000275626, ENSG00000275083, ENSG00000275136, ENSG00000275262, ENSG00000275970, ENSG00000275269, ENSG00000278726, ENSG00000275416, ENSG00000276882, ENSG00000275629, ENSG00000276357, ENSG00000278809, ENSG00000278656, ENSG00000278403, ENSG00000278361, ENSG00000277709, ENSG00000275838, ENSG00000228978, ENSG00000278415, ENSG00000278707, ENSG00000276739, ENSG00000276701, ENSG00000276336, ENSG00000273911, ENSG00000275564, ENSG00000274580, ENSG00000282019, ENSG00000204490, ENSG00000275566, ENSG00000223952, ENSG00000277181, ENSG00000228849, ENSG00000228321, ENSG00000158270, ENSG00000230108, ENSG00000102755, ENSG00000284017, ENSG00000125810, ENSG00000206439, ENSG00000284466, ENSG00000284046, ENSG00000283750, ENSG00000156273, ENSG00000284192, ENSG00000134489, ENSG00000283951, ENSG00000284384, ENSG00000284381, ENSG00000284213, ENSG00000284245, ENSG00000284004, ENSG00000284980, ENSG00000283975, ENSG00000159216, ENSG00000284295, ENSG00000284063, ENSG00000283953, ENSG00000288299, ENSG00000284528, ENSG00000284061, ENSG00000124212, ENSG00000147853, ENSG00000154822, ENSG00000137265, ENSG00000059804, ENSG00000105866, ENSG00000150045, ENSG00000109046, ENSG00000110848, ENSG00000101400, ENSG00000179388, ENSG00000060982, ENSG00000110048, ENSG00000038945, ENSG00000134762, ENSG00000188677, ENSG00000197930,

ENSG00000074800, ENSG00000085063, ENSG00000104765, ENSG00000124813, ENSG00000120889, ENSG00000125968, ENSG00000171314, ENSG00000108691, ENSG00000115738, ENSG00000148737, ENSG00000105825, ENSG00000183963, ENSG00000114480, ENSG00000074410, ENSG00000124731, ENSG00000110218, ENSG00000106799, ENSG00000170323, ENSG00000103089, ENSG00000075426, ENSG00000156103, ENSG00000240972, ENSG00000163735, ENSG00000176171, ENSG00000134352, ENSG00000049249, ENSG00000105369, ENSG00000105220, ENSG00000025156, ENSG00000240403, ENSG00000186431, ENSG00000134333, ENSG00000163734, ENSG00000164761, ENSG00000162433, ENSG00000146072, ENSG00000159399, ENSG00000114013, ENSG00000125740, ENSG00000184730, ENSG00000145675, ENSG00000122779, ENSG00000117525, ENSG00000138764, ENSG00000135622, ENSG00000136688, ENSG00000125735, ENSG00000113578, ENSG00000143398, ENSG00000102144, ENSG00000187325, ENSG00000181634, ENSG00000196083, ENSG00000149925, ENSG00000104419, ENSG00000121966, ENSG00000148926, ENSG00000067057, ENSG00000051382, ENSG00000113520, ENSG00000105376, ENSG00000248098, ENSG00000105509, ENSG00000232810, ENSG00000184371, ENSG00000119950, ENSG00000135245, ENSG00000117475, ENSG00000105976, ENSG00000147852, ENSG00000169136, ENSG00000112715, ENSG00000168329, ENSG00000115966, ENSG00000187764, ENSG00000111674, ENSG00000111669, ENSG00000075785, ENSG00000151651, ENSG00000064989, ENSG00000123342, ENSG00000074660, ENSG00000122641, ENSG00000217555, ENSG00000003436, ENSG00000155090, ENSG00000134107, ENSG00000144476, ENSG00000108342, ENSG00000115008, ENSG00000109861, ENSG00000122884, ENSG00000136244, ENSG00000135766, ENSG00000117984, ENSG00000023608, ENSG00000155189, ENSG00000196878, ENSG00000178562, ENSG00000135503, ENSG00000181163, ENSG00000114251, ENSG00000090104, ENSG00000196611, ENSG00000111640, ENSG00000153234, ENSG00000072682, ENSG00000118762, ENSG00000081041, ENSG00000075711, ENSG00000120738, ENSG00000115548, ENSG00000109107, ENSG00000118432, ENSG00000152256, ENSG00000267467, ENSG00000240184, ENSG00000145730, ENSG00000096696, ENSG00000118785, ENSG00000001167, ENSG00000196352, ENSG00000115009, ENSG00000114268, ENSG00000110944, ENSG00000008300, ENSG00000114270, ENSG00000144645, ENSG00000108788, ENSG00000090376, ENSG00000120798, ENSG00000160791, ENSG00000136068, ENSG00000117394, ENSG00000072694, ENSG00000143226, ENSG00000116977, ENSG00000203747, ENSG00000143384, ENSG00000115414, ENSG00000177606, ENSG00000170425, ENSG00000167850, ENSG00000240871, ENSG00000168758, ENSG00000073756, ENSG00000136689, ENSG00000019169, ENSG00000079308

Shi 2007

- ***HGNC***: B4GALT5, CLCN7, CXCL13, DEDD, DTX2, FBLN2, FGFR4, FMC1, IFIH1, KCNJ5, LMAN1, PBLD, MLC1, MMP2, NCOR2, NDUFS2, PLCH2, POLE2, ACOT8, RAB3A, ROCK2, SCN5A, GPBP1L1, SRP54, STK25, TMEM9, UBA7, VEGFA, GDI1, ZNF33A, ZPBP
- ***ENSG ID***: ENSG00000276429, ENSG00000282379, ENSG00000074695, ENSG00000103249, ENSG00000158470, ENSG00000042813, ENSG00000149527,

ENSG00000100479, ENSG00000189180, ENSG00000163520, ENSG00000182179, ENSG00000108187, ENSG00000203879, ENSG00000101473, ENSG00000087245, ENSG00000100427, ENSG00000164898, ENSG00000196498, ENSG00000120457, ENSG00000100883, ENSG00000160867, ENSG00000105649, ENSG00000112715, ENSG00000115694, ENSG00000134318, ENSG00000091073, ENSG00000183873, ENSG00000156234, ENSG00000158796, ENSG00000158864, ENSG00000115267, ENSG00000159592, ENSG00000116857

Sung 2007

- ***HGNC***: ADM, AK3, ALDOC, ARRDC3, BHLHE40, BNIP3, BNIP3L, BTG1, DEPP1, OLMALINC, CA9, CCNG2, CDKN1C, CLEC2B, CLK1, CLK3, CNOT8, DDIT4, DHRS3, TSC22D3, DUSP1, FAM162A, EDN2, EFNA1, EFNA3, EGLN3, ELF3, ENO2, ERO1A, FAM13A, TMEM45A, WDR54, FOS, FOSL2, FOXD1, FOXO3, FUT11, GADD45B, GOLGA8A, HES1, HEY1, HILPDA, HK2, IER3, IGFBP3, INSIG1, INSIG2, ITGA5, KDM3A, KDM4B, LOX, LOXL2, LRP2BP, MAFF, MAFK, ERRFI1, MOB3A, MUC1, MXI1, NDRG1, NFIL3, NOL3, P4HA2, PAG1, NAMPT, PDK1, PFKFB3, PFKFB4, PGK1, PLOD2, RAB20, RAB40C, GPRC5A, RARA, SEMA4B, SERPINE1, SERTAD2, SLC2A1, SLC2A14, SLC2A3, SLC6A8, STC1, STC2, TBC1D3, TIPARP, TPBG, TRRAP, WSB1, ZNF292, ZNF395
- ***ENSG ID***: ENSG00000273707, ENSG00000235030, ENSG00000237155, ENSG00000230128, ENSG00000227231, ENSG00000283085, ENSG00000206478, ENSG00000185022, ENSG00000170525, ENSG00000197562, ENSG00000147853, ENSG00000139832, ENSG00000059804, ENSG00000172081, ENSG00000198517, ENSG00000109046, ENSG00000127663, ENSG00000129521, ENSG00000170345, ENSG00000159167, ENSG00000116285, ENSG00000197930, ENSG00000162496, ENSG00000134013, ENSG00000186918, ENSG00000165030, ENSG00000104765, ENSG00000165507, ENSG00000131759, ENSG00000106366, ENSG00000076641, ENSG00000161638, ENSG00000188994, ENSG00000127129, ENSG00000164683, ENSG00000133639, ENSG00000075426, ENSG00000179833, ENSG00000176171, ENSG00000130821, ENSG00000235823, ENSG00000113369, ENSG00000107159, ENSG00000114023, ENSG00000159399, ENSG00000118689, ENSG00000186480, ENSG00000138640, ENSG00000138764, ENSG00000120129, ENSG00000152952, ENSG00000196367, ENSG00000146242, ENSG00000102144, ENSG00000163659, ENSG00000125629, ENSG00000157514, ENSG00000196968, ENSG00000104419, ENSG00000114315, ENSG00000148926, ENSG00000119950, ENSG00000135245, ENSG00000168209, ENSG00000140939, ENSG00000111674, ENSG00000181458, ENSG00000099860, ENSG00000163435, ENSG00000134107, ENSG00000105835, ENSG00000251493, ENSG00000113083, ENSG00000185033, ENSG00000146674, ENSG00000137331, ENSG00000155508, ENSG00000179335, ENSG00000109771, ENSG00000072682, ENSG00000175265, ENSG00000005448, ENSG00000115548, ENSG00000109107, ENSG00000113739, ENSG00000173262, ENSG00000143590, ENSG00000169242, ENSG00000185499, ENSG00000152256, ENSG00000114268, ENSG00000110852, ENSG00000013588, ENSG00000117394, ENSG00000129757, ENSG00000274611, ENSG00000013441

Winter 2007

- ***HGNC***: NTMT1, ADORA2B, AKR7A2P1, AK3, ALDOA, ANGPTL4, ANKRD9, ANLN, B4GALT2, BCAR1, BMS1, BNIP3, SLIRP, HAUS2, MRGBP, CA12, CA9, CDCA4, COL4A5, CORO1C, TNS4, TANC2, DPM2, EIF2S1, GAPDH, GMFB, GSS, HES2, HILPDA, HOMER1, CNIH4, CXCL8, IGF2BP2, KCTD11, TRMT5, KRT17, PEDS1, LDHA, LDLR, LRP2BP, GPN3, C16orf74, ECE2, METTL22, MIF, MNAT1, MRPL14, MRPS17, MTX1, NDRG1, NME1, NUDT15, P4HA1, PAWR, PDZD11, PFKFB4, PGAM1, PGF, PGK1, PLAU, PLEKHG3, PPARD, PPM1J, PPP4R1, PSMA7, PSMB7, PSMD2, PTGFRN, PVR, PYGL, RAN, RNF24, RNPS1, RUVBL2, S100A10, S100A3, ZEB2, SLC16A1, SLC2A1, SLC6A10P, SLC6A8, SLCO1B3, TMTC3, SNX24, SPTB, TEAD4, TFAP2C, TIMM23, TMEM30B, TPBG, TPD52L2, TPI1, TUBB2A, VAPB, VEGFA, VEZT, XPO5
- ***ENSG ID***: ENSG00000276701, ENSG00000281917, ENSG00000280751, ENSG00000283085, ENSG00000285460, ENSG00000288299, ENSG00000240849, ENSG00000288399, ENSG00000136159, ENSG00000147853, ENSG00000101182, ENSG00000124164, ENSG00000182107, ENSG00000205937, ENSG00000100983, ENSG00000087510, ENSG00000020426, ENSG00000126814, ENSG00000011426, ENSG00000197905, ENSG00000165733, ENSG00000154845, ENSG00000101150, ENSG00000067365, ENSG00000214617, ENSG00000156381, ENSG00000171314, ENSG00000239789, ENSG00000170779, ENSG00000137814, ENSG00000074410, ENSG00000128422, ENSG00000188153, ENSG00000136930, ENSG00000131746, ENSG00000177425, ENSG00000073008, ENSG00000240972, ENSG00000101189, ENSG00000176171, ENSG00000050820, ENSG00000130821, ENSG00000119705, ENSG00000100504, ENSG00000119630, ENSG00000197045, ENSG00000134333, ENSG00000101236, ENSG00000122861, ENSG00000169429, ENSG00000154102, ENSG00000107159, ENSG00000152413, ENSG00000028203, ENSG00000120509, ENSG00000170921, ENSG00000064652, ENSG00000139324, ENSG00000148335, ENSG00000146242, ENSG00000134247, ENSG00000136908, ENSG00000134001, ENSG00000102144, ENSG00000130164, ENSG00000132341, ENSG00000155367, ENSG00000149925, ENSG00000104419, ENSG00000265354, ENSG00000111700, ENSG00000135245, ENSG00000180992, ENSG00000124571, ENSG00000112715, ENSG00000167772, ENSG00000073792, ENSG00000111669, ENSG00000126822, ENSG00000070182, ENSG00000175166, ENSG00000183207, ENSG00000122884, ENSG00000110880, ENSG00000069812, ENSG00000137267, ENSG00000112033, ENSG00000111640, ENSG00000109771, ENSG00000145194, ENSG00000239672, ENSG00000111231, ENSG00000169554, ENSG00000114268, ENSG00000117394, ENSG00000213859, ENSG00000143771, ENSG00000170425, ENSG00000155380, ENSG00000229020, ENSG00000117411, ENSG00000173171, ENSG00000197747, ENSG00000188015

Seigneuric 2007 (common)

- ***HGNC***: ACACA, AMH, BACH1, CCT2, GRM3, HMMR, IFI6, KDM4A, NIT1, PAPPA, PEX14, SLC5A12, TIMP2, TTLL5
- ***ENSG ID***: ENSG00000275176, ENSG00000156273, ENSG00000104899, ENSG00000142655, ENSG00000148942, ENSG00000278540, ENSG00000182752,

ENSG00000166226, ENSG00000066135, ENSG00000119685, ENSG00000035862, ENSG00000072571, ENSG00000198822, ENSG00000158793, ENSG00000126709

Seigneuric 2007 (early0)

- ***HGNC***: ACACA, ACOX1, ACSS1, AIF1, AMH, ATF3, DMAC2L, ATXN7L1, BACH1, BET1L, BTBD7, NAA25, PIANP, CCL26, CCNH, CCT2, ACAP2, COL6A3, CPEB4, DUSP3, EIF4EBP2, CIBAR1, GAS6, GATAD2B, GFOD2, GRK6, GRM3, H2AC16, HMMR, IFI6, IGF1R, IGFBP1, IGSF11, IL22RA2, KDM4A, RALGAPB, LY86, MAML2, NCAPH2, NIT1, NSD1, OSTM1, PAPPA, PARG, PCF11, PCSK1, PEX14, PHF10, PIK3R4, PRKAG2, HACD3, RAB4B, RBM4, RBPMS, RHOBTB3, RNASE4, SLC5A12, SOX12, SSH2, ST3GAL1, TERT, TIMP2, TLE3, TRNT1, TTLL5, TTLL10, ZNF117, ZNF664
- ***ENSG ID***: ENSG00000261992, ENSG00000235588, ENSG00000275176, ENSG00000206428, ENSG00000277222, ENSG00000234836, ENSG00000235985, ENSG00000237727, ENSG00000156273, ENSG00000170471, ENSG00000183087, ENSG00000154930, ENSG00000112799, ENSG00000104899, ENSG00000142655, ENSG00000152926, ENSG00000164362, ENSG00000011114, ENSG00000148942, ENSG00000162571, ENSG00000278540, ENSG00000184384, ENSG00000074696, ENSG00000134480, ENSG00000165494, ENSG00000108861, ENSG00000188343, ENSG00000141098, ENSG00000182752, ENSG00000175426, ENSG00000025770, ENSG00000164485, ENSG00000144847, ENSG00000177732, ENSG00000157110, ENSG00000258818, ENSG00000166226, ENSG00000167578, ENSG00000066135, ENSG00000081087, ENSG00000106617, ENSG00000164292, ENSG00000008513, ENSG00000140443, ENSG00000130024, ENSG00000179195, ENSG00000140332, ENSG00000119685, ENSG00000114331, ENSG00000143614, ENSG00000113742, ENSG00000035862, ENSG00000072571, ENSG00000204472, ENSG00000125375, ENSG00000198055, ENSG00000141298, ENSG00000227345, ENSG00000148730, ENSG00000196455, ENSG00000006606, ENSG00000165671, ENSG00000162772, ENSG00000276903, ENSG00000146776, ENSG00000146678, ENSG00000161533, ENSG00000177951, ENSG00000198822, ENSG00000139200, ENSG00000072756, ENSG00000173933, ENSG00000111300, ENSG00000158793, ENSG00000126709, ENSG00000163359

Seigneuric 2007 (early2)

- ***HGNC***: ACACA, ALKBH1, AMH, ATXN1, BACH1, C1GALT1C1, TCAIM, CCT2, CDR2, DTL, DDIAS, GRM3, HMMR, KAT5, ICMT, IFI6, ING1, ITGB4, KDM4A, MAGEA6, MALAT1, MUC1, NIT1, PAPPA, PEX14, SCD, SGK1, SLC5A12, SPAG5, TIMP2, DCAF8
- ***ENSG ID***: ENSG00000275176, ENSG00000156273, ENSG00000288368, ENSG00000153487, ENSG00000104899, ENSG00000142655, ENSG00000197172, ENSG00000179152, ENSG00000148942, ENSG00000099194, ENSG00000251562, ENSG00000278540, ENSG00000140743, ENSG00000116237, ENSG00000182752, ENSG00000100601, ENSG00000166226, ENSG00000066135, ENSG00000171155, ENSG00000035862, ENSG00000072571, ENSG00000132716, ENSG00000143476, ENSG00000172977, ENSG00000132470, ENSG00000165490, ENSG00000198822,

ENSG00000076382, ENSG00000185499, ENSG00000124788, ENSG00000118515, ENSG00000158793, ENSG00000126709

Beyer 2008

- ***HGNC***: ABCB6, ADM, ALDOC, ANG, ANGPTL4, ANKZF1, AP1G1, AQP3, BCKDK, BHLHE40, BNIP3, BNIP3L, BTG1, DEPP1, THEMIS2, C1R, FAM162A, C4orf19, CA11, CA12, CA9, CD24, CDA, CDK19, CEBPD, CIDEB, CKB, COL11A1, CRABP2, CSRP2, CXCR4, ACKR3, DAPK1, DDIT4, DUSP1, EFEMP2, EFNA1, ELL2, ENO2, ENPEP, EPOR, ERO1A, FAM13A, FGB, FGF14, FSTL3, GAD1, GAL3ST1, GBE1, GBP2, GLUL, GOLGA8G, GPI, GYS1, HCFC1R1, HILPDA, H2AC6, H2BC5, H2AC18, H2AC19, HLA-DQB1, HNRNPD, HOXC10, IFITM1, IFITM2, INSIG2, ITPR1, JAM2, KDM3A, KDM4B, JUNB, KCNK3, ERV3-2, KLF9, LBP, LGALS8, LOX, LOXL2, LTBP3, MALL, MOAP1, MPI, MT1M, MT1X, MUC1, MXI1, NDRG1, NOL3, NPEPPS, NRN1, OBSL1, OPN3, P4HA1, P4HA2, CDHR1, CDK18, PDK1, PDK3, PDZK1IP1, PER1, PEX11A, PFKL, PFKP, PGK1, PLAGL1, PLOD1, PLOD2, PMP22, PNMA2, POU5F1P3, POU5F1P4, PPFIA4, PPP1R13L, PPP1R3C, PTEN, RAB20, RAD54B, RALGDS, RLF, RNASE4, RNASET2, RORA, RRAD, SELENOP, SERPINA3, SERPINA5, SERPINE1, SERPING1, SH3BP2, SIPA1L1, SLC16A3, SLC25A36, SLC2A1, SLC2A14, SLC2A3, SLC2A5, SLCO2B1, SMPDL3A, SPAG4, SSBP2, SSPN, STBD1, STC1, TBC1D8, TGFA, TIPARP, TMCC1, TMEM45A, TMEM47, TNS1, TSC22D3, VEGFA, VKORC1, VLDLR, YEATS2, ZFP36, ZNF292, ZNF395
- ***ENSG ID***: ENSG00000273651, ENSG00000277090, ENSG00000282019, ENSG00000278268, ENSG00000206237, ENSG00000231939, ENSG00000231286, ENSG00000154721, ENSG00000284792, ENSG00000225824, ENSG00000206302, ENSG00000233209, ENSG00000147027, ENSG00000070404, ENSG00000288512, ENSG00000102466, ENSG00000141959, ENSG00000139832, ENSG00000285199, ENSG00000059804, ENSG00000127663, ENSG00000235602, ENSG00000129988, ENSG00000159167, ENSG00000142583, ENSG00000119938, ENSG00000183629, ENSG00000197930, ENSG00000134013, ENSG00000186918, ENSG00000179344, ENSG00000104765, ENSG00000165507, ENSG00000087266, ENSG00000196730, ENSG00000240694, ENSG00000148600, ENSG00000154274, ENSG00000180818, ENSG00000114480, ENSG00000074410, ENSG00000171303, ENSG00000106366, ENSG00000178802, ENSG00000188994, ENSG00000272398, ENSG00000133639, ENSG00000128242, ENSG00000165272, ENSG00000176171, ENSG00000117000, ENSG00000137491, ENSG00000165943, ENSG00000138668, ENSG00000250722, ENSG00000188488, ENSG00000105220, ENSG00000196136, ENSG00000067992, ENSG00000221869, ENSG00000061656, ENSG00000197275, ENSG00000145687, ENSG00000258818, ENSG00000155111, ENSG00000172594, ENSG00000214274, ENSG00000166165, ENSG00000107159, ENSG00000114023, ENSG00000167397, ENSG00000103507, ENSG00000175183, ENSG00000069667, ENSG00000205364, ENSG00000187193, ENSG00000118495, ENSG00000138640, ENSG00000104812, ENSG00000163235, ENSG00000136305, ENSG00000114120, ENSG00000120129, ENSG00000103145, ENSG00000128016, ENSG00000152952, ENSG00000060718, ENSG00000118985, ENSG00000102144, ENSG00000163659, ENSG00000125629, ENSG00000157514, ENSG00000104419, ENSG00000121966, ENSG00000148926,

ENSG00000160271, ENSG00000067057, ENSG00000172638, ENSG00000163872, ENSG00000124785, ENSG00000158373, ENSG00000119950, ENSG00000135245, ENSG00000159403, ENSG00000168209, ENSG00000147852, ENSG00000112715, ENSG00000167772, ENSG00000140939, ENSG00000063180, ENSG00000187266, ENSG00000166821, ENSG00000171862, ENSG00000111674, ENSG00000181458, ENSG00000119138, ENSG00000134107, ENSG00000171223, ENSG00000171564, ENSG00000144476, ENSG00000143847, ENSG00000179094, ENSG00000122884, ENSG00000123096, ENSG00000197555, ENSG00000149131, ENSG00000128683, ENSG00000168056, ENSG00000138792, ENSG00000113083, ENSG00000109099, ENSG00000166747, ENSG00000072682, ENSG00000180573, ENSG00000185885, ENSG00000166592, ENSG00000118804, ENSG00000115548, ENSG00000026297, ENSG00000158825, ENSG00000109107, ENSG00000173262, ENSG00000169242, ENSG00000185499, ENSG00000185201, ENSG00000152256, ENSG00000104881, ENSG00000162366, ENSG00000115657, ENSG00000135821, ENSG00000163516, ENSG00000141279, ENSG00000083444, ENSG00000144063, ENSG00000141526, ENSG00000130775, ENSG00000117394, ENSG00000237872, ENSG00000150995, ENSG00000124006, ENSG00000116977, ENSG00000172765, ENSG00000204634, ENSG00000162645, ENSG00000143320, ENSG00000203812, ENSG00000272196, ENSG00000079308, ENSG00000117266, ENSG00000054277

Hu 2009

- ***HGNC***: ADM, ANGPTL4, FLVCR2, DDIT4, FABP5, GAL, NDRG1, ZNF384, PLOD1, RRAGD, SLC16A3, UCHL1, VEGFA
- ***ENSG ID***: ENSG00000154277, ENSG00000069482, ENSG00000164687, ENSG00000025039, ENSG00000119686, ENSG00000104419, ENSG00000148926, ENSG00000168209, ENSG00000112715, ENSG00000167772, ENSG00000083444, ENSG00000141526, ENSG00000126746

Benita 2009

- ***HGNC***: RRAGD, SPAG4, P4HA2, RSBN1, PLOD1, GADD45B, CRKL, PGK1, NDRG1, GYS1, NAMPT, TGFBR1, KDM4C, DHX40, GOSR2, WSB1, ALDOC, GAPDH, LOX, NR3C1, STC2, KDM3A, MXI1, CXCR4, P4HA1, INSIG2, KDM4B, RARA, PER2, SEC61G, LDHA, PIM1, ASCC1, PHLDA1, NARF, KLF10, SLC16A1, ATF3, RAB8B, DDIT4, RBPJ, EFNA1, PGAM1, EIF1, BNIP3, CLK3, TMEM45A, ANKRD37, ERO1A, ASPH, PJA2, VDAC1, PPME1, MIF
- ***ENSG ID***: ENSG00000105835, ENSG00000173812, ENSG00000104419, ENSG00000179335, ENSG00000099942, ENSG00000169242, ENSG00000139289, ENSG00000109107, ENSG00000127663, ENSG00000107077, ENSG00000132432, ENSG00000111640, ENSG00000109046, ENSG00000141562, ENSG00000113580, ENSG00000104812, ENSG00000197930, ENSG00000168214, ENSG00000186352, ENSG00000288299, ENSG00000134333, ENSG00000113083, ENSG00000276701, ENSG00000240972, ENSG00000198363, ENSG00000119950, ENSG00000099860, ENSG00000162772, ENSG00000122884, ENSG00000138303, ENSG00000125629, ENSG00000214517, ENSG00000166128, ENSG00000171314, ENSG00000102144, ENSG00000137193, ENSG00000083444, ENSG00000168209, ENSG00000081019, ENSG00000181458, ENSG00000115548, ENSG00000025039, ENSG00000131759,

ENSG00000281917, ENSG00000155380, ENSG00000176171, ENSG00000061656, ENSG00000106799, ENSG00000155090, ENSG00000213585, ENSG00000121966, ENSG00000108406, ENSG00000113739, ENSG00000132326, ENSG00000072682, ENSG00000108433, ENSG00000198961

Fardin 2009

- ***HGNC***: AK4, ALDOC, BNIP3, DDIT4, FAM162A, PDK1, VEGFA, WDR5B
- ***ENSG ID***: ENSG00000176171, ENSG00000196981, ENSG00000162433, ENSG00000114023, ENSG00000168209, ENSG00000112715, ENSG00000109107, ENSG00000152256

Lendhal 2009

- ***HGNC***: ADM, ALDOC, ATF7IP, BNIP3, BNIP3L, DDIT4, ENO2, ERO1A, KDM3A, GLRX, HILPDA, HK2, HMMR, INSIG2, MXI1, MYCBP, NDRG1, P4HA1, PDK1, PPFIA4, SLC7A1, VEGFA, ZNF654
- ***ENSG ID***: ENSG00000139514, ENSG00000197930, ENSG00000104765, ENSG00000173221, ENSG00000176171, ENSG00000214114, ENSG00000159399, ENSG00000125629, ENSG00000104419, ENSG00000148926, ENSG00000072571, ENSG00000119950, ENSG00000135245, ENSG00000168209, ENSG00000112715, ENSG00000111674, ENSG00000143847, ENSG00000122884, ENSG00000115548, ENSG00000109107, ENSG00000175105, ENSG00000152256, ENSG00000171681

Buffa 2010

- ***HGNC***: ACOT7, ADM, AK4, ALDOA, ANKRD37, ANLN, BNIP3, MRGBP, CA9, CDKN3, CHCHD2, CORO1C, CTSV, DDIT4, ENO1, ESRP1, GAPDH, GPI, HILPDA, HK2, KIF20A, KIF4A, LDHA, LRRC42, MAD2L2, MAP7D1, MCTS1, MIF, MRPL13, MRPL15, MRPS17, NDRG1, ZNF384, P4HA1, PFKP, PGAM1, PGK1, PSMA7, PSRC1, SEC61G, SHCBP1, SLC16A1, SLC25A32, SLC2A1, TPI1, TUBA1B, TUBA1C, TUBB6, UTP11, VEGFA, YKT6
- ***ENSG ID***: ENSG00000276701, ENSG00000282019, ENSG00000281917, ENSG00000288299, ENSG00000288295, ENSG00000090889, ENSG00000101182, ENSG00000011426, ENSG00000176014, ENSG00000074800, ENSG00000132432, ENSG00000171314, ENSG00000239789, ENSG00000171241, ENSG00000240972, ENSG00000183520, ENSG00000101189, ENSG00000176171, ENSG00000105220, ENSG00000100526, ENSG00000134333, ENSG00000106153, ENSG00000116212, ENSG00000104413, ENSG00000162433, ENSG00000164933, ENSG00000107159, ENSG00000172172, ENSG00000137547, ENSG00000159399, ENSG00000232119, ENSG00000116670, ENSG00000134222, ENSG00000112984, ENSG00000102144, ENSG00000149925, ENSG00000104419, ENSG00000186352, ENSG00000148926, ENSG00000067057, ENSG00000136943, ENSG00000135245, ENSG00000168209, ENSG00000112715, ENSG00000123416, ENSG00000111669, ENSG00000106636, ENSG00000122884, ENSG00000110880, ENSG00000111640, ENSG00000167553, ENSG00000117394, ENSG00000126746, ENSG00000097021, ENSG00000155380, ENSG00000116871

Ghorbel 2010

- ***HGNC***: ABRA, ACTL6A, ACTR3, ADK, ALS2, ANK1, ANXA7, AP1S2, APOBEC3G, ATE1, ATP2C1, B3GALNT1, BCL10, BLZF1, BPNT1, BTF3, IDI2-AS1, TIMM21, NADK2, TCIM, CALM3, CALR, CBWD1, CD8A, CDKN2C, CEP170, CHORDC1, CNN3, COMMD10, COMMD8, COX11, CPE, CPNE4, CYP51A1, RBM48, DPYD, S1PR1, EFCAB2, EFEMP1, MICU2, MICU3, EGFR, EGR1, EIF1AXP1, EIF1AX, EIF4E, EPOR, TCAF1, MIGA1, FAM76B, FCF1, FGD4, FGF7, FGL2, FLCN, TRAPPC13, OGFOD3, GAPDHP73, GATM, GJB4, GNG10, GOLT1B, GULP1, HNRNPH1, HSDL2, IGFBP7, IL33, ITM2A, JAG1, JRKL, TTC37, UFL1, KLHL20, KPNA2, LACTB2, LCE1E, LRRC25, LUM, LYPLA1, LYRM7, LYZ, MAD2L1, MANSC1, MEFV, MRPL50, MTHFD2L, MYOZ2, NAP1L1, NDUFA5, NFATC2IP, NIPSNAP3A, NLRP1, OMA1, OTUD6B, PAIP1, PAK2, PEX7, PHACTR2, BLOC1S6, POLR2K, POMP, PPM1B, PPP1CB, PSMC2, PTP4A1, RAB12, RABGGTB, RAP2C, RBM3, RPL15, RPL23AP7, RPL37A, RPL38, SAT1, SCOC, CAVIN2, SELENOP, SEPTIN7, SERPINB9, SGMS1, SH3BGRL, SKAP2, SLC12A2, SLC2A13, SLMAP, SP3, SPP1, STAT5B, STK17B, STOML1, STRN3, STXBP3, SUCLG2, SUMO1, ELOC, THAP5, THBS1, TM2D1, TMEM123, TMTC3, TNFSF10, TNRC6B, TPMT, TSPAN12, TUBE1, UBB, UBE2D1, USP28, SLC22A25, WNT6, WSB2, XPNPEP3, YIPF4, PTGR2, ZEB2, ZFX, ZNF552, ZNF654
- ***ENSG ID***: ENSG00000276725, ENSG00000281818, ENSG00000165487, ENSG00000132963, ENSG00000284254, ENSG00000100354, ENSG00000101384, ENSG00000173674, ENSG00000285023, ENSG00000102317, ENSG00000196236, ENSG00000072401, ENSG00000131171, ENSG00000091592, ENSG00000155970, ENSG00000078596, ENSG00000111261, ENSG00000239713, ENSG00000137364, ENSG00000005020, ENSG00000137801, ENSG00000151229, ENSG00000127993, ENSG00000170315, ENSG00000075336, ENSG00000169019, ENSG00000206418, ENSG00000113597, ENSG00000177683, ENSG00000213639, ENSG00000119471, ENSG00000138279, ENSG00000155100, ENSG00000115380, ENSG00000090382, ENSG00000152558, ENSG00000250722, ENSG00000067221, ENSG00000198420, ENSG00000163681, ENSG00000147669, ENSG00000176907, ENSG00000226540, ENSG00000123080, ENSG00000162600, ENSG00000182287, ENSG00000123728, ENSG00000120992, ENSG00000112419, ENSG00000180488, ENSG00000048028, ENSG00000136897, ENSG00000106025, ENSG00000110172, ENSG00000005889, ENSG00000138032, ENSG00000164109, ENSG00000172239, ENSG00000169255, ENSG00000139324, ENSG00000189433, ENSG00000162604, ENSG00000176953, ENSG00000121858, ENSG00000117519, ENSG00000172809, ENSG00000166260, ENSG00000147592, ENSG00000188641, ENSG00000170989, ENSG00000064651, ENSG00000074935, ENSG00000145781, ENSG00000119820, ENSG00000115091, ENSG00000160014, ENSG00000240356, ENSG00000001630, ENSG00000170542, ENSG00000178935, ENSG00000029534, ENSG00000081320, ENSG00000136783, ENSG00000130066, ENSG00000154582, ENSG00000107669, ENSG00000109472, ENSG00000077458, ENSG00000171766, ENSG00000104164, ENSG00000103313, ENSG00000175489, ENSG00000172785, ENSG00000172399, ENSG00000156110, ENSG00000119616, ENSG00000112357, ENSG00000117475, ENSG00000163453, ENSG00000232656, ENSG00000139329, ENSG00000174748, ENSG00000187266, ENSG00000198964, ENSG00000197756, ENSG00000161057, ENSG00000187109,

ENSG00000198677, ENSG00000203666, ENSG00000196792, ENSG00000127951, ENSG00000186687, ENSG00000174429, ENSG00000128609, ENSG00000140043, ENSG00000242616, ENSG00000115596, ENSG00000154803, ENSG00000136518, ENSG00000140285, ENSG00000137033, ENSG00000017260, ENSG00000143702, ENSG00000172340, ENSG00000176871, ENSG00000162813, ENSG00000116030, ENSG00000144366, ENSG00000145741, ENSG00000181396, ENSG00000152620, ENSG00000180370, ENSG00000153130, ENSG00000151247, ENSG00000169045, ENSG00000179218, ENSG00000146648, ENSG00000076321, ENSG00000120738, ENSG00000163738, ENSG00000122545, ENSG00000183340, ENSG00000175105, ENSG00000182481, ENSG00000116266, ENSG00000118785, ENSG00000169554, ENSG00000173757, ENSG00000112245, ENSG00000014123, ENSG00000196600, ENSG00000111711, ENSG00000153563, ENSG00000139132, ENSG00000142867, ENSG00000196353, ENSG00000003393, ENSG00000168497, ENSG00000137955, ENSG00000172845, ENSG00000186226, ENSG000002366985

Sorensen 2010

- ***HGNC***: ADM, AK4, ALDOA, ANKRD37, ARRDC3, BNIP3, BNIP3L, FAM210A, FAM162A, CCNG2, CSRP2, EGLN1, EGLN3, ERO1A, FOSL2, GPI, IGFBP3, KDM3A, KCTD11, P4HA1, P4HA2, PDK1, PFKFB3, SLC2A1, SLC6A8, ZNF395
- ***ENSG ID***: ENSG00000282019, ENSG00000170525, ENSG00000288399, ENSG00000129521, ENSG00000177150, ENSG00000197930, ENSG00000186918, ENSG00000104765, ENSG00000075426, ENSG00000176171, ENSG00000130821, ENSG00000105220, ENSG00000113369, ENSG00000162433, ENSG00000114023, ENSG00000175183, ENSG00000138764, ENSG00000149925, ENSG00000186352, ENSG00000148926, ENSG00000122884, ENSG00000135766, ENSG00000146674, ENSG00000072682, ENSG00000115548, ENSG00000152256, ENSG00000117394, ENSG00000213859

Van Malenstein 2010

- ***HGNC***: CCNG2, EGLN3, ERO1A, WDR45B
- ***ENSG ID***: ENSG00000129521, ENSG00000197930, ENSG00000138764, ENSG00000141580

Fardin 2010

- ***HGNC***: AK4, ALDOC, ATP5MC1, ATP5ME, ATP5PF, ATP5MG, ATP6AP1L, ATP6V0B, ATP6V0E1, ATP6V1B1, ATP6V1D, BNIP3, BNIP3L, DDIT4, DDX11, FAM162A, EGLN3, PPP4R3B, NDUFAF1, NDUFB2, NDUFB3, NDUFB6, NDUFB8, NDUFS1, NDUFS6, NDUFS7, NDUFS8, NDUFV3, PDK1, PGK1, PPP4R3B, UQCR10, UQCRC2, VEGFA, WDR5B
- ***ENSG ID***: ENSG00000283447, ENSG00000160194, ENSG00000154723, ENSG00000129521, ENSG00000104765, ENSG00000169020, ENSG00000110717, ENSG00000013573, ENSG00000184076, ENSG00000176171, ENSG00000116039, ENSG00000275052, ENSG00000196981, ENSG00000162433, ENSG00000114023, ENSG00000140740, ENSG00000113732, ENSG00000102144, ENSG00000100554, ENSG00000090266, ENSG00000115286, ENSG00000159199, ENSG00000165264, ENSG00000168209, ENSG00000112715, ENSG00000166136, ENSG00000023228,

ENSG00000145494, ENSG00000137806, ENSG00000167283, ENSG00000109107, ENSG00000205464, ENSG00000152256, ENSG00000117410, ENSG00000119013

Ghazoui 2011

- ***HGNC***: AK4, ANKRD37, ATP1B3, ATP5MC3, BNIP3, CMC2, CCDC167, CHCHD2, EMC8, CTSV, DDIT4, PSMG1, EIF2S2, EIF4EBP1, ENO1, ENY2, GAPDH, GARS1, GMPS, GOLT1B, GPI, HCCS, CXCL8, IMPA2, BPNT2, KIF18A, KIF20A, LDHA, LRRC59, LSM4, MMP1, MRPL13, MRPL15, MRPS17, MTCH2, MTFR1, MTHFD2, NDRG1, NFIL3, ZNF384, NUP93, P4HA1, PDIA6, PFKP, PGAM1, PGK1, PLOD1, PRDX4, PSMA5, PSMB5, PSMD8, RALA, RANBP1, ESRP1, TMEM158, RNASEH2A, SEC61G, SF3B5, SHMT2, SLC16A3, SLC2A1, SLC7A5, SRD5A1, ELOC, TMEM70, UBE2S, VEGFA, YEATS2
- ***ENSG ID***: ENSG00000282019, ENSG00000141401, ENSG00000004961, ENSG00000285121, ENSG00000288299, ENSG00000183527, ENSG00000125977, ENSG00000074800, ENSG00000187840, ENSG00000165030, ENSG00000099901, ENSG00000182199, ENSG00000132432, ENSG00000171314, ENSG00000239789, ENSG00000102900, ENSG00000104331, ENSG00000100804, ENSG00000176171, ENSG00000175606, ENSG00000249992, ENSG00000105220, ENSG00000103257, ENSG00000134333, ENSG00000169429, ENSG00000106153, ENSG00000103121, ENSG00000121621, ENSG00000104413, ENSG00000162433, ENSG00000172172, ENSG00000131148, ENSG00000137547, ENSG00000169976, ENSG00000069849, ENSG00000163655, ENSG00000066855, ENSG00000112984, ENSG00000143106, ENSG00000123131, ENSG00000102144, ENSG00000106105, ENSG00000006451, ENSG00000104419, ENSG00000186352, ENSG00000108106, ENSG00000154582, ENSG00000067057, ENSG00000163872, ENSG00000136943, ENSG00000130520, ENSG00000120533, ENSG00000168209, ENSG00000112715, ENSG00000145545, ENSG00000198937, ENSG00000104889, ENSG00000122884, ENSG00000109919, ENSG00000196611, ENSG00000111640, ENSG00000099341, ENSG00000143870, ENSG00000111711, ENSG00000083444, ENSG00000141526, ENSG00000117394, ENSG00000126746, ENSG00000108829, ENSG00000065911, ENSG00000154518

Toustrup 2011

- ***HGNC***: ADM, ALDOA, ANKRD37, BNIP3, BNIP3L, FAM162A, EGLN3, KCTD11, LOX, NDRG1, P4HA1, P4HA2, PDK1, PFKFB3, SLC2A1
- ***ENSG ID***: ENSG00000170525, ENSG00000288399, ENSG00000129521, ENSG00000104765, ENSG00000176171, ENSG00000114023, ENSG00000149925, ENSG00000104419, ENSG00000186352, ENSG00000148926, ENSG00000122884, ENSG00000113083, ENSG00000072682, ENSG00000152256, ENSG00000117394, ENSG00000213859

Starmans 2012

- ***HGNC***: AARS1, ABCA5, ABCB6, ABHD4, ABI1, ACAD11, ACADVL, ACBD3, ACVR1, ADAM17, ADM, ADORA2B, AFTPH, GPAT3, AGR2, AHR, AKAP12, AKAP8L, ALDOA, ALDOC, ALG2, ANG, ANGPTL4, ANKRD10, ANKRD12, ANKRD37, ANKZF1, ANO6, ANXA5, ARCN1, ARF4, ARFGAP3, ARG2, ARHGAP29, ARHGEF2, ARL1, ARL13B, ARL5B, ARMCX3, ARRDC3, ARRDC4,

ASB3, ASNS, ASS1, ATF2, ATF3, ATF6, ATG13, ATG14, ATG5, ATG9A, ATXN1, AVL9, AVPI1, AZI2, B4GALT4, BBX, BCKDHA, BCKDK, BCL10, BCL6, BET1, BET1L, BHLHE40, BIK, BIRC3, BNIP1, BNIP3L, BRAP, BRD2, BTG1, BTN2A1, INTS13, TMEM263, FAM216A, TIGAR, C12orf57, GPATCH2L, C14orf28, RUSF1, ADPRM, MIR22HG, FAM210A, MYDGF, ODR4, SDE2, RSRP1, SUCO, OSER1, C2orf49, C4orf3, SMIM14, CREBRF, SNHG32, BMT2, HILPDA, FAM220A, LURAP1L, FAM219A, MSANTD3, CA9, CALR, CALU, CAMLG, CANX, CARS1, CAST, CBLB, CFAP36, CCDC107, CCDC28A, CEP95, CCDC47, MIX23, CCDC6, CCNB1IP1, CCNG2, CCNYL1, NOCT, CD55, CDC37L1, CDC6, CDC73, CDK2AP2, CDK7, CDKN1A, CDKN1B, CDR2, CDRT4, CDV3, CEBPG, CEP120, CEP350, CEP57, CFDP1, CHAC1, CHD2, CHIC2, CIR1, CLCN3, CLDN12, CLDND1, CLGN, CLIC4, CLK1, CLK2, CLK3, CLK4, CNOT8, COG6, COPA, COPB1, COPB2, COPG1, COQ10B, CPEB4, CREB3, CREB3L2, CRELD1, CRELD2, CRKL, CRY1, CSNK2A2, CSRNP1, CSRP2, CTH, CTNNAL1, CUL4B, CXCR4, BCLAF3, EOLA2, CYP20A1, CYP26A1, DAGLB, DCUN1D2, DDIT3, DDIT4, DDT, DDX41, DDX50, DDX59, DEDD2, DENND5A, DERL1, DERL2, DHRS3, DHRS7, DHX40, DMTF1, DNAJB11, DNAJB2, DNAJB9, DNAJC1, DNAJC10, DNAJC25, DNAJC3, DNTTIP1, DPCD, DUSP1, DUSP10, DUSP11, DUSP5, DYM, DYRK4, EAF1, EAF2, EDEM1, EDEM2, EDEM3, EDN2, EFNA1, EFL1, EGLN1, EGLN3, EIF1, EIF2AK3, AGO3, ELL2, EMD, ENO2, EPAS1, EPRS1, ERLEC1, ERO1A, ERO1B, ERP44, ERRFI1, ESCO1, ESRP1, ETS2, F3, FADS3, FAM107B, FAM114A1, EEF1AKMT3, FAM13A, FAM162A, FBXO25, FBXO32, FBXO42, FBXO8, FEM1B, FEM1C, FHL2, FICD, FKBP11, FKBP14, FNDC3B, FNIP1, FNTA, FOSL1, FOSL2, FOXD1, FOXO3, FUT11, FYN, GABARAPL1, GADD45A, GADD45B, GALK2, GALNT18, GARS1, GATA6, GBE1, GCH1, GCLC, GDF15, GEM, GFPT1, GLRX, GMPPA, GMPPB, GOLGA4, GOLGA5, GOLGA8A, GOLGB1, GOLPH3L, GOLT1A, GOLT1B, GORAB, GOSR2, GOT1, GPI, GPT2, GRB10, GRB7, GTF2B, GTPBP2, GUK1, GYS1, HBEGF, HBP1, HELZ, HERPUD1, HEY1, HIVEP2, HK2, HLA-A, HLA-B, HLA-C, HLA-E, HMGCL, HMGCR, HOOK1, HPS5, HSP90B1, HSPA13, HSPA5, HYOU1, IBTK, ICA1, IFRD1, IFT20, IGFBP3, IKZF5, IL1RAP, IL20RB, ING1, ING2, INSIG1, INSIG2, IRF6, IRF7, IRS2, ISCA1, ISG20, KDM7A, JMY, JUN, KCTD11, KDELR2, KDELR3, KDM3A, KDM4B, KDM4C, KDM5B, GARRE1, ARFGEF3, KANSL1, LNPK, KIFAP3, KLF10, KLF4, KLF6, KLHL21, KLHL24, KLHL28, KRT15, KTN1, LAMB3, LAMP3, LARP1B, LARP6, LCOR, LDLR, LETMD1, LGALS8, LHFPL2, LIF, LIMCH1, LIN37, LINS1, LNX2, LONP1, LONRF1, LPIN2, LRRC49, LSR, LYSMD3, LZTFL1, MAFF, MAFK, MAGT1, MALAT1, MANBA, MANF, MAP1LC3B, MAP2K1, MBD6, MBNL2, MDM4, MED8, MEF2A, NIM1K, MIA3, MIS12, MKNK2, MKRN1, KMT2E, MNT, MOCOS, MON1B, MORC3, MOSPD1, MOSPD2, MPI, MT1F, MT1X, MT2A, MTFR1, MTHFD2, MTMR11, MXD1, MXI1, MYLIP, MYO1B, NAMPT, NARF, NBEAL1, NBR1, NCK1, NCKIPSD, EPB41L4A-AS1, ZFAS1, NDEL1, NDRG1, NEAT1, NEU1, NFAT5, NFE2L1, NFIL3, NFKBIE, NFKBIZ, NFXL1, NGLY1, NKAP, NOL3, NPC1, NR1D2, NR4A2, NRAS, NRBF2, NUCB2, NUP58, OASL, ORAI3, OS9, OSBP, OSTC, OTUD1, OXSR1, P4HA1, P4HA2, P4HB, PABPC1L, PATL1, PCK2, PCMTD1, PDIA3, PDIA4, PDIA5, PDIA6, PDK1, PDP1, PEA15, PEAR1, PELO, PFDN2, PFKFB3, PFKFB4, PFKP, PGK1, PGM1, PGM3, PHGDH, PIGA, PIM1, PIM3, PJA2, PLAC8, PLAUR, PLIN2, KIZ, PLOD1, PLOD2, PMAIP1,

PNPLA8, PNRC1, POLR3D, PLPP5, PPIB, PPIL4, PPM1D, PPME1, PPP1R15A, PPP1R2, PPP1R3B, PPP2R2A, PPP2R5B, PPP4R2, PPTC7, PREB, PREPL, PRKCZ, PRPF39, PRRC1, PRSS16, PSAP, PSAT1, PTP4A1, PTPN6, PTPRH, QSOX1, RAB11FIP5, RAB20, RAB24, RAB2B, RAB33B, RAB5A, RAB9A, RABAC1, SLC50A1, RASEF, RB1CC1, RBKS, RBM18, RBM7, RBPJ, RETSAT, RHOQ, RINT1, RIOK3, RIT1, RLF, RNASE4, RND3, RNF111, RNF113A, RNF181, RNF183, RNF19A, RNF19B, RNF24, RNF41, RNMT, RP9, RPS6KC1, RRAGC, RRAS, RSBN1, RSRC2, RWDD2A, RYBP, SARS1, SAT1, SAV1, MSMO1, SC5D, SCAPER, SCFD1, SDC4, SDF2L1, SDR16C5, SEC11C, SEC22A, SEC23A, SEC23B, SEC24A, SEC24D, SEC31A, SEC61A1, SEC61G, SEC63, SEL1L, SELENOK, SELENOS, SEMA4B, SERINC1, SERP1, SERTAD1, SERTAD2, SESN2, SREK1, SGPP2, SH3GL3, SH3GLB1, SHCBP1, SHMT2, SIAH1, SIAH2, SIL1, SIRT1, SLC17A5, SLC25A36, SLC25A37, SLC2A1, SLC30A1, SLC31A1, SLC33A1, SLC35B1, SLC35E1, SLC37A3, SLC39A7, SLC3A2, SLC41A2, SLC7A11, SLC9A7, SLCO4A1, SMAP1, SNAPC1, SNAPC3, SNHG1, SNHG12, SNHG8, SNIP1, SNX16, SNX9, SOX9, SPAG1, SPAG4, SPCS3, SPRY1, SPTY2D1, SRP19, SRP54, SRP68, SRP72, SRPRB, SSR1, SSR3, STAT3, STBD1, STK17B, STK19, STT3A, STX5, STYK1, SURF4, SYTL1, SYVN1, TAF1A, TAF1D, TARS1, TAX1BP1, TBC1D15, TBC1D23, TBC1D8B, TBK1, TES, TGIF1, THAP1, THAP5, THAP8, THAP9, THOC6, TIFA, TIPARP, TMC4, TMCO3, TMED2, TMED7, TMED9, TMEM125, TMEM158, LDAF1, TMEM167B, TMEM170A, TMEM182, TMEM214, TMEM38B, TMEM39A, TMEM41B, TMEM45A, TMEM45B, TMEM50B, MACO1, TMEM65, TMF1, TMOD3, TNFAIP3, TNFAIP8, TNFRSF10B, TNIP1, TOM1L1, TPBG, TPD52, TRAM1, TRIB3, TRIM39, TRIT1, TRPT1, TSC22D2, TSC22D3, TSPYL2, TSTD2, TTC33, TTLL7, TUBE1, TUFT1, TULP3, TXNDC15, TXNDC16, TXNIP, U2AF1L4, UAP1, UBE2J1, UBXN4, UBXN6, UCHL3, UFM1, UFSP2, UGGT1, UGP2, UHRF2, ULBP2, ULK1, UNC5B, UPP1, UPRT, USO1, USP37, USP53, VAMP4, VEGFA, VLDLR, VPS37A, WARS1, WASF2, WBP2, WDR41, WDR45, WDR45B, WDR47, WDR54, WFS1, WIPI1, CCN5, WSB1, XBP1, XPNPEP1, YEATS2, YIPF4, YIPF5, YIPF6, YPEL2, YPEL5, ZBTB41, ZBTB8A, ZFAND1, ZFAND2A, ZFP36, ZFYVE1, ZMYM5, ZNF133, ZNF212, ZNF222, ZNF292, ZBTB21, ZNF383, ZNF395, ZNF451, ZNF654, KRBOX4, ZNF697, ZNF805, ZRSR2

- ***ENSG ID***: ENSG00000279389, ENSG00000263160, ENSG00000276561, ENSG00000261893, ENSG00000274873, ENSG00000277667, ENSG00000277996, ENSG00000184494, ENSG00000274905, ENSG00000206450, ENSG00000236250, ENSG00000276260, ENSG00000282800, ENSG00000275003, ENSG00000224994, ENSG00000234343, ENSG00000278363, ENSG00000277789, ENSG00000275867, ENSG00000278191, ENSG00000228691, ENSG00000229300, ENSG00000227715, ENSG00000206495, ENSG00000226437, ENSG00000277273, ENSG00000206493, ENSG00000206505, ENSG00000278458, ENSG00000235307, ENSG00000282019, ENSG00000234704, ENSG00000282735, ENSG00000232839, ENSG00000274384, ENSG00000229802, ENSG00000223980, ENSG00000230678, ENSG00000229215, ENSG00000224320, ENSG00000229929, ENSG00000273722, ENSG00000206435, ENSG00000223532, ENSG00000234728, ENSG00000228964, ENSG00000230308, ENSG00000225201, ENSG00000233841, ENSG00000206288, ENSG00000236632, ENSG00000231834, ENSG00000227402, ENSG00000206380, ENSG00000226033,

ENSG00000230254, ENSG00000236227, ENSG00000280682, ENSG00000139517,
ENSG00000227315, ENSG00000206342, ENSG00000226257, ENSG00000235657,
ENSG00000280951, ENSG00000224608, ENSG00000234507, ENSG00000229252,
ENSG00000283068, ENSG00000232126, ENSG00000226710, ENSG00000283085,
ENSG00000237022, ENSG00000120686, ENSG00000142188, ENSG00000206452,
ENSG00000233904, ENSG00000159256, ENSG00000223957, ENSG00000234947,
ENSG00000234846, ENSG00000226614, ENSG00000283777, ENSG00000283847,
ENSG00000281766, ENSG00000215077, ENSG00000101745, ENSG00000100219,
ENSG00000228299, ENSG00000227129, ENSG00000225691, ENSG00000224399,
ENSG00000139496, ENSG00000185022, ENSG00000285069, ENSG00000128342,
ENSG00000133103, ENSG00000157557, ENSG00000101310, ENSG00000102580,
ENSG00000132950, ENSG00000130150, ENSG00000101782, ENSG00000185950,
ENSG00000285343, ENSG00000141458, ENSG00000173276, ENSG00000285360,
ENSG00000166562, ENSG00000067082, ENSG00000150401, ENSG00000088448,
ENSG00000150403, ENSG00000170525, ENSG00000288445, ENSG00000165997,
ENSG00000198355, ENSG00000178381, ENSG00000155304, ENSG00000065923,
ENSG00000088298, ENSG00000285241, ENSG00000242247, ENSG00000101104,
ENSG00000100290, ENSG00000075643, ENSG00000288399, ENSG00000003147,
ENSG00000094841, ENSG00000072849, ENSG00000110328, ENSG00000118939,
ENSG00000177426, ENSG00000100196, ENSG00000141682, ENSG00000136770,
ENSG00000106993, ENSG00000110756, ENSG00000184205, ENSG00000141627,
ENSG00000087301, ENSG00000139832, ENSG00000153487, ENSG00000139793,
ENSG00000106546, ENSG00000167842, ENSG00000173281, ENSG00000102401,
ENSG00000106541, ENSG00000196365, ENSG00000198517, ENSG00000154359,
ENSG00000147854, ENSG00000112763, ENSG00000147121, ENSG00000100612,
ENSG00000153714, ENSG00000179476, ENSG00000139112, ENSG00000109046,
ENSG00000204599, ENSG00000165861, ENSG00000127663, ENSG00000124783,
ENSG00000206503, ENSG00000173681, ENSG00000129521, ENSG00000067606,
ENSG00000141448, ENSG00000155975, ENSG00000107731, ENSG00000116285,
ENSG00000124145, ENSG00000060140, ENSG00000112812, ENSG00000141446,
ENSG00000108091, ENSG00000074842, ENSG00000177150, ENSG00000101654,
ENSG00000135069, ENSG00000110048, ENSG00000165105, ENSG00000101577,
ENSG00000144655, ENSG00000007944, ENSG00000148572, ENSG00000132823,
ENSG00000010219, ENSG00000106070, ENSG00000184164, ENSG00000147454,
ENSG00000197930, ENSG00000197021, ENSG00000162496, ENSG00000144674,
ENSG00000166889, ENSG00000185246, ENSG00000186918, ENSG00000163512,
ENSG00000119523, ENSG00000165030, ENSG00000168300, ENSG00000101928,
ENSG00000095209, ENSG00000164970, ENSG00000104765, ENSG00000133138,
ENSG00000183696, ENSG00000144566, ENSG00000037637, ENSG00000221914,
ENSG00000064102, ENSG00000164975, ENSG00000128965, ENSG00000078246,
ENSG00000120053, ENSG00000105829, ENSG00000119986, ENSG00000251562,
ENSG00000172939, ENSG00000023318, ENSG00000166171, ENSG00000141738,
ENSG00000196470, ENSG00000071537, ENSG00000120889, ENSG00000136925,
ENSG00000182199, ENSG00000088970, ENSG00000132432, ENSG00000127526,
ENSG00000095596, ENSG00000094804, ENSG00000066455, ENSG00000169504,
ENSG00000140743, ENSG00000151694, ENSG00000140105, ENSG00000166123,

ENSG00000131931, ENSG00000137193, ENSG00000013375, ENSG00000158195, ENSG00000112200, ENSG00000245532, ENSG00000117305, ENSG00000197712, ENSG00000153879, ENSG00000171241, ENSG00000166012, ENSG00000114480, ENSG00000130766, ENSG00000128228, ENSG00000005700, ENSG00000142765, ENSG00000070669, ENSG00000095574, ENSG00000126777, ENSG00000162413, ENSG00000013392, ENSG00000025796, ENSG00000178802, ENSG00000135249, ENSG00000169018, ENSG00000023445, ENSG00000136826, ENSG00000135070, ENSG00000147872, ENSG00000177683, ENSG00000171346, ENSG00000188994, ENSG00000043514, ENSG00000005483, ENSG00000127129, ENSG00000064042, ENSG00000165195, ENSG00000100439, ENSG00000152409, ENSG00000102908, ENSG00000116954, ENSG00000164683, ENSG00000173221, ENSG00000135269, ENSG00000104497, ENSG00000135241, ENSG00000146232, ENSG00000133639, ENSG00000112305, ENSG00000075426, ENSG00000144747, ENSG00000001084, ENSG00000164949, ENSG00000159885, ENSG00000108039, ENSG00000036054, ENSG00000179833, ENSG00000164951, ENSG00000101457, ENSG00000239704, ENSG00000119844, ENSG00000198833, ENSG00000114423, ENSG00000163683, ENSG00000101187, ENSG00000011422, ENSG00000104231, ENSG00000117000, ENSG00000144802, ENSG00000164253, ENSG00000114439, ENSG00000103111, ENSG00000140941, ENSG00000128590, ENSG00000151748, ENSG00000168152, ENSG00000249992, ENSG00000095139, ENSG00000105220, ENSG00000176142, ENSG00000224032, ENSG00000173812, ENSG00000099942, ENSG00000156958, ENSG00000145050, ENSG00000068912, ENSG00000061656, ENSG00000198961, ENSG00000177410, ENSG00000163877, ENSG00000133606, ENSG00000188554, ENSG00000101236, ENSG00000174738, ENSG00000080822, ENSG00000064205, ENSG00000138678, ENSG00000109220, ENSG00000179454, ENSG00000153774, ENSG00000099977, ENSG00000100814, ENSG00000153914, ENSG00000164096, ENSG00000164603, ENSG00000169249, ENSG00000121578, ENSG00000170260, ENSG00000269893, ENSG00000115239, ENSG00000258818, ENSG00000116016, ENSG00000104413, ENSG00000108588, ENSG00000113369, ENSG00000135631, ENSG00000214274, ENSG00000182158, ENSG00000160124, ENSG00000159884, ENSG00000163605, ENSG00000107159, ENSG00000204524, ENSG00000217128, ENSG00000107175, ENSG00000105404, ENSG00000156804, ENSG00000134910, ENSG00000138674, ENSG00000177453, ENSG00000140386, ENSG00000109929, ENSG00000131652, ENSG00000105778, ENSG00000065485, ENSG00000174564, ENSG00000151135, ENSG00000108433, ENSG00000137941, ENSG00000101255, ENSG00000116717, ENSG00000114023, ENSG00000057663, ENSG00000168894, ENSG00000108406, ENSG00000159399, ENSG00000023287, ENSG00000010818, ENSG00000140598, ENSG00000145390, ENSG00000131016, ENSG00000125846, ENSG00000138709, ENSG00000034677, ENSG00000198856, ENSG00000058262, ENSG00000150961, ENSG00000157800, ENSG00000131979, ENSG00000141198, ENSG00000175938, ENSG00000119801, ENSG00000118689, ENSG00000119777, ENSG00000103507, ENSG00000145780, ENSG00000166173, ENSG00000175183, ENSG00000151012, ENSG00000135974, ENSG00000134970, ENSG00000125148, ENSG00000008405, ENSG00000198417, ENSG00000187193, ENSG00000158092, ENSG00000179178, ENSG00000097033, ENSG00000181788, ENSG00000140450, ENSG00000186480, ENSG00000134709, ENSG00000100889, ENSG00000198265,

ENSG00000068305, ENSG00000006459, ENSG00000158290, ENSG00000067167, ENSG00000152684, ENSG00000146278, ENSG00000138640, ENSG00000082641, ENSG00000075420, ENSG00000112379, ENSG00000145685, ENSG00000119899, ENSG00000137962, ENSG00000119326, ENSG00000131013, ENSG00000068971, ENSG00000140600, ENSG00000118503, ENSG00000104812, ENSG00000079739, ENSG00000125352, ENSG00000101882, ENSG00000170836, ENSG00000173575, ENSG00000111897, ENSG00000080031, ENSG00000145817, ENSG00000113638, ENSG00000164463, ENSG00000136986, ENSG00000117525, ENSG00000105856, ENSG00000066855, ENSG00000114120, ENSG00000138764, ENSG00000074935, ENSG00000120742, ENSG00000181704, ENSG00000129472, ENSG00000120129, ENSG00000145779, ENSG00000066697, ENSG00000119820, ENSG00000164117, ENSG00000123595, ENSG00000170417, ENSG00000196998, ENSG00000042445, ENSG00000010810, ENSG00000138073, ENSG00000107077, ENSG00000138594, ENSG00000128016, ENSG00000152952, ENSG00000174780, ENSG00000109572, ENSG00000124762, ENSG00000173230, ENSG00000167004, ENSG00000120805, ENSG00000146242, ENSG00000090861, ENSG00000118985, ENSG00000031698, ENSG00000147535, ENSG00000166822, ENSG00000168556, ENSG00000115641, ENSG00000006652, ENSG00000119446, ENSG00000078081, ENSG00000089916, ENSG00000102144, ENSG00000169359, ENSG00000138166, ENSG00000104450, ENSG00000130164, ENSG00000102119, ENSG00000163659, ENSG00000081181, ENSG00000197019, ENSG00000196233, ENSG00000100883, ENSG00000114796, ENSG00000081320, ENSG00000106105, ENSG00000169032, ENSG00000140688, ENSG00000129128, ENSG00000196083, ENSG00000089234, ENSG00000092108, ENSG00000149925, ENSG00000125629, ENSG00000102158, ENSG00000109775, ENSG00000130066, ENSG00000157514, ENSG00000120725, ENSG00000087074, ENSG00000052802, ENSG00000196968, ENSG00000104419, ENSG00000121966, ENSG00000186352, ENSG00000148926, ENSG00000168522, ENSG00000067057, ENSG00000113742, ENSG00000170448, ENSG00000166987, ENSG00000135164, ENSG00000148248, ENSG00000135506, ENSG00000166037, ENSG00000113916, ENSG00000163872, ENSG00000175592, ENSG00000123427, ENSG00000184203, ENSG00000044574, ENSG00000184840, ENSG00000169228, ENSG00000136754, ENSG00000144320, ENSG00000024862, ENSG00000248098, ENSG00000221968, ENSG00000168374, ENSG00000137947, ENSG00000147364, ENSG00000014914, ENSG00000145088, ENSG00000121542, ENSG00000130513, ENSG00000106080, ENSG00000119950, ENSG00000151014, ENSG00000172007, ENSG00000100934, ENSG00000135245, ENSG00000154265, ENSG00000183258, ENSG00000164056, ENSG00000166398, ENSG00000116260, ENSG00000126775, ENSG00000157224, ENSG00000204256, ENSG00000168209, ENSG00000110619, ENSG00000164111, ENSG00000051108, ENSG00000161265, ENSG00000075945, ENSG00000267796, ENSG00000147852, ENSG00000107625, ENSG00000122218, ENSG00000172432, ENSG00000112715, ENSG00000197746, ENSG00000167772, ENSG00000140939, ENSG00000134285, ENSG00000162298, ENSG00000115966, ENSG00000157450, ENSG00000137821, ENSG00000118197, ENSG00000070540, ENSG00000163082, ENSG00000204387, ENSG00000126070, ENSG00000117533, ENSG00000111674, ENSG00000111678, ENSG00000111679, ENSG00000181458, ENSG00000126458, ENSG00000204386, ENSG00000151092, ENSG00000204592, ENSG00000143507,

ENSG00000143367, ENSG00000167671, ENSG00000129083, ENSG00000204525, ENSG00000265972, ENSG00000168495, ENSG00000078237, ENSG00000165312, ENSG00000135837, ENSG00000065809, ENSG00000117595, ENSG00000099860, ENSG00000070770, ENSG00000186594, ENSG00000125398, ENSG00000131871, ENSG00000185507, ENSG00000121073, ENSG00000170786, ENSG00000117139, ENSG00000059769, ENSG00000164983, ENSG00000130707, ENSG00000130340, ENSG00000255717, ENSG00000178397, ENSG00000164535, ENSG00000136240, ENSG00000113070, ENSG00000153037, ENSG00000086598, ENSG00000176018, ENSG00000128595, ENSG00000214517, ENSG00000168944, ENSG00000155090, ENSG00000134107, ENSG00000134109, ENSG00000136731, ENSG00000162236, ENSG00000198855, ENSG00000149743, ENSG00000140471, ENSG00000136868, ENSG00000165188, ENSG00000138768, ENSG00000114850, ENSG00000172183, ENSG00000166794, ENSG00000157181, ENSG00000113734, ENSG00000096717, ENSG00000091527, ENSG00000122884, ENSG00000135766, ENSG00000145901, ENSG00000076554, ENSG00000094975, ENSG00000162772, ENSG00000023608, ENSG00000105835, ENSG00000170222, ENSG00000113407, ENSG00000153113, ENSG00000160570, ENSG00000111011, ENSG00000173540, ENSG00000117143, ENSG00000196878, ENSG00000171174, ENSG00000145365, ENSG00000234745, ENSG00000251493, ENSG00000113615, ENSG00000164615, ENSG00000113621, ENSG00000076053, ENSG00000145287, ENSG00000185033, ENSG00000136643, ENSG00000086619, ENSG00000141562, ENSG00000168214, ENSG00000128641, ENSG00000146674, ENSG00000164610, ENSG00000153132, ENSG00000136628, ENSG00000184432, ENSG00000155508, ENSG00000179335, ENSG00000153234, ENSG00000072682, ENSG00000240303, ENSG00000119729, ENSG00000175265, ENSG00000113161, ENSG00000116406, ENSG00000132471, ENSG00000113240, ENSG00000109323, ENSG00000155660, ENSG00000167881, ENSG00000119004, ENSG00000163602, ENSG00000177951, ENSG00000070444, ENSG00000179218, ENSG00000127022, ENSG00000005448, ENSG00000166579, ENSG00000118804, ENSG00000011243, ENSG00000115548, ENSG00000106052, ENSG00000164244, ENSG00000134058, ENSG00000109083, ENSG00000161277, ENSG00000134371, ENSG00000116761, ENSG00000109107, ENSG00000151715, ENSG00000167608, ENSG00000177888, ENSG00000169242, ENSG00000169241, ENSG00000138078, ENSG00000099875, ENSG00000105699, ENSG00000175105, ENSG00000184014, ENSG00000112473, ENSG00000166471, ENSG00000152256, ENSG00000143870, ENSG00000144867, ENSG00000176444, ENSG00000183735, ENSG00000131015, ENSG00000258890, ENSG00000050426, ENSG00000177169, ENSG00000215717, ENSG00000168003, ENSG00000143622, ENSG00000124788, ENSG00000116584, ENSG00000109501, ENSG00000166598, ENSG00000196352, ENSG00000135114, ENSG00000115657, ENSG00000177119, ENSG00000169379, ENSG00000198925, ENSG00000163516, ENSG00000204856, ENSG00000196850, ENSG00000149428, ENSG00000162734, ENSG00000144597, ENSG00000168610, ENSG00000188283, ENSG00000090520, ENSG00000204344, ENSG00000181852, ENSG00000114268, ENSG00000112245, ENSG00000167797, ENSG00000213672, ENSG00000111711, ENSG00000144048, ENSG00000070081, ENSG00000083444, ENSG00000121749, ENSG00000163249, ENSG00000163818, ENSG00000185624, ENSG00000163703, ENSG00000143498, ENSG00000154305, ENSG00000163001, ENSG00000169764,

ENSG00000113811, ENSG00000120370, ENSG00000143067, ENSG00000092621, ENSG00000120071, ENSG00000179119, ENSG00000142867, ENSG00000117394, ENSG00000111276, ENSG00000115520, ENSG00000175197, ENSG00000213859, ENSG00000136052, ENSG00000141580, ENSG00000196428, ENSG00000138433, ENSG00000175224, ENSG00000181789, ENSG00000116977, ENSG00000065911, ENSG00000135913, ENSG00000143751, ENSG00000182827, ENSG00000072778, ENSG00000143457, ENSG00000177606, ENSG00000170425, ENSG00000143256, ENSG00000175155, ENSG00000077232, ENSG00000143774, ENSG00000085433, ENSG00000116514, ENSG00000198380, ENSG00000170385, ENSG00000118217, ENSG00000115963, ENSG00000159479, ENSG00000013441, ENSG00000059728, ENSG00000115170, ENSG00000135924, ENSG00000144224, ENSG00000117616, ENSG00000204178, ENSG00000172071, ENSG00000144591, ENSG00000081019, ENSG00000144426, ENSG00000197989, ENSG00000174567, ENSG00000198625, ENSG00000160062, ENSG00000213281, ENSG00000187800

Halle 2012

- ***HGNC***: ALDOA, AK2, AK4, B3GNT4, SCARB1, CLK3, MRGBP, ECE2, ERO1A, GAPDH, HMOX1, ISG15, PFKFB4, P4HA2, PYGL, RPL36A, UPK1A, DDIT3, KCTD11, PVR, RHOC, STC2, ATP5MJ, C19orf53, C4orf3, FGF11, SH3GL3, SNTA1, SPAG7, S100A2, TRAPPC1
- ***ENSG ID***: ENSG00000283903, ENSG00000288399, ENSG00000091640, ENSG00000101400, ENSG00000197930, ENSG00000100292, ENSG00000073008, ENSG00000101189, ENSG00000241343, ENSG00000100504, ENSG00000105668, ENSG00000156411, ENSG00000164096, ENSG00000162433, ENSG00000140600, ENSG00000149925, ENSG00000176383, ENSG00000161958, ENSG00000170043, ENSG00000111640, ENSG00000179335, ENSG00000072682, ENSG00000145194, ENSG00000104979, ENSG00000113739, ENSG00000073060, ENSG00000114268, ENSG00000155366, ENSG00000175197, ENSG00000213859, ENSG00000004455, ENSG00000187608, ENSG00000196754

Eustace 2013

- ***HGNC***: ALDOA, ANGPTL4, ANLN, BNC1, MRGBP, CA9, CDKN3, COL4A6, DCBLD1, ENO1, FAM83B, FOSL1, GNAI1, HILPDA, KCTD11, KRT17, LDHA, P4HA1, PGAM1, PGK1, SDC1, SLC16A1, SLC2A1, TPI1, VEGFA
- ***ENSG ID***: ENSG00000276701, ENSG00000100290, ENSG00000186895, ENSG00000106366, ENSG00000127129, ENSG00000240972, ENSG00000116717, ENSG00000123384, ENSG00000112715, ENSG00000146674, ENSG00000137745

Boidot 2014 (Continuous Hypoxia)

- ***HGNC***: TMEM258, KDM3A, RPS28, SEC61G, ANKZF1, ALDOA, P4HA1, C4orf3, H4C1, DDT, INSIG2, RPLP2, PFKFB4, BNIP3L, P4HA2, ISCA1, KPNB1, DDIT4, RBX1, H1-2, PYCR3, ANP32D, ANKRD37, PNRC1, GPATCH4, PPFIA4, ANP32C, NDUFAF2, RPS2, CLK3, RPS13, ANP32A, VEGFA, ACAP1, PHPT1, PGK1, PRMT3, MXI1, MAP1LC3B, KCTD11, RRP15, PTPRCAP, H4C2, H4C3, CMSS1, RNASEH1, WSB1, AIDA, H2AC4, SSNA1, UTP20, TIMM23, NDRG1, PFKFB3, FAM162A, MRPL19, MRPL3, RPPH1, PSME3, HCFC1R1, H2AC6, PSMD7,

EBNA1BP2, HINT1, PGAM1, PSMC3, DDIT3, FUT11, TNFSF8, CEP83, KLHDC10, CCNG2, RPL39, HTR5A, PNPT1, UBTF, KRTAP10-12, PPP2R5B, SNORA68, GPI, TAGAP, TMEM88, AK4, MAD1L1, H3C2, DIPK2A, CHD1, NDUFA1, PDK1, WBP1, POU5F1P3, MAP1LC3B2, ATF4

- ***ENSG ID***: ENSG00000275003, ENSG00000276657, ENSG00000282019, ENSG00000170525, ENSG00000288169, ENSG00000288399, ENSG00000140988, ENSG00000100387, ENSG00000189169, ENSG00000110700, ENSG00000109046, ENSG00000235602, ENSG00000185238, ENSG00000104765, ENSG00000132432, ENSG00000171314, ENSG00000171865, ENSG00000135070, ENSG00000140350, ENSG00000184220, ENSG00000131467, ENSG00000140941, ENSG00000105220, ENSG00000128272, ENSG00000128607, ENSG00000138035, ENSG00000099977, ENSG00000277209, ENSG00000164096, ENSG00000103035, ENSG00000162433, ENSG00000258102, ENSG00000115364, ENSG00000114023, ENSG00000114686, ENSG00000176101, ENSG00000146278, ENSG00000068971, ENSG00000198918, ENSG00000125356, ENSG00000138764, ENSG00000054148, ENSG00000177600, ENSG00000103145, ENSG00000207166, ENSG00000120800, ENSG00000173588, ENSG00000102144, ENSG00000149925, ENSG00000125629, ENSG00000196968, ENSG00000104419, ENSG00000265354, ENSG00000186352, ENSG00000248546, ENSG00000157219, ENSG00000134825, ENSG00000278637, ENSG00000286522, ENSG00000278463, ENSG00000187837, ENSG00000119950, ENSG00000168209, ENSG00000112715, ENSG00000233927, ENSG00000181744, ENSG00000106952, ENSG00000169567, ENSG00000104524, ENSG00000213402, ENSG00000067533, ENSG00000143847, ENSG00000167874, ENSG00000122884, ENSG00000153922, ENSG00000002822, ENSG00000164182, ENSG00000072818, ENSG00000164691, ENSG00000179335, ENSG00000072682, ENSG00000180573, ENSG00000239779, ENSG00000115548, ENSG00000197061, ENSG00000152256, ENSG00000108424, ENSG00000278705, ENSG00000165916, ENSG00000163516, ENSG00000114268, ENSG00000108312, ENSG00000186063, ENSG00000139223, ENSG00000175197, ENSG00000213859, ENSG00000117395, ENSG00000160818

Boidot 2014 (Cyclic Hypoxia)

- ***HGNC***:BIRC5, ZGPAT, LSM5, PFDN2, FCN1, NACA, PTPRCAP, TMED1, IGBP1, EIF4B, H1-2, SLIRP, HTR1B, GSTA3, SAC3D1, NTHL1, MPPED1, NEUROD2, KCNJ12, LHX5, TMEM160, HMX1, TMEM258, MARCO, TSG101, UTF1, LLPH, H2AC6, CBLN1, NDUFA13, TLX2, AURKAIP1, ANTKMT, ZSWIM1, TPSD1, UBE2D2, EDF1, MRPL17, CEND1, SSNA1, RPS28, SRPK3, NELFB, SDHC, MED31, FAM30A, SNF8, POU3F4, GORASP1, C2orf42, RBX1, MKNK1, SNRPD3, ABHD17A, H3-4, RPL19, RPS13, ZFTA, MRPS7, GNG5, PRCC, CHMP1B, LMO2, TCF15, COPZ1, MAD1L1, H4C1, H4C3, RPPH1, MALAT1, NKX6-3, POM121L8P, CIAO2A, TMIGD2, MRPL36, TMEM61, KTI12, RBFOX3, RAX2, LRRC45, LGALS7, SPDYE7P, LCN12, TRIM65, MFSD14A, C2CD4B, TIMMDC1, COMTD1, GSX1, CELF4
- ***ENSG ID***: ENSG00000277059, ENSG00000255112, ENSG00000101489, ENSG00000229266, ENSG00000283082, ENSG00000285435, ENSG00000288204, ENSG00000169840, ENSG00000186732, ENSG00000100387, ENSG00000100028, ENSG00000196767, ENSG00000110700, ENSG00000065057, ENSG00000108590,

ENSG00000184185, ENSG00000089289, ENSG00000167664, ENSG00000171421, ENSG00000175756, ENSG00000165644, ENSG00000186010, ENSG00000251562, ENSG00000168061, ENSG00000215612, ENSG00000171532, ENSG00000106355, ENSG00000085265, ENSG00000184343, ENSG00000168612, ENSG00000119705, ENSG00000143001, ENSG00000113845, ENSG00000165066, ENSG00000277209, ENSG00000198841, ENSG00000197114, ENSG00000205502, ENSG00000174021, ENSG00000130748, ENSG00000125878, ENSG00000135363, ENSG00000131508, ENSG00000159210, ENSG00000184925, ENSG00000176101, ENSG00000188986, ENSG00000184524, ENSG00000107223, ENSG00000174156, ENSG00000135312, ENSG00000173976, ENSG00000103254, ENSG00000134825, ENSG00000278637, ENSG00000187837, ENSG00000166797, ENSG00000169683, ENSG00000233927, ENSG00000095917, ENSG00000171794, ENSG00000129968, ENSG00000102924, ENSG00000213402, ENSG00000063046, ENSG00000226777, ENSG00000099203, ENSG00000002822, ENSG00000196531, ENSG00000111481, ENSG00000141569, ENSG00000180573, ENSG00000115297, ENSG00000158042, ENSG00000197061, ENSG00000179994, ENSG00000125445, ENSG00000108298, ENSG00000205076, ENSG00000089116, ENSG00000089685, ENSG00000114745, ENSG00000139233, ENSG00000074319, ENSG00000079277, ENSG00000168148, ENSG00000143256, ENSG00000143252, ENSG00000167281, ENSG00000019169, ENSG00000143294, ENSG00000156875, ENSG00000115998

Ragnum 2015

- ***HGNC***: ASF1B, ASPM, BIRC5, BUB3, CENPE, CENPU, CMTM3, DONSON, DTL, FOXM1, G6PD, HJURP, MCM2, MEP1A, MTMR2, TRIP13, ZWINT, TDG, UNG, XRCC6, ADM, DDIT4, DSP, FER1L4, HILPDA, P4HA1, PGAM4, PKM, RIMKLA, RNASE4, SCD, SPAG4
- ***ENSG ID***: ENSG00000159147, ENSG00000288210, ENSG00000122952, ENSG00000088340, ENSG00000196419, ENSG00000071539, ENSG00000099194, ENSG00000160211, ENSG00000177181, ENSG00000154473, ENSG00000061656, ENSG00000138778, ENSG00000258818, ENSG00000073111, ENSG00000226784, ENSG00000148926, ENSG00000111206, ENSG00000087053, ENSG00000135245, ENSG00000168209, ENSG00000140931, ENSG00000112818, ENSG00000143476, ENSG00000151725, ENSG00000105011, ENSG00000122884, ENSG00000067225, ENSG00000076248, ENSG00000123485, ENSG00000096696, ENSG00000139372, ENSG00000089685, ENSG00000066279

Fjeldbo 2016

- ***HGNC***: DDIT3, ERO1A, KCTD11, P4HA2, STC2, UPK1A
- ***ENSG ID***: ENSG00000288399, ENSG00000197930, ENSG00000105668, ENSG00000072682, ENSG00000113739, ENSG00000175197, ENSG00000213859

Suh 2017

- ***HGNC***: TMPRSS11D, GJB6, IL20RB, LOX, APOL1, FBXO45, S100A7, NCOA7, TM4SF1, CDCP1, NAMPT, TFRC, SOD2, PFKFB3, SRPK1, RUNX3, SYNGR2, ORAI2, POU2AF1, LIFR, TPM4

- ***ENSG ID***: ENSG00000170525, ENSG00000121742, ENSG00000100342, ENSG00000167460, ENSG00000096063, ENSG00000110777, ENSG00000111912, ENSG00000153802, ENSG00000163814, ENSG00000174564, ENSG00000169908, ENSG00000113594, ENSG00000112096, ENSG00000143556, ENSG00000174013, ENSG00000105835, ENSG00000113083, ENSG00000160991, ENSG00000072274, ENSG00000108639, ENSG00000020633

Yang 2017

- ***HGNC***: CAV1, COL5A1, ITGA5, P4HA2, SLC16A1, TGFBI, DPYSL2, SRPX, TRAM2, SYDE1, LRP1, PDLIM2, SAV1, AHNAK2, CAD, CYP1B1, DAAM1, DSC2, SLC2A3, FUT11, GLG1, GULP1, LDLR, THBS4
- ***ENSG ID***: ENSG00000281917, ENSG00000101955, ENSG00000059804, ENSG00000120913, ENSG00000092964, ENSG00000105137, ENSG00000065308, ENSG00000185567, ENSG00000134755, ENSG00000138061, ENSG00000161638, ENSG00000130635, ENSG00000084774, ENSG00000151748, ENSG00000120708, ENSG00000113296, ENSG00000090863, ENSG00000123384, ENSG00000100592, ENSG00000130164, ENSG00000196968, ENSG00000105974, ENSG00000144366, ENSG00000072682, ENSG00000155380

Ye 2018

- ***HGNC***: AK4, ALDOA, ALDOC, ANGPTL4, ANKRD37, BHLHE40, BNIP3, BNIP3L, C4orf3, C4orf47, CA9, CASP14, DARS1-AS1, DDIT4, DNAH11, EGLN1, EGLN3, FGF11, FUT11, GBE1, HK2, KDM3A, LDHA, LOX, MIR210HG, NDRG1, P4HA1, PDK1, PFKFB4, PGK1, PPFIA4, SDAD1P1, SLC2A1, SPAG4, STC1, TCAF2, TMEM45A, VEGFA, VLDLR
- ***ENSG ID***: ENSG00000282810, ENSG00000283903, ENSG00000288299, ENSG00000105877, ENSG00000129521, ENSG00000159167, ENSG00000228451, ENSG00000104765, ENSG00000105141, ENSG00000114480, ENSG00000176171, ENSG00000061656, ENSG00000134333, ENSG00000164096, ENSG00000162433, ENSG00000107159, ENSG00000159399, ENSG00000102144, ENSG00000149925, ENSG00000196968, ENSG00000104419, ENSG00000186352, ENSG00000168209, ENSG00000147852, ENSG00000112715, ENSG00000167772, ENSG00000181458, ENSG00000231890, ENSG00000247095, ENSG00000134107, ENSG00000143847, ENSG00000161958, ENSG00000122884, ENSG00000135766, ENSG00000113083, ENSG00000115548, ENSG00000170379, ENSG00000109107, ENSG00000152256, ENSG00000205129, ENSG00000114268, ENSG00000117394

Yang 2018 (Prostate)

- ***HGNC***: ADAMTS4, ATF3, BHLHE40, BTG2, CSRNP1, CCN1, EGR1, EGR2, EGR3, FOSB, FOSL2, GEM, JUNB, KLF10, KLF6, LIF, MCL1, NR4A3, PPP1R15A, RHOB, SELE, SIK1, SLC2A14, SLC2A3, SOCS3, THBS1, TIPARP, ZFP36
- ***ENSG ID***: ENSG00000128342, ENSG00000067082, ENSG00000142178, ENSG00000059804, ENSG00000179388, ENSG00000144655, ENSG00000137801, ENSG00000119508, ENSG00000075426, ENSG00000164949, ENSG00000125740, ENSG00000128016, ENSG00000163659, ENSG00000087074, ENSG00000007908, ENSG00000159388, ENSG00000155090, ENSG00000122877, ENSG00000134107,

ENSG00000171223, ENSG00000162772, ENSG00000120738, ENSG00000173262, ENSG00000184557, ENSG00000158859, ENSG00000142871, ENSG00000143878, ENSG00000143384

Yang 2018 (Sarcoma)

- ***HGNC***: ENO2, SLC2A1, BNIP3, PDK1, NDRG1, PFKFB4, FAM162A, VEGFA, ZNF395, DDIT4, ANKRD37, MXI1, SLC2A3, PPFIA4, GBE1, ALDOC, CDK18, ANG, PRSS53, INSIG2, VLDLR, P4HA1, BNIP3L, BHLHE40
- ***ENSG ID***: ENSG00000059804, ENSG00000186918, ENSG00000104765, ENSG00000114480, ENSG00000176171, ENSG00000214274, ENSG00000114023, ENSG00000151006, ENSG00000125629, ENSG00000104419, ENSG00000186352, ENSG00000119950, ENSG00000168209, ENSG00000147852, ENSG00000112715, ENSG00000111674, ENSG00000134107, ENSG00000143847, ENSG00000122884, ENSG00000109107, ENSG00000152256, ENSG00000114268, ENSG00000117394, ENSG00000117266

Trong 2018

- ***HGNC***: LYVE1, FAM162A, WNT6, OTP, PLOD1
- ***ENSG ID***: ENSG00000114023, ENSG00000115596, ENSG00000133800, ENSG00000171540, ENSG00000083444

Chen 2019

- ***HGNC***: ALDOA, ALDOC, ANGPTL4, DDIT4, DRD4, ENO2, GPI, GREB1L, LOXL2, PDK1, PFKP, PGK1, PLOD2, SLC2A1, SPOCK1, STOX1, SUV39H2
- ***ENSG ID***: ENSG00000276825, ENSG00000282019, ENSG00000152455, ENSG00000141449, ENSG00000134013, ENSG00000105220, ENSG00000152952, ENSG00000102144, ENSG00000149925, ENSG00000067057, ENSG00000152377, ENSG00000168209, ENSG00000165730, ENSG00000167772, ENSG00000111674, ENSG00000069696, ENSG00000109107, ENSG00000152256, ENSG00000117394

Zou 2019

- ***HGNC***: MDM2, VEGFA, ORAI3, MVD, TRAF3, CYB5R3, ZBTB44, CASP6, FBP1, CCNG1, FAM117B, PRELID2, RRP1B, GAS6
- ***ENSG ID***: ENSG00000160208, ENSG00000183087, ENSG00000100243, ENSG00000131323, ENSG00000167508, ENSG00000196323, ENSG00000175938, ENSG00000113328, ENSG00000138439, ENSG00000112715, ENSG00000165140, ENSG00000138794, ENSG00000186314, ENSG00000135679

Zhang 2020

- ***HGNC***: PDSS1, CDCA8, SLC7A11
- ***ENSG ID***: ENSG00000148459, ENSG00000134690, ENSG00000151012

Wang 2020

- ***HGNC***: PFKL, P4HA2, GRHPR, SDC3, PPP1R15A, SIAH2, NDRG1, BTG1, TPD52, MAFF, ISG20, LALBA, ERRFI1, VHL

- ***ENSG ID***: ENSG00000185022, ENSG00000141959, ENSG00000137106, ENSG00000116285, ENSG00000162512, ENSG00000133639, ENSG00000181788, ENSG00000087074, ENSG00000104419, ENSG00000134086, ENSG00000172183, ENSG00000076554, ENSG00000072682, ENSG00000167531

Shou 2020

- ***HGNC***: ABCA12, PTK6, FERMT1, GSDMC, KRT2, CSTA, SPRR2F
- ***ENSG ID***: ENSG00000285114, ENSG00000101213, ENSG00000101311, ENSG00000172867, ENSG00000121552, ENSG00000244094, ENSG00000147697, ENSG00000144452

Lin 2000

- ***HGNC***: VEGFA, HK2, JUN, LDHA, GAPDH
- ***ENSG ID***: ENSG00000288299, ENSG00000134333, ENSG00000159399, ENSG00000112715, ENSG00000111640, ENSG00000177606

Mo 2020

- ***HGNC***: XPNPEP1, ANGPTL4, SLC2A1, PFKP
- ***ENSG ID***: ENSG00000108039, ENSG00000067057, ENSG00000167772, ENSG00000117394

Tardon 2020

- ***HGNC***: ADM, ALDOC, ANKRD37, ARRDC3, BHLHE40, CA9, DDIT4, EGLN3, HAS2, HILPDA, HK2, PGK1, NDRG1, PDK1, SLC2A1, SLC2A3, STC1, TMEM45A, VEGFA
- ***ENSG ID***: ENSG00000129521, ENSG00000159167, ENSG00000059804, ENSG00000113369, ENSG00000159399, ENSG00000170961, ENSG00000102144, ENSG00000104419, ENSG00000148926, ENSG00000168209, ENSG00000135245, ENSG00000152256, ENSG00000181458, ENSG00000107159, ENSG00000134107, ENSG00000112715, ENSG00000109107, ENSG00000186352, ENSG00000117394

Appendix 2

	Median	Mean	GSVA	ssGSEA
Aprelikova	100.00	16.00	96.00	0.00
Benita	96.00	68.00	84.00	0.00
Beyer	100.00	20.00	96.00	0.00
Boidot Cont. *†	100.00	20.00	16.00	0.00
Boidot Cyc. *†	88.00	0.00	12.00	0.00
Bosco	100.00	56.00	92.00	0.00
Buffa †	100.00	84.00	24.00	0.00
Chen	100.00	64.00	24.00	0.00
Chi	100.00	32.00	92.00	0.00
Denko	100.00	56.00	96.00	4.00
Detwiller	100.00	28.00	36.00	4.00
Elvidge	100.00	24.00	96.00	0.00
Eustace †	100.00	40.00	12.00	4.00
Fardin 2009	100.00	68.00	0.00	4.00
Fardin 2010	96.00	24.00	4.00	0.00
Fjeldbo †	88.00	8.00	0.00	0.00
Ghazoui †	100.00	84.00	16.00	0.00
Ghorbel †	92.00	0.00	12.00	0.00
Halle †	100.00	80.00	36.00	0.00
Hu †	100.00	68.00	0.00	0.00
Jogi	100.00	28.00	80.00	0.00
Koong	100.00	72.00	0.00	0.00
Lendhal	100.00	40.00	56.00	0.00
Lin †	100.00	92.00	0.00	0.00
Manalo	100.00	16.00	72.00	0.00
Mense	100.00	16.00	96.00	8.00
Mo †	76.00	36.00	0.00	4.00
Ning	92.00	44.00	16.00	0.00
Peters	100.00	28.00	36.00	0.00
Ragnum †	100.00	28.00	0.00	0.00
Seigneuric C.	16.00	8.00	0.00	0.00
Seigneuric E0	72.00	0.00	4.00	0.00
Seigneuric E2	84.00	8.00	0.00	0.00
Shi	76.00	0.00	4.00	4.00
Shou *†	8.00	0.00	0.00	0.00
Suh *†	76.00	0.00	0.00	0.00
Sorensen	100.00	24.00	84.00	0.00
Starmans	100.00	24.00	84.00	4.00
Sun *†	8.00	4.00	0.00	0.00
Sung	100.00	32.00	96.00	0.00
Tardon	100.00	76.00	76.00	0.00
Toustrup	100.00	28.00	52.00	0.00
Trong *	0.00	0.00	0.00	0.00
Van Malenstein	20.00	8.00	0.00	4.00
Wang 2005	100.00	36.00	76.00	0.00
Wang 2020 *†	96.00	16.00	0.00	0.00
Winter †	96.00	68.00	28.00	0.00
Yang 2017 †	96.00	0.00	16.00	0.00
Yang Prostate *†	96.00	0.00	4.00	0.00
Yang Sarcoma †	100.00	56.00	92.00	0.00
Ye	100.00	64.00	96.00	0.00
Zhang †	0.00	0.00	0.00	0.00
Zou *†	72.00	0.00	0.00	0.00

Supplementary Table S1: *Percentage accuracy of determining hypoxic samples from normoxic samples in breast non-cancer cell lines for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in Table 4.2

	Median	Mean	GSVA	ssGSEA
Aprelikova	83.24	72.00	88.76	0.95
Benita	85.24	79.90	81.33	1.05
Beyer	81.33	66.95	83.33	1.33
Boidot Cont. *†	70.10	62.57	45.62	1.05
Boidot Cyc. *†	12.57	5.05	3.90	0.29
Bosco	53.52	74.57	74.86	0.86
Buffa †	66.00	76.19	40.38	0.86
Chen	74.67	71.90	45.52	0.38
Chi	79.71	69.81	81.05	1.14
Denko	74.86	68.86	75.43	0.86
Detwiller	64.10	62.67	38.10	0.76
Elvidge	85.24	68.19	85.33	0.95
Eustace †	69.90	72.10	50.19	0.76
Fardin 2009	69.71	70.29	33.14	0.95
Fardin 2010	46.67	45.90	8.10	0.76
Fjeldbo †	53.05	43.52	20.95	0.48
Ghazoui †	48.48	65.71	21.05	0.76
Ghorbel †	15.52	6.86	11.05	0.10
Halle †	64.67	67.05	44.95	0.67
Hu †	63.71	62.19	20.48	0.57
Jogi	62.57	66.76	63.71	0.38
Koong	40.10	53.81	23.05	0.76
Lendhal	84.29	70.48	61.14	0.57
Lin †	72.76	78.95	28.86	0.86
Manalo	49.62	57.71	70.76	0.95
Mense	86.95	69.81	84.76	1.05
Mo †	52.67	41.24	13.62	0.67
Ning	37.62	37.81	11.05	0.29
Peters	27.90	26.19	10.38	0.19
Ragnum †	35.43	35.33	8.00	0.86
Seigneuric C.	4.95	1.62	0.19	0.19
Seigneuric E0	12.86	1.24	0.86	0.10
Seigneuric E2	13.24	11.90	2.29	0.00
Shi	8.10	11.14	4.10	0.38
Shou *†	2.03	2.22	1.16	0.48
Suh *†	19.43	21.90	7.71	0.57
Sorensen	87.52	69.90	77.81	0.76
Starmans	76.67	64.10	81.81	0.76
Sun *†	1.14	1.33	0.86	0.19
Sung	86.95	71.33	86.10	0.48
Tardon	86.10	77.14	70.48	1.33
Toustrup	83.71	73.62	64.95	1.24
Trong *	3.33	16.95	4.67	0.76
Van Malenstein	39.81	36.38	13.90	0.76
Wang 2005	78.95	69.43	69.62	0.67
Wang 2020 *†	47.14	48.29	19.24	0.86
Winter †	41.62	66.19	23.52	0.76
Yang 2017 †	44.67	40.29	38.19	0.76
Yang Prostate *†	37.52	40.48	33.90	0.67
Yang Sarcoma †	85.24	72.48	76.86	1.24
Ye	85.43	78.86	82.86	0.95
Zhang †	0.19	1.14	0.57	0.00
Zou *†	12.95	17.43	4.95	0.76

Supplementary Table S2: *Percentage accuracy of determining hypoxic samples from normoxic samples in cancer cell lines, immortalised non-cancer cell lines and a cancer-associated cell experiment for the 53 signatures across four hypoxia scores*

The percentage accuracy is shown in different shades of blue from lowest (light blue) to highest (dark blue). Conventions as in Table 4.2.

Cell line	Hypoxic samples	Normoxic samples	Pairs
MCF-7	91	54	327
MDA-MB-231	22	14	60
HCT116	20	17	58
501mel	16	3	48
HepG2	16	10	46
HeLa	13	14	42
A2780	6	6	36
U87	12	6	36
SiHa	8	8	28
IGR39	9	3	27
LN229	9	3	27
PC-3	9	9	23
Huh-7	12	4	21
MCF10A	7	7	21
EPC2	4	4	16
HT29	15	3	15
DU145	14	2	14
LNCaP	6	6	14
T-47D	5	5	11
Hep3B	4	4	10
Astrocytes-fetal-brain	3	3	9
L3.6pl	3	3	9
HKC8	3	3	9
AsPC-1	3	3	9
SK-MEL-28	3	3	9
SKOV3ip.1	3	3	9
DCs	3	3	9
UFH-001	3	3	9
WM793B	3	3	9
SK-OV-3	3	3	9
Normal gastric myofibroblasts	3	3	9

FG	3	3	9
LM	4	2	8
Cancer-associated myofibroblasts gastric tumour	2	3	6
ZR-75-1	5	2	5
HCC1806	3	3	5
P493-6	2	2	4
LM2	2	2	4
DKO3	2	2	4
TC-252	2	2	4
DAOY	3	1	3
CaSki	2	2	2
MKN28	2	1	2
MDA-MB-468	1	1	1
MCF12A	1	1	1
HS578T	1	1	1
HME2	1	1	1
HCC38	1	1	1
MDA-MB-157	1	1	1
HCC1937	1	1	1
MDA-MB-175	1	1	1
MDA-MB-231-PSOC	1	1	1
MDA-MB-436	1	1	1
SK-HEP-1	1	1	1
SUM149	1	1	1
SKBR3	1	1	1
SUM1315	1	1	1
HCC1428	1	1	1
SUM159	1	1	1
SUM185	1	1	1
SUM225CWN	1	1	1
SUM229	1	1	1
hTERT-HME	1	1	1

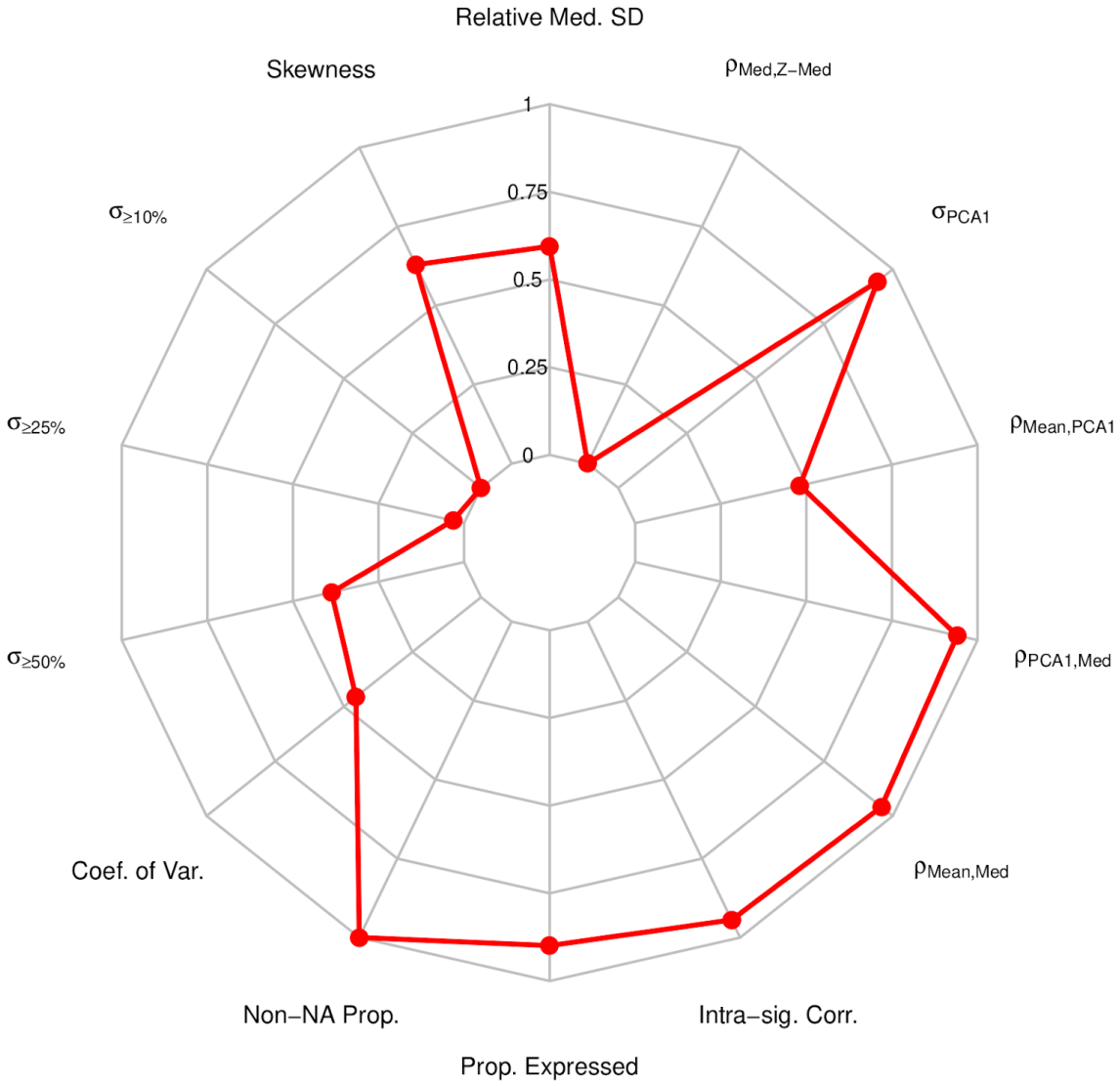
SMMC-7721	1	1	1
NCH660h	1	1	1
NCH644	1	1	1
NCH601	1	1	1
NCH421k	1	1	1
HCC1569	1	1	1
BT549	1	1	1
HBL100	1	1	1
DU4475	1	1	1
WiDr	1	1	1
22Rv1	1	1	1
PLC-PRF-5	1	1	1
LoVo	1	1	1
HT1080	1	1	1
HCT-15	1	1	1
COLO-205	1	1	1
A10.7	1	1	1
A125	1	1	1
A13D	1	1	1
A2.4	1	1	1
A32.4	1	1	1
A38.44	1	1	1
A38.5	1	1	1
A6L	1	1	1
C-33	1	1	1
C-41	1	1	1
HT-3	1	1	1
ME-180	1	1	1
SW756	1	1	1
JJN3	1	1	1
BT20	1	1	1
BT474	1	1	1
U373	1	1	1

CAMA1	1	1	1
A38.41	1	1	1

Supplementary Table S3: *Cell lines used in hypoxia experiments in GEO and included in the pan-cancer hypoxia signature analysis and across all experimental conditions*

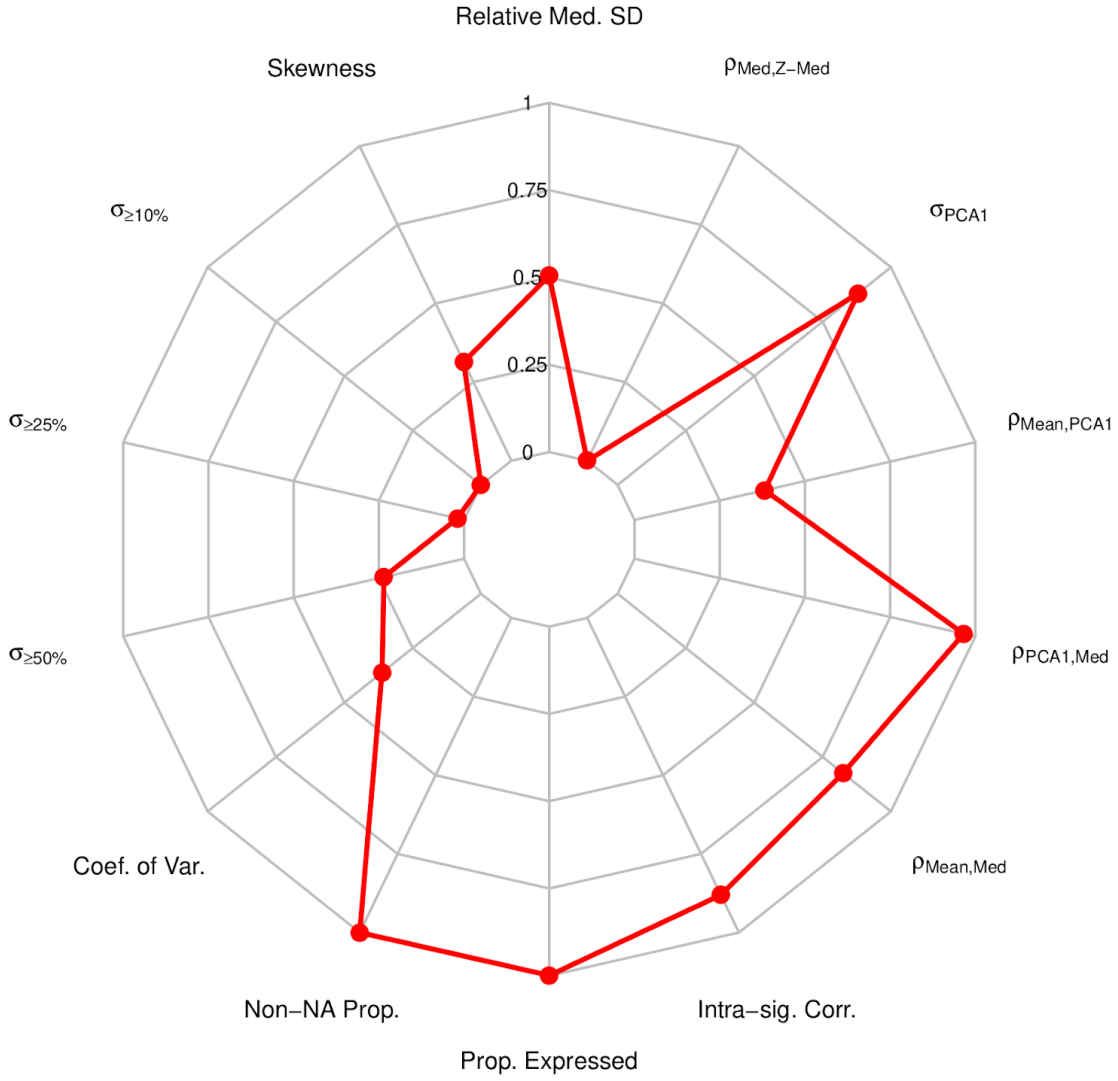
Details of the hypoxia versus normoxia experiments in GEO using all cancer, cancer-associated and non-cancer cell lines across all the experiments. The table reports a total number of 1090 pairs.

Appendix 3



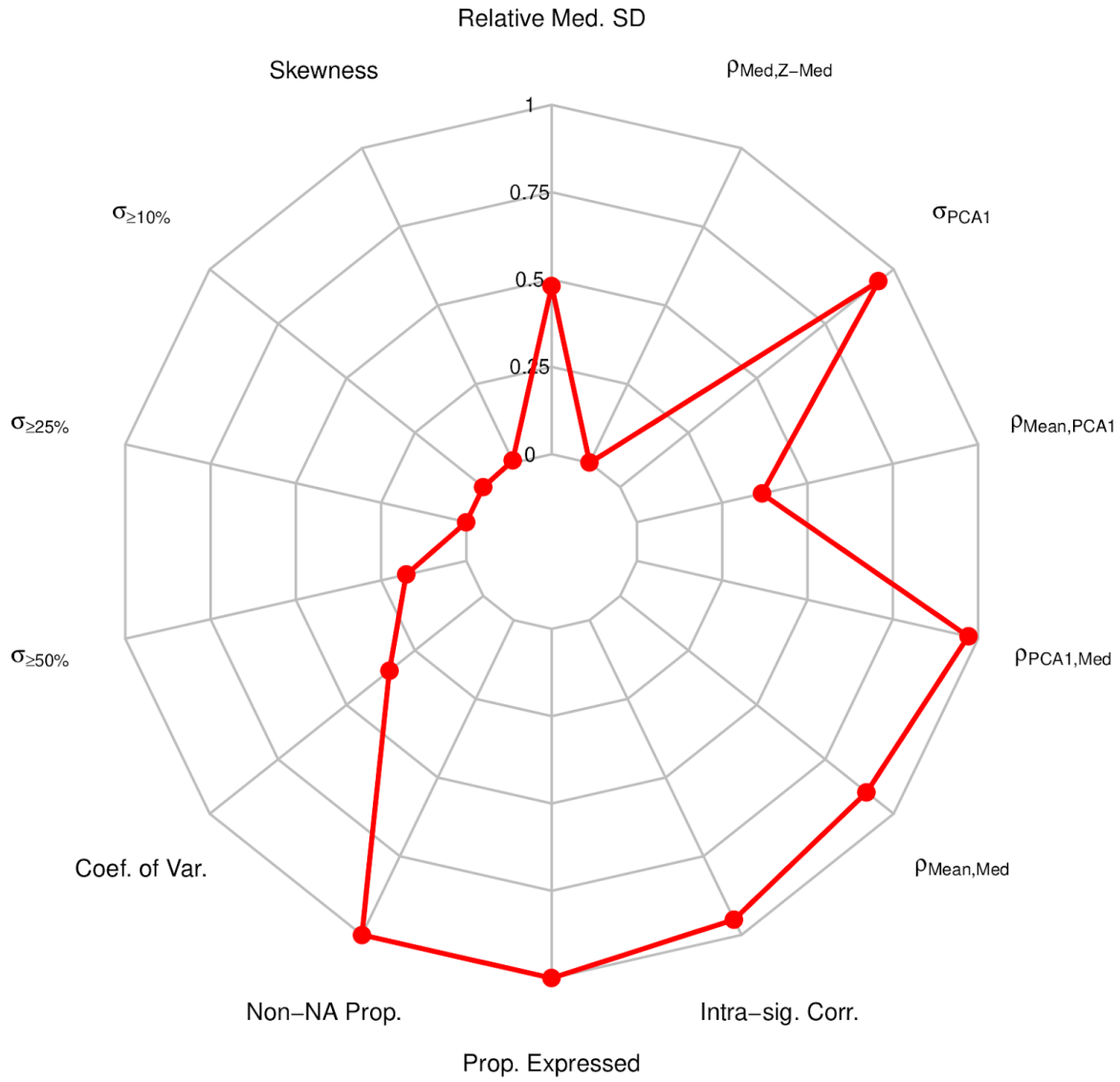
Supplementary Figure S1: *Radar plot representing the performance of Ragnum signature in breast cancer samples from the TCGA*

The radar plot shows the performance of the Ragnum signature based on several measures described fully in sigQC. Notably, the proportion of expressed genes in the signature is very high and there are not any NA values in the signature genes. More information on sigQC metrics can be found in Chapter 3.



Supplementary Figure S2: *Radar plot representing the performance of Buffa signature in breast cancer samples from the TCGA*

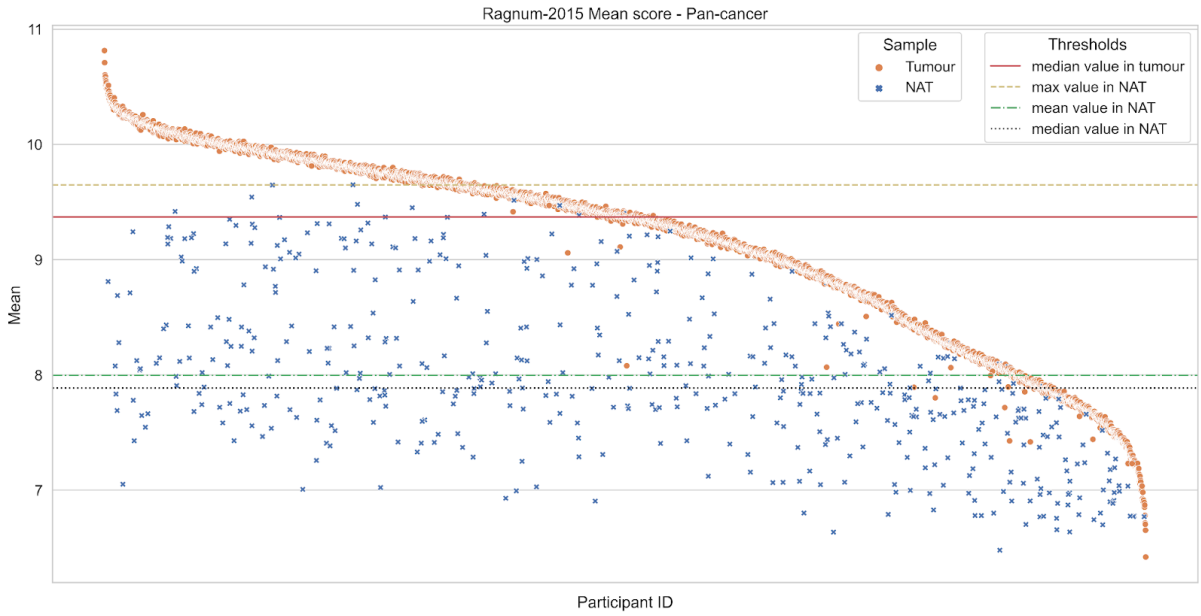
The radar plot shows the performance of the Buffa signature based on several measures described fully in sigQC. Notably, the proportion of expressed genes in the signature is very high and there are not any NA values in the signature genes. More information on sigQC metrics can be found in Chapter 3.



Supplementary Figure S3: *Radar plot representing the performance of Ghazoui signature in breast cancer samples from the TCGA*

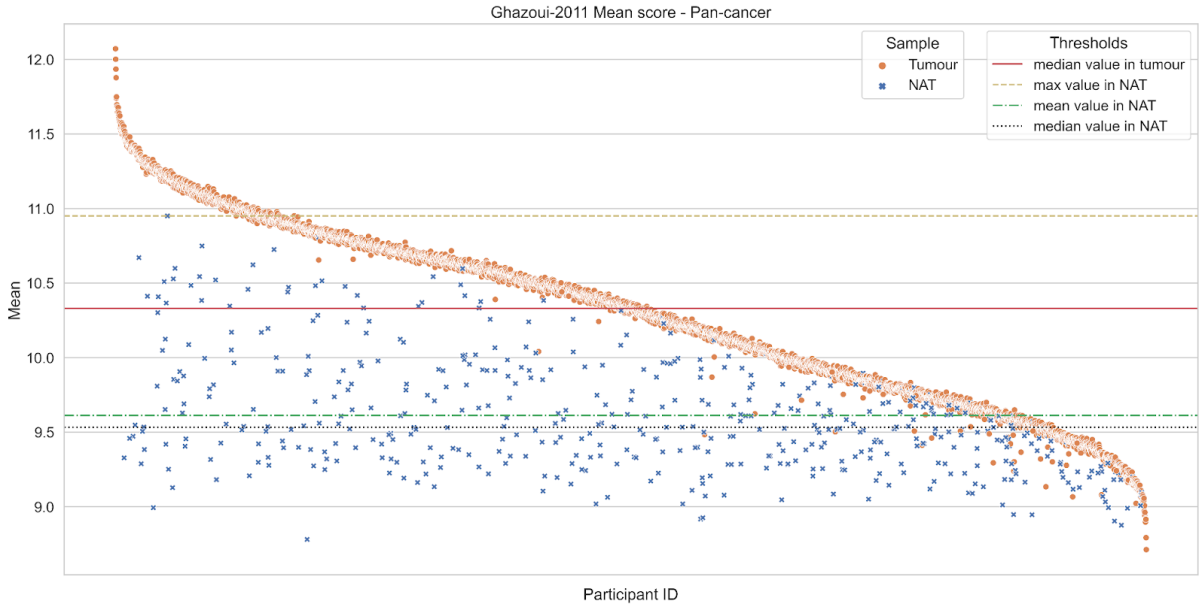
The radar plot shows the performance of the Ghazoui signature based on several measures described fully in sigQC. Notably, the proportion of expressed genes in the signature is very high and there are not any NA values in the signature genes. More information on sigQC metrics can be found in Chapter 3.

Appendix 4



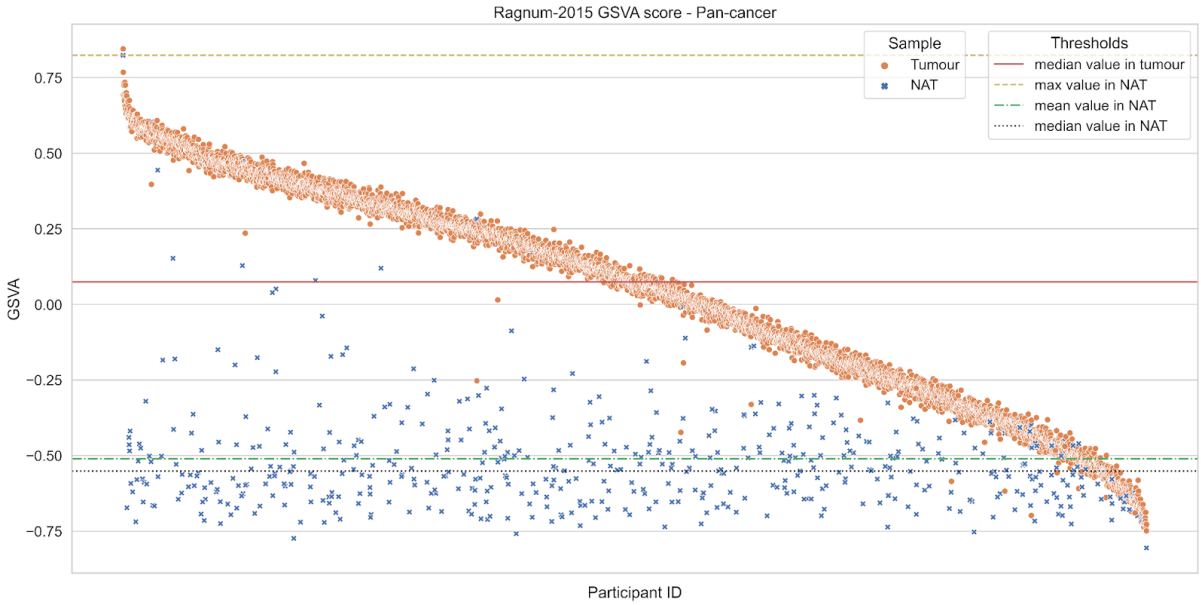
Supplementary Figure S4: *Line plot representing the distribution of hypoxia scores for each samples in the TCGA cohort (10 cancer types) including NAT using mean score with Ragnum signature*

Each data point represents a sample in the TCGA cohort. All NAT samples are shown with a blue cross and all tumour samples are shown with an orange circle. Conventions as in *Fig. 6.4*.



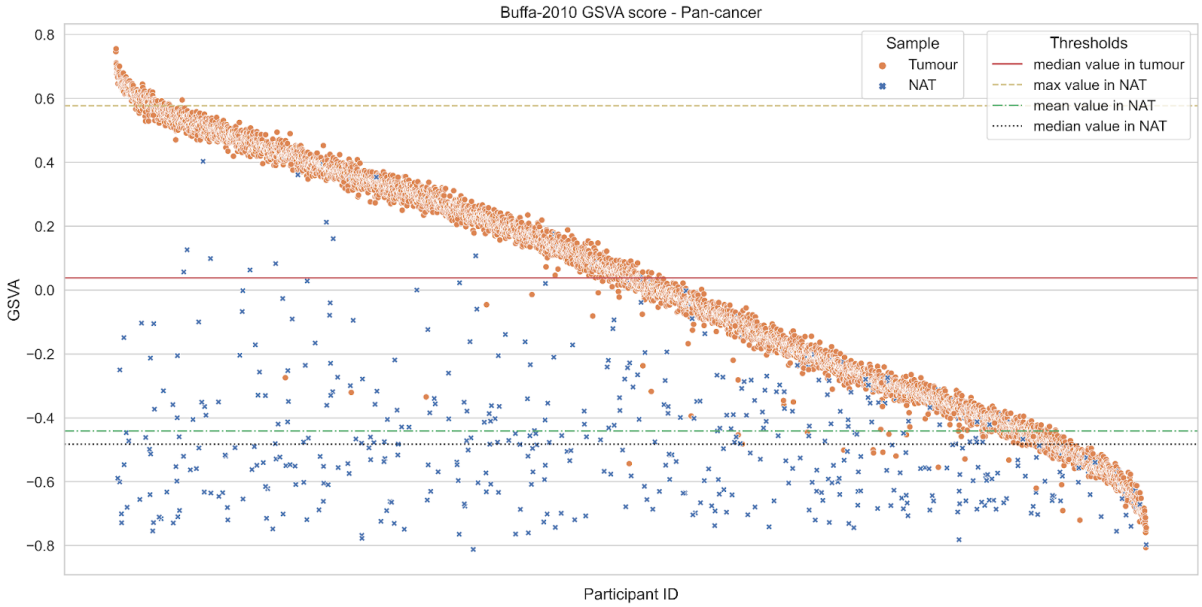
Supplementary Figure S5: *Line plot representing the distribution of hypoxia scores for each samples in the TCGA cohort (10 cancer types) including NAT using mean score with Ghazoui signature*

Each data point represents a sample in the TCGA cohort. All NAT samples are shown with a blue cross and all tumour samples are shown with an orange circle. Conventions as in Fig. 6.4.



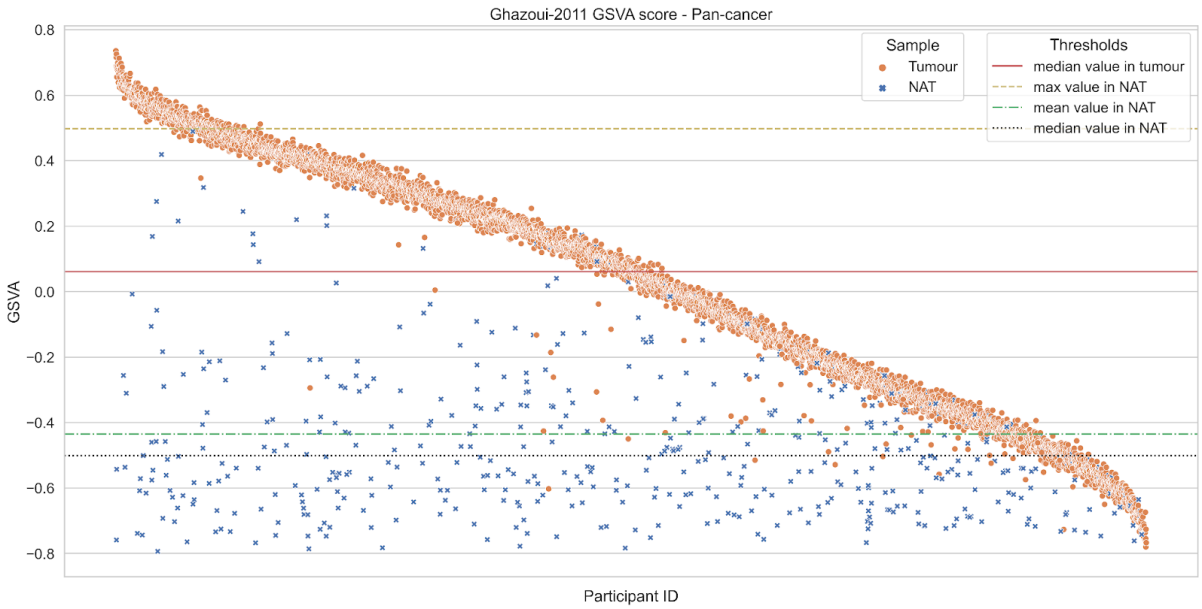
Supplementary Figure S6: *Line plot representing the distribution of hypoxia scores for each samples in the TCGA cohort (10 cancer types) including NAT using GSVA score with Ragnum signature*

Each data point represents a sample in the TCGA cohort. All NAT samples are shown with a blue cross and all tumour samples are shown with an orange circle. Conventions as in *Fig. 6.4*.



Supplementary Figure S7: *Line plot representing the distribution of hypoxia scores for each samples in the TCGA cohort (10 cancer types) including NAT using GSVA score with Buffa signature*

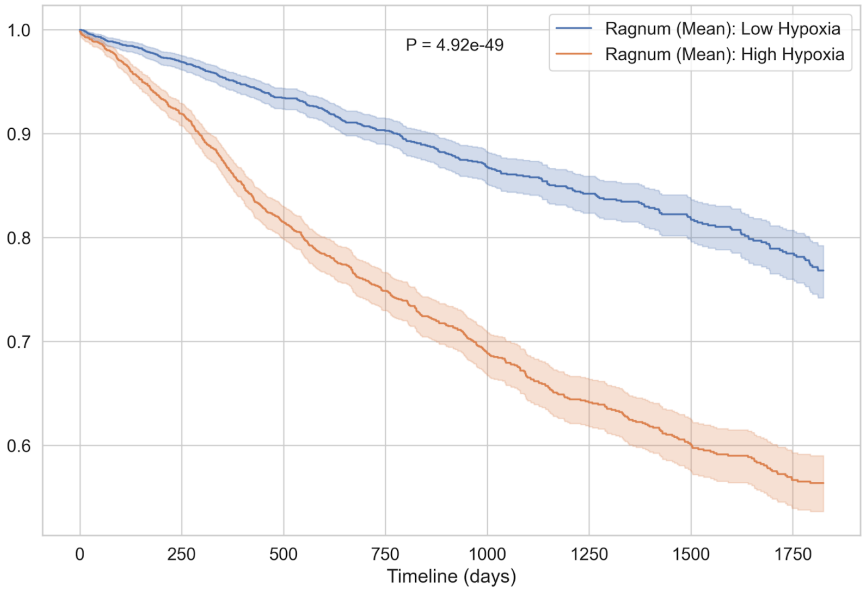
Each data point represents a sample in the TCGA cohort. All NAT samples are shown with a blue cross and all tumour samples are shown with an orange circle. Conventions as in Fig. 6.4.



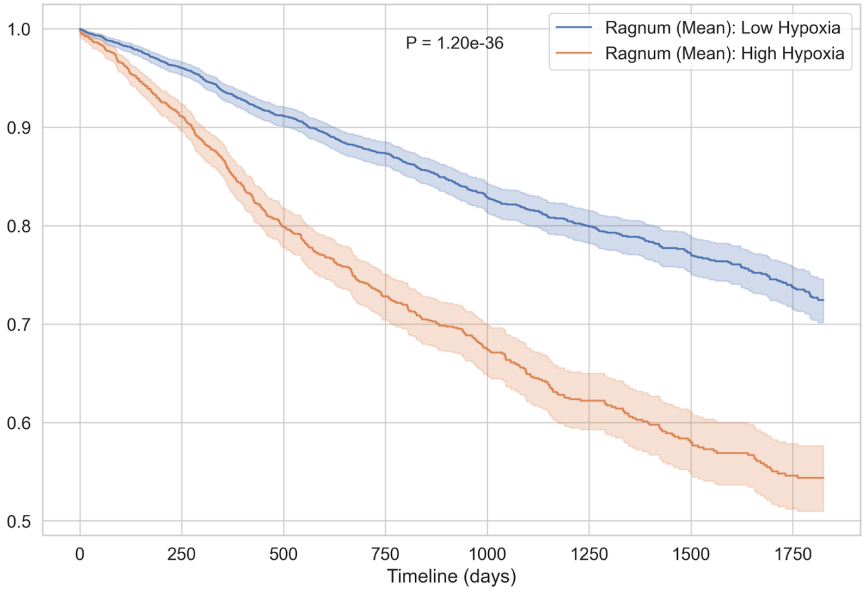
Supplementary Figure S8: *Line plot representing the distribution of hypoxia scores for each samples in the TCGA cohort (10 cancer types) including NAT using GSVA score with Ghazoui signature*

Each data point represents a sample in the TCGA cohort. All NAT samples are shown with a blue cross and all tumour samples are shown with an orange circle. Conventions as in *Fig. 6.4*.

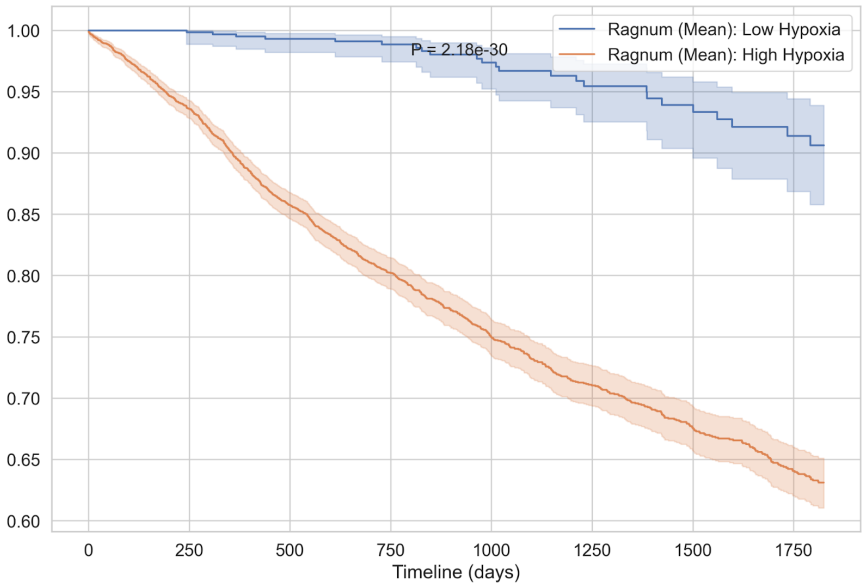
a)


P = 4.92e-49
Ragnum (Mean): Low Hypoxia
Ragnum (Mean): High Hypoxia
1.0
0.9
0.8
0.7
0.6
0
250
500
750
1000
1250
1500
1750
Timeline (days)

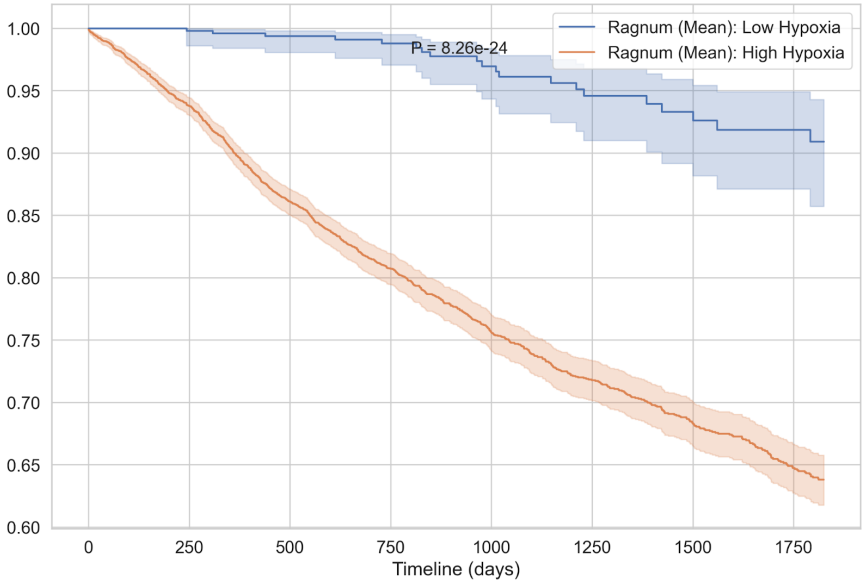
b)


P = 1.20e-36
Ragnum (Mean): Low Hypoxia
Ragnum (Mean): High Hypoxia
1.0
0.9
0.8
0.7
0.6
0.5
0
250
500
750
1000
1250
1500
1750
Timeline (days)

c)



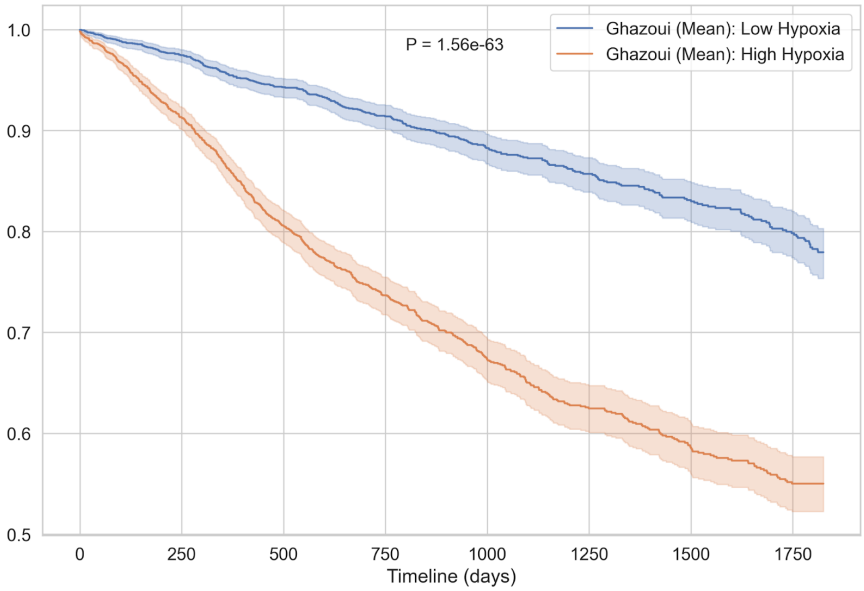
d)



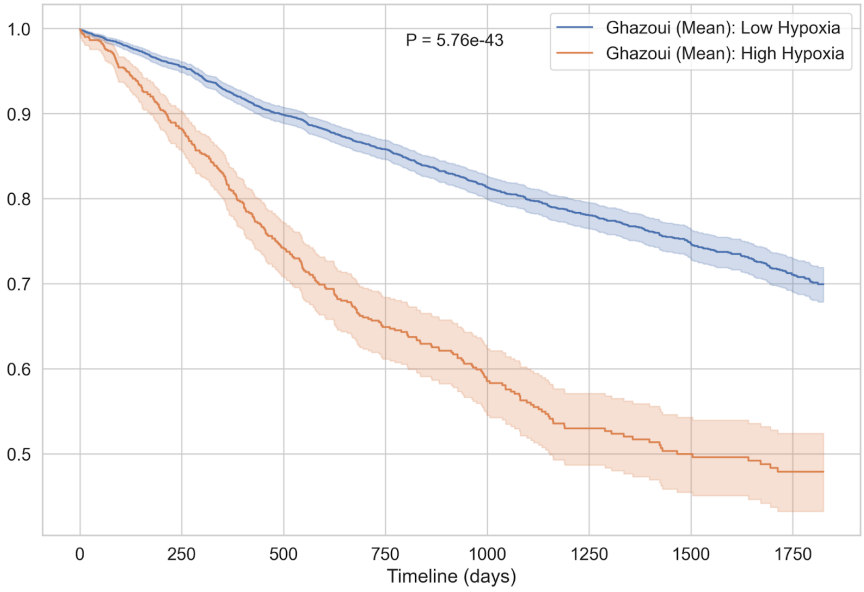
Supplementary Figure S9: *Kaplan-Meier survival curve compared with the log-rank test obtained using the three methods to split samples in high and low mean hypoxia score using the Ragnum signature.*

KM survival curve for all tumour and NAT samples in the TCGA cohort using the (a) traditional method, (b) maximum on NAT, (c) mean on NAT and (d) median on NAT with Ragnum signature and the Mean hypoxia score. Conventions as per *Fig. 6.5*.

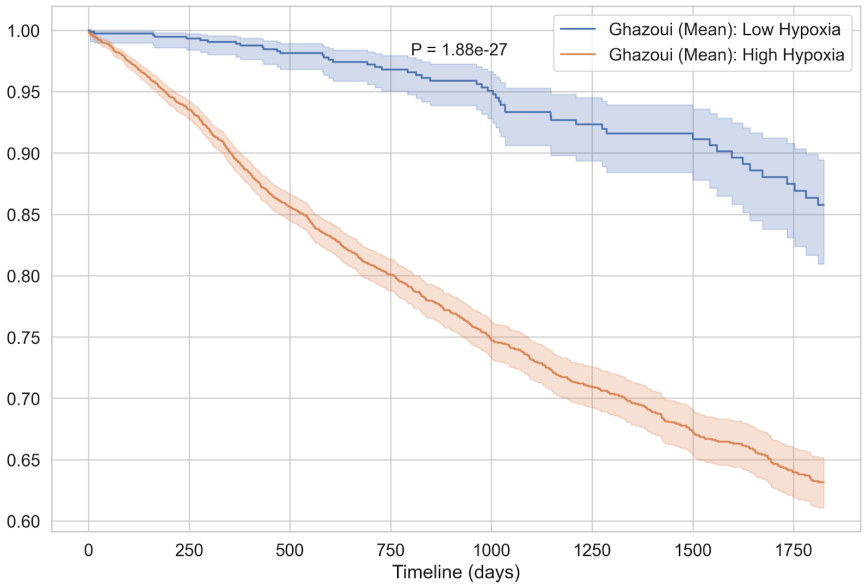
a)


P = 1.56e-63
Ghazoui (Mean): Low Hypoxia
Ghazoui (Mean): High Hypoxia
1.0
0.9
0.8
0.7
0.6
0.5
0
250
500
750
1000
1250
1500
1750
Timeline (days)

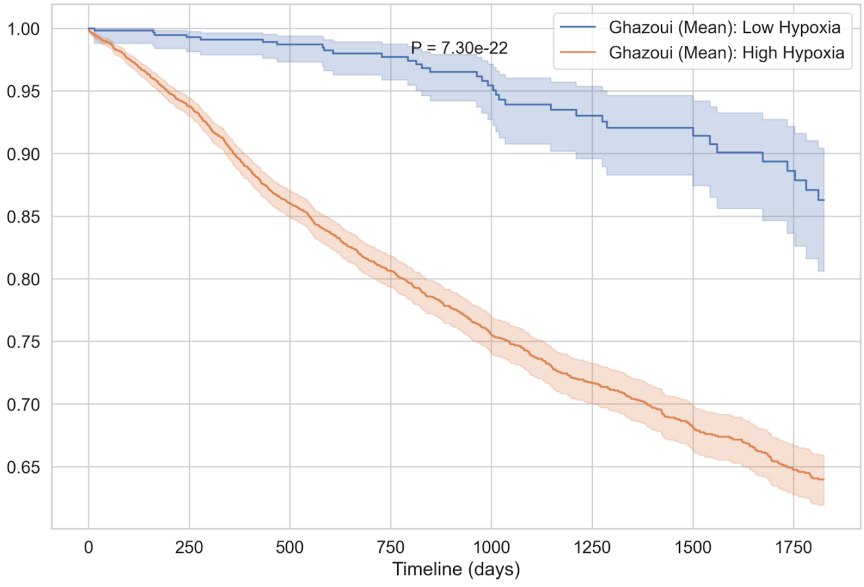
b)


P = 5.76e-43
Ghazoui (Mean): Low Hypoxia
Ghazoui (Mean): High Hypoxia
1.0
0.9
0.8
0.7
0.6
0.5
0
250
500
750
1000
1250
1500
1750
Timeline (days)

c)



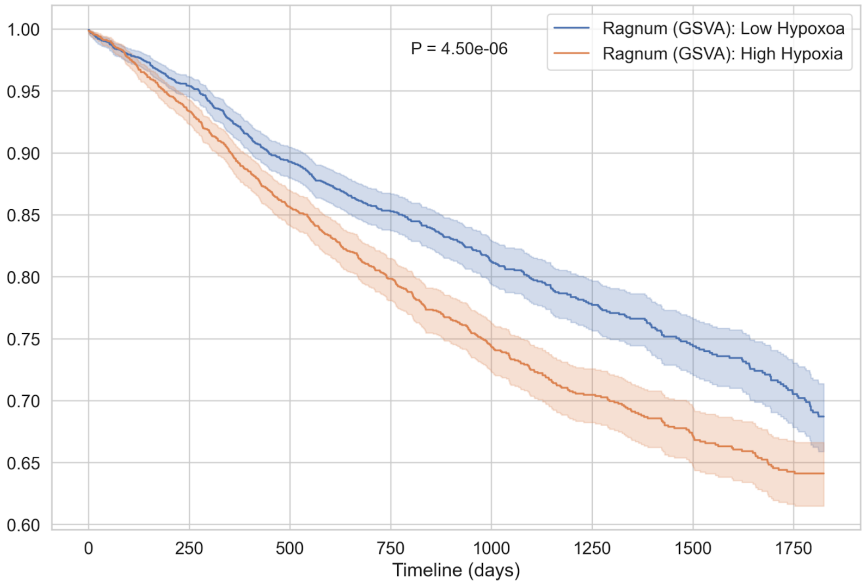
d)



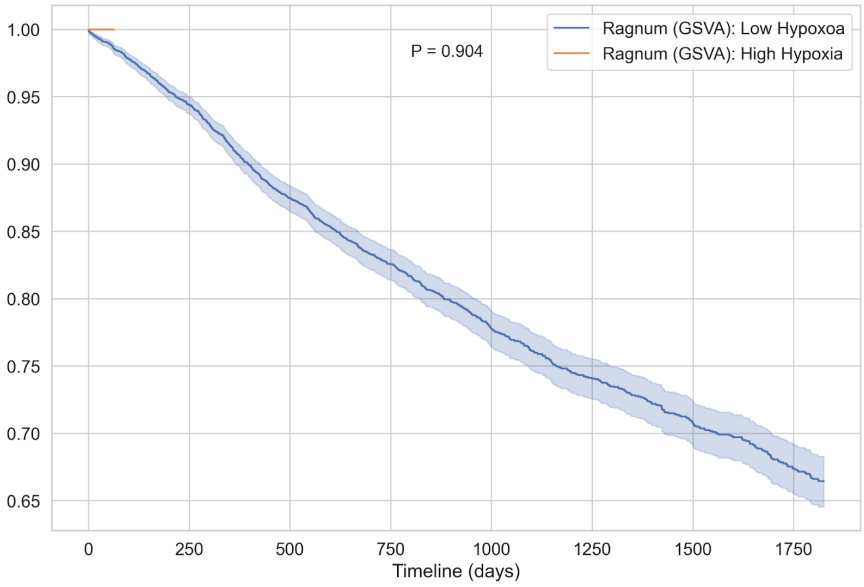
Supplementary Figure S10: *Kaplan-Meier survival curve compared with the log-rank test obtained using the three methods to split samples in high and low mean hypoxia score using the Ghazoui signature.*

KM survival curve for all tumour and NAT samples in the TCGA cohort using the (a) traditional method, (b) maximum on NAT, (c) mean on NAT and (d) median on NAT with Ghazoui signature and the mean hypoxia score. Conventions as per *Fig. 6.5*.

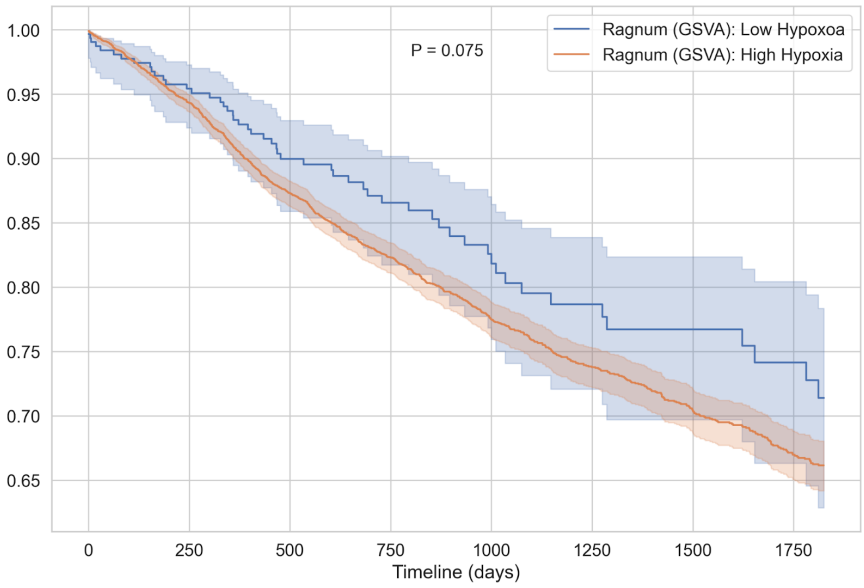
a)


Ragnum (GSVA): Low Hypoxoa
Ragnum (GSVA): High Hypoxia
P = 4.50e-06
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
250
500
750
1000
1250
1500
1750
Timeline (days)

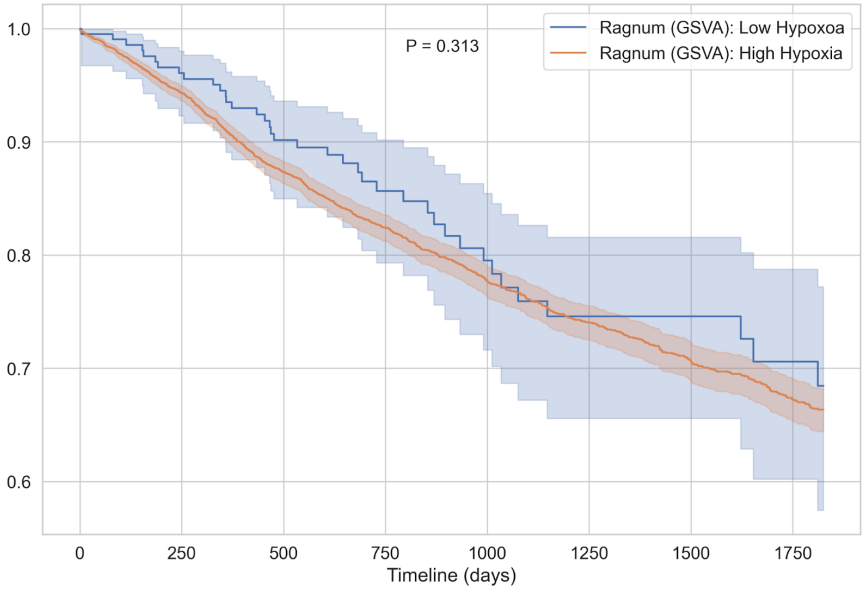
b)


Ragnum (GSVA): Low Hypoxoa
Ragnum (GSVA): High Hypoxia
P = 0.904
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0
250
500
750
1000
1250
1500
1750
Timeline (days)

c)



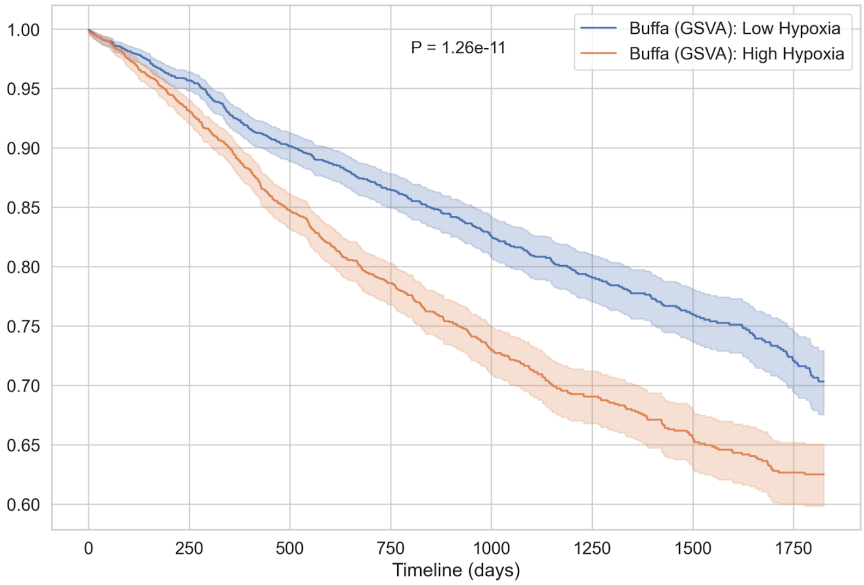
d)



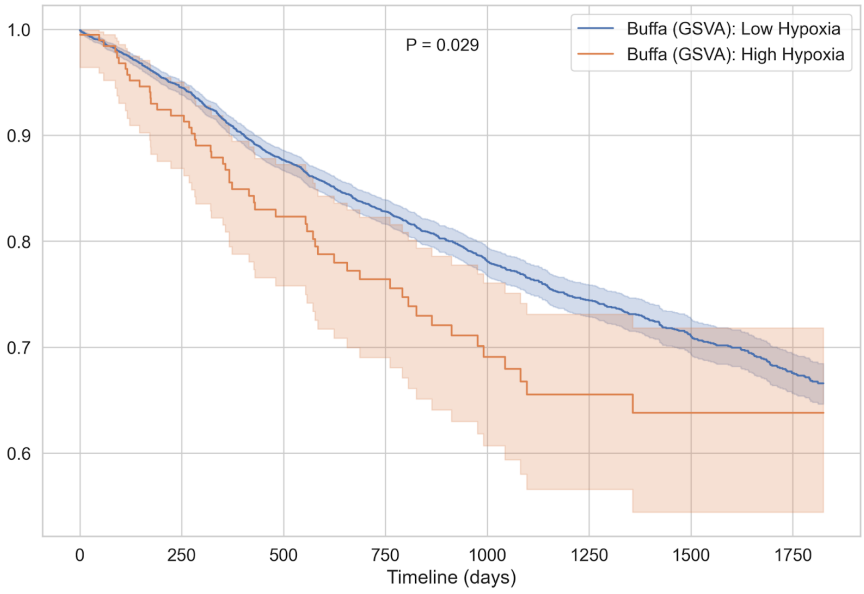
Supplementary Figure S11: *Kaplan-Meier survival curve compared with the log-rank test obtained using the three methods to split samples in high and low GSVA hypoxia score using the Ragnum signature.*

KM survival curve for all tumour and NAT samples in the TCGA cohort using the (a) traditional method, (b) maximum on NAT, (c) mean on NAT and (d) median on NAT with Ragnum signature and the GSVA hypoxia score. Conventions as per *Fig. 6.5*.

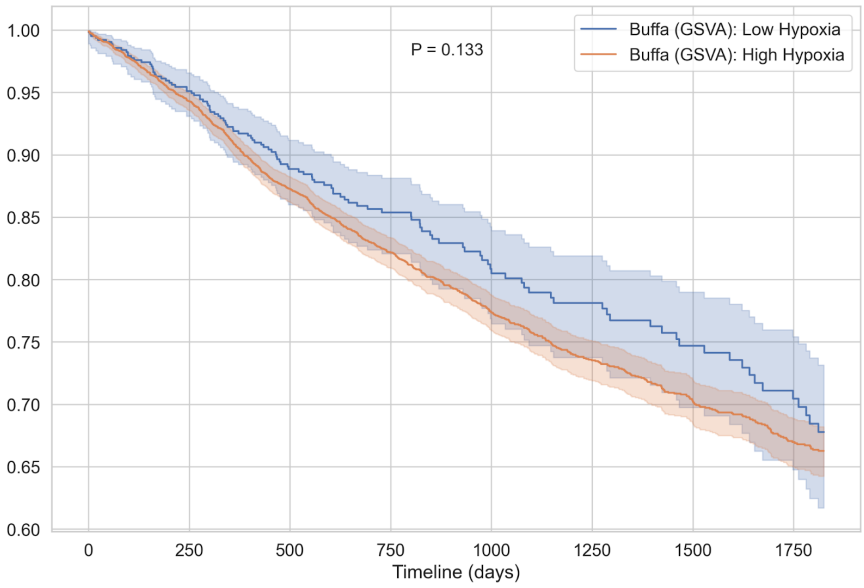
a)


P = 1.26e-11
Buffa (GSVA): Low Hypoxia
Buffa (GSVA): High Hypoxia
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
250
500
750
1000
1250
1500
1750
Timeline (days)

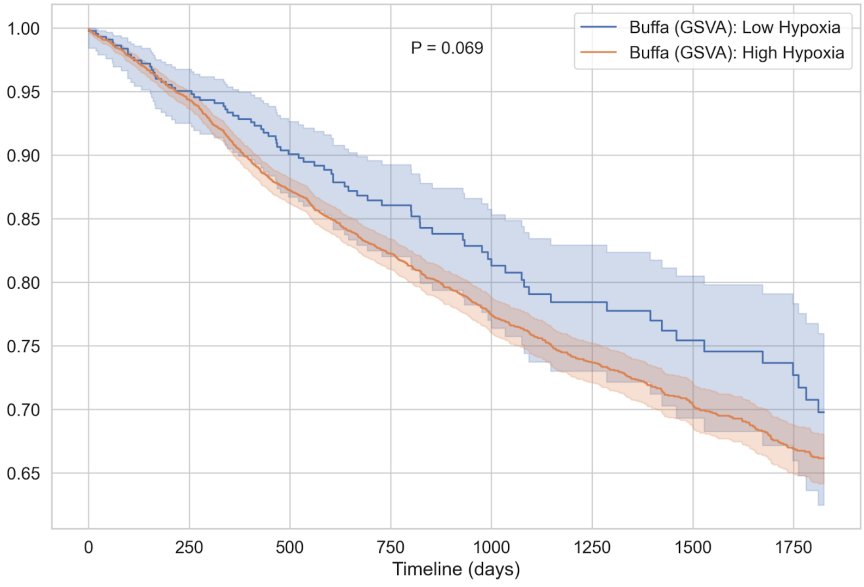
b)


P = 0.029
Buffa (GSVA): Low Hypoxia
Buffa (GSVA): High Hypoxia
1.0
0.9
0.8
0.7
0.6
0
250
500
750
1000
1250
1500
1750
Timeline (days)

c)



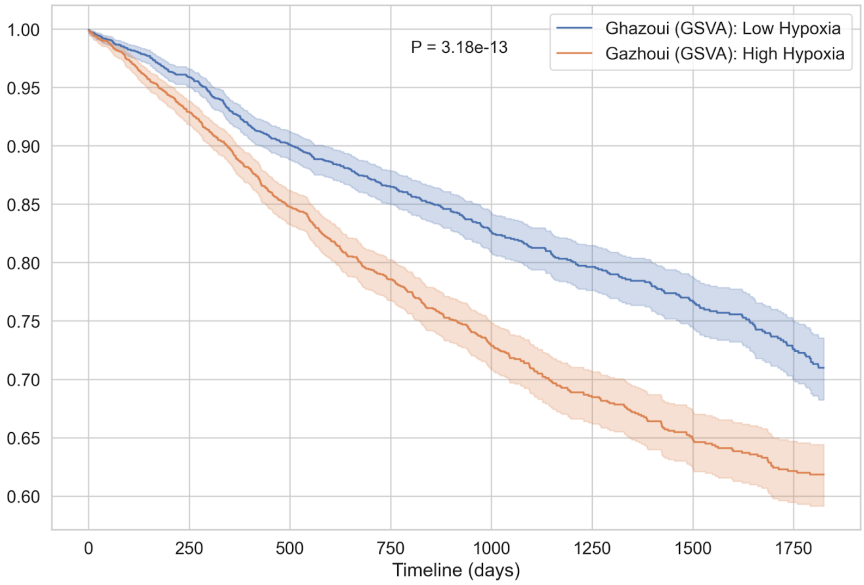
d)



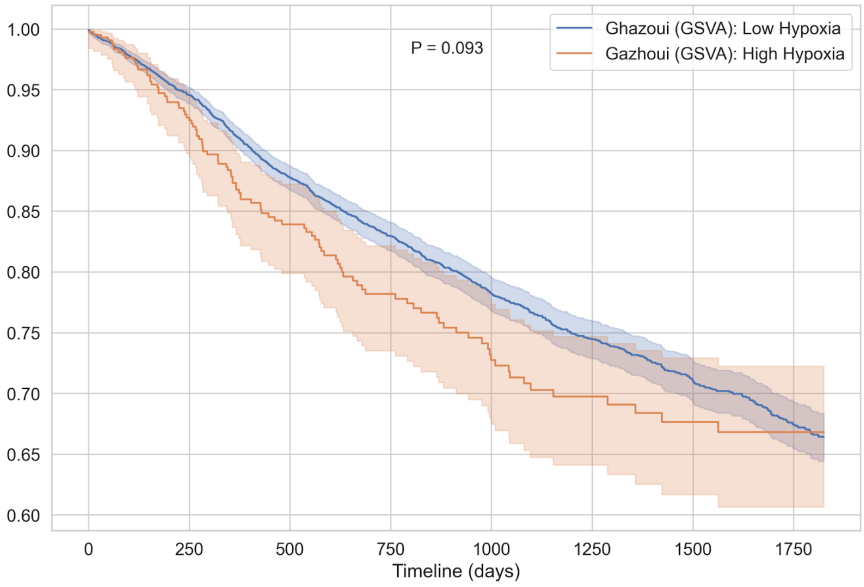
Supplementary Figure S12: *Kaplan-Meier survival curve compared with the log-rank test obtained using the three methods to split samples in high and low GSVA hypoxia score using the Buffa signature.*

KM survival curve for all tumour and NAT samples in the TCGA cohort using the (a) traditional method, (b) maximum on NAT, (c) mean on NAT and (d) median on NAT with Buffa signature and the GSVA hypoxia score. Conventions as per *Fig. 6.5*.

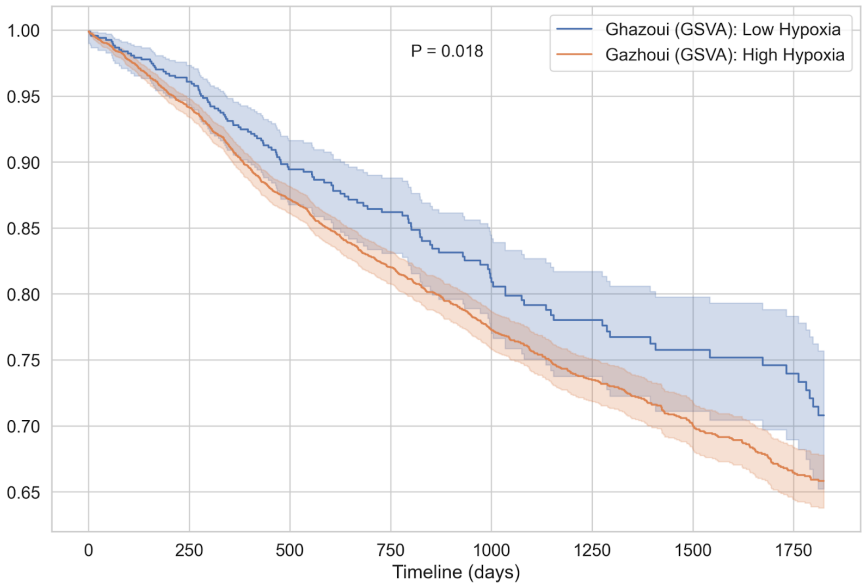
a)


P = 3.18e-13
Ghazoui (GSVA): Low Hypoxia
Gazhoui (GSVA): High Hypoxia
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
250
500
750
1000
1250
1500
1750
Timeline (days)

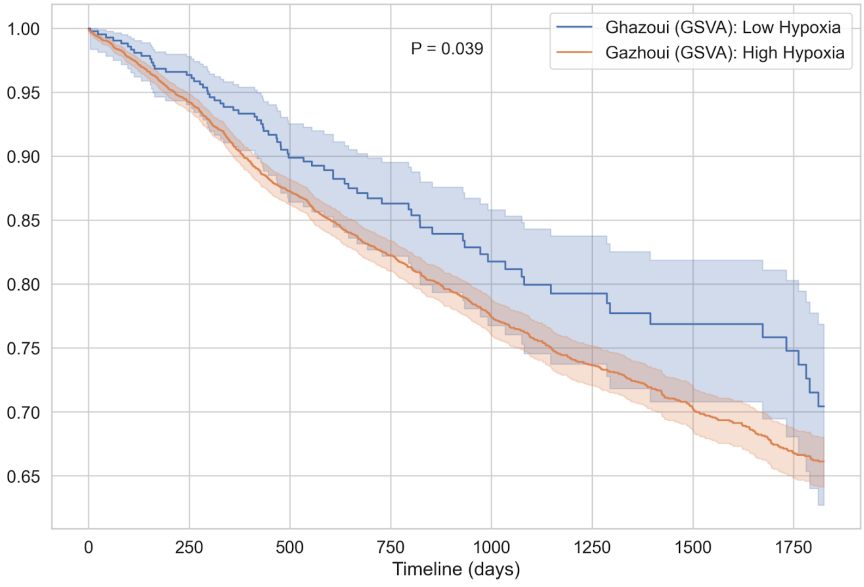
b)


P = 0.093
Ghazoui (GSVA): Low Hypoxia
Gazhoui (GSVA): High Hypoxia
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
250
500
750
1000
1250
1500
1750
Timeline (days)

c)



d)

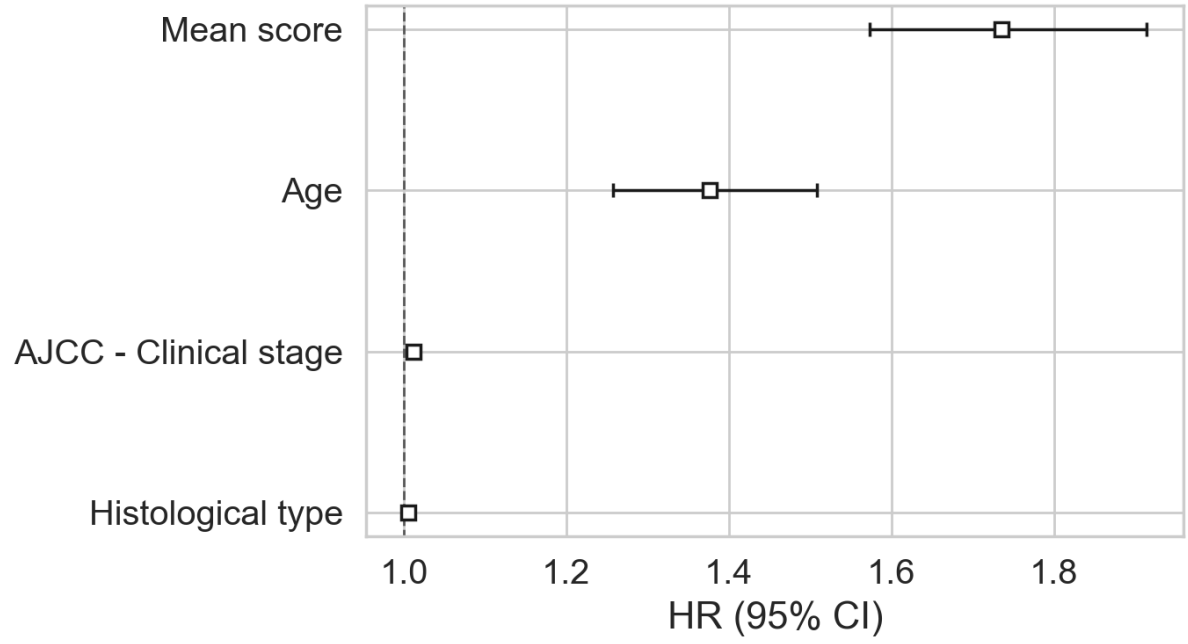


Supplementary Figure S13: *Kaplan-Meier survival curve compared with the log-rank test obtained using the three methods to split samples in high and low GSVA hypoxia score using the Ghazoui signature.*

KM survival curve for all tumour and NAT samples in the TCGA cohort using the (a) traditional method, (b) maximum on NAT, (c) mean on NAT and (d) median on NAT with Ghazoui signature and the GSVA hypoxia score. Conventions as per *Fig. 6.5*.

Appendix 5

a)



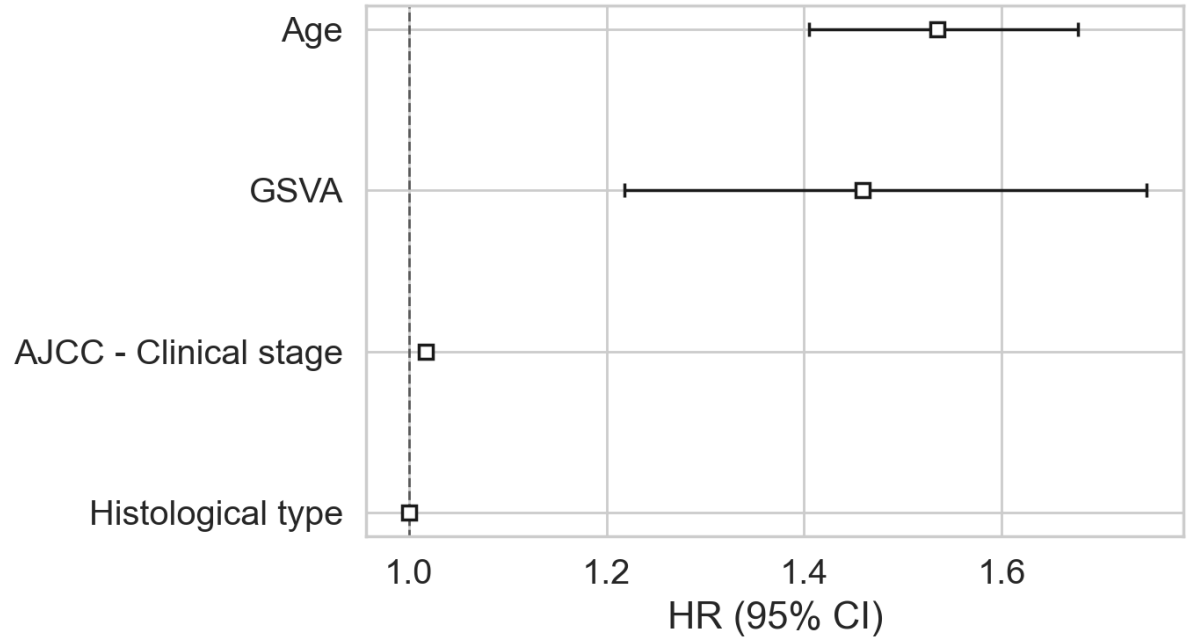
b)

covariate	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
AJCC - Clinical stage	0.011518	1.011585	0.001738	0.008111	0.014925	1.008144	1.015037	0.0	6.626322	3.441532e-11	34.758158
Age	0.319764	1.376803	0.046421	0.228781	0.410748	1.257067	1.507945	0.0	6.888368	5.643597e-12	37.366522
Histological type	0.005748	1.005764	0.003047	-0.000224	0.011720	0.999776	1.011789	0.0	1.886343	5.924870e-02	4.077073
Mean score	0.551349	1.735593	0.049966	0.453417	0.649282	1.573680	1.914165	0.0	11.034412	2.607503e-28	91.631318

Supplementary Figure S14: *Cox-proportional hazards model on the TCGA tumour samples included in the study across ten cancer types using the mean hypoxia score on Ragnum signature*

HR values for the Pan-cancer Cox-proportional hazards model (a) using the mean score with Ragnum signature. The summary table reporting coefficients, standard error, z and p-values are shown in (b).

a)



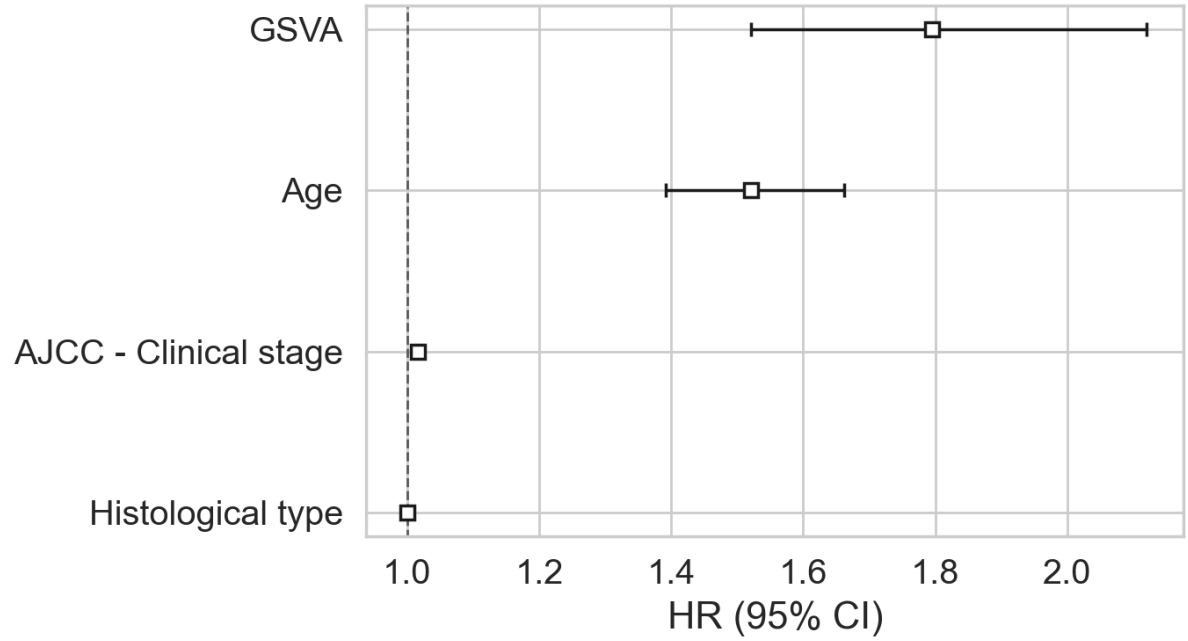
b)

covariate	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
AJCC - Clinical stage	0.017004	1.017149	0.001627	0.013815	0.020193	1.013911	1.020398	0.0	10.451560	1.441344e-25	82.520788
Age	0.428333	1.534697	0.045215	0.339714	0.516952	1.404545	1.676908	0.0	9.473322	2.710826e-21	68.321758
Histological type	0.000229	1.000229	0.002839	-0.005334	0.005793	0.994680	1.005810	0.0	0.080745	9.356447e-01	0.095967
GSVA	0.377602	1.458783	0.091857	0.197565	0.557640	1.218432	1.746545	0.0	4.110740	3.943937e-05	14.630004

Supplementary Figure S15: *Cox-proportional hazards model on the TCGA tumour samples included in the study across ten cancer types using the GSVA hypoxia score on Ragnum signature*

HR values for the Pan-cancer Cox-proportional hazards model (a) using the GSVA score with Ragnum signature. The summary table reporting coefficients, standard error, z and p-values are shown in (b).

a)



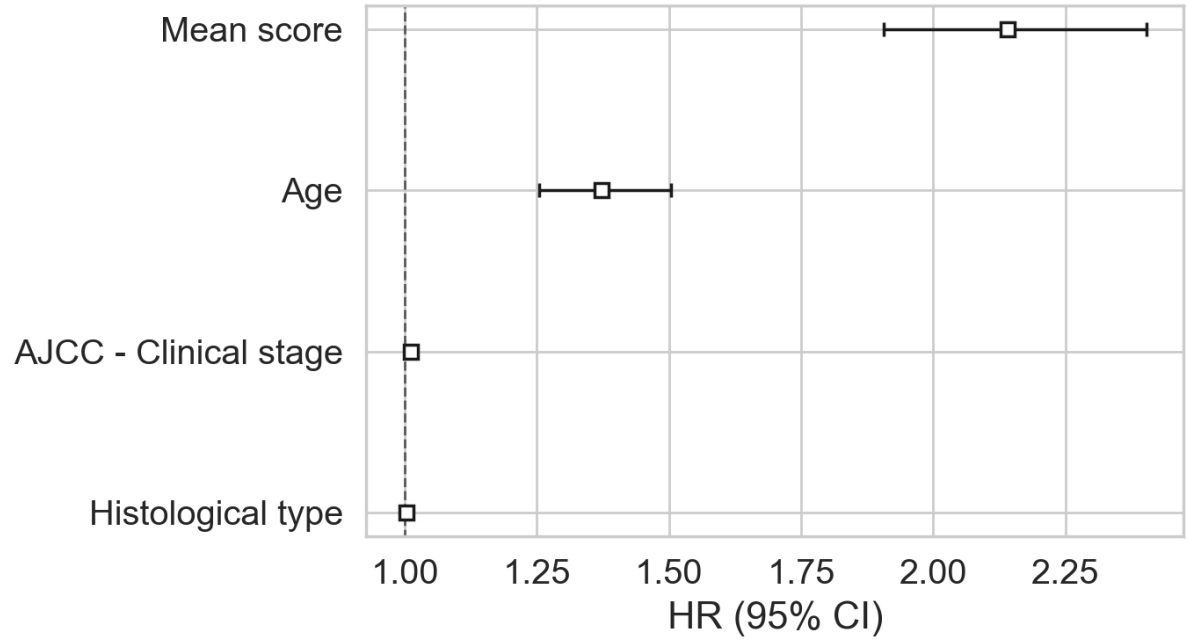
b)

covariate	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
AJCC - Clinical stage	0.016915	1.017059	0.001623	0.013734	0.020095	1.013829	1.020298	0.0	10.423642	1.934030e-25	82.096592
Age	0.419068	1.520544	0.045203	0.330472	0.507665	1.391625	1.661407	0.0	9.270788	1.847759e-20	65.552786
Histological type	0.000195	1.000195	0.002789	-0.005271	0.005660	0.994743	1.005677	0.0	0.069858	9.443068e-01	0.082672
GSVA	0.585274	1.795483	0.084755	0.419157	0.751392	1.520678	2.119949	0.0	6.905443	5.004717e-12	37.539849

Supplementary Figure S16: *Cox-proportional hazards model on the TCGA tumour samples included in the study across ten cancer types using the GSVA hypoxia score on Buffa signature*

HR values for the Pan-cancer Cox-proportional hazards model (a) using the GSVA score with Buffa signature. The summary table reporting coefficients, standard error, z and p-values are shown in (b).

a)



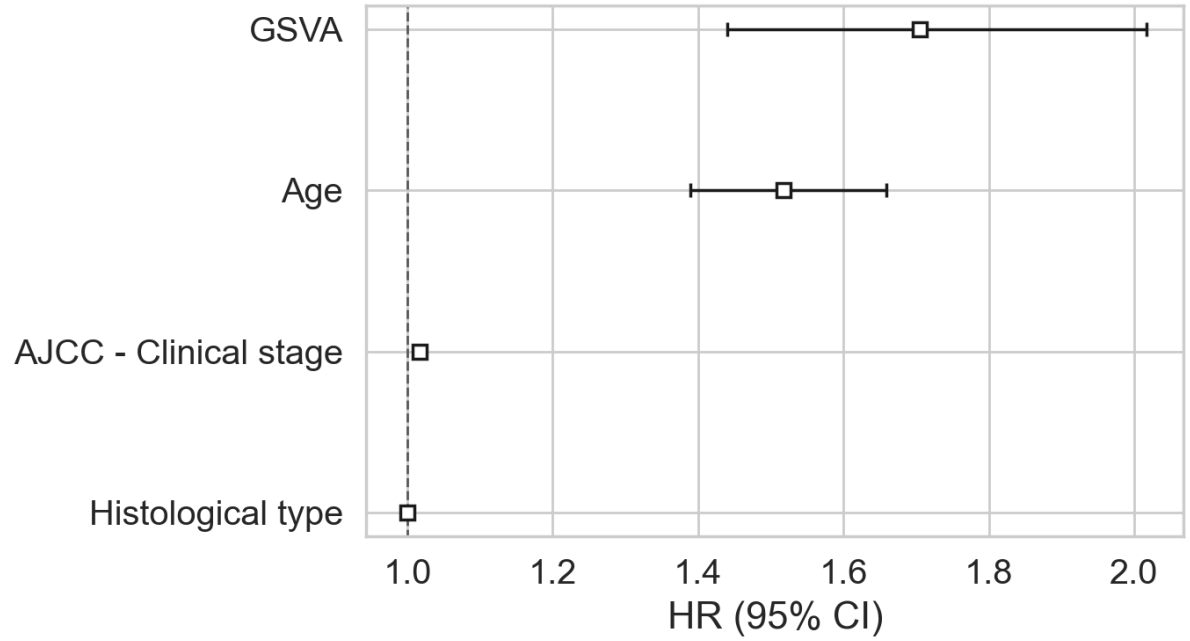
b)

covariate	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
AJCC - Clinical stage	0.011796	1.011866	0.001668	0.008527	0.015064	1.008564	1.015178	0.0	7.073518	1.510540e-12	39.268073
Age	0.317057	1.373081	0.046244	0.226420	0.407694	1.254103	1.503347	0.0	6.856159	7.073653e-12	37.040682
Histological type	0.003751	1.003758	0.003018	-0.002164	0.009666	0.997838	1.009713	0.0	1.242877	2.139133e-01	2.224902
Mean_Scores	0.761418	2.141310	0.059048	0.645686	0.877150	1.907295	2.404039	0.0	12.894869	4.810825e-38	123.966983

Supplementary Figure S17: *Cox-proportional hazards model on the TCGA tumour samples included in the study across ten cancer types using the Mean hypoxia score on Ghazoui signature*

HR values for the Pan-cancer Cox-proportional hazards model (a) using the Mean score with Ghazoui signature. The summary table reporting coefficients, standard error, z and p-values are shown in (b).

a)



b)

covariate	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
AJCC - Clinical stage	0.017021	1.017167	0.001626	0.013834	0.020209	1.013930	1.020414	0.0	10.465956	1.238205e-25	82.739952
Age	0.417489	1.518144	0.045191	0.328916	0.506061	1.389461	1.658745	0.0	9.238297	2.504519e-20	65.114028
Histological type	0.000393	1.000393	0.002802	-0.005099	0.005886	0.994914	1.005903	0.0	0.140371	8.883672e-01	0.170772
GSVA	0.533338	1.704613	0.085897	0.364984	0.701692	1.440491	2.017163	0.0	6.209071	5.329858e-10	30.805184

Supplementary Figure S18: *Cox-proportional hazards model on the TCGA tumour samples included in the study across ten cancer types using the GSVA hypoxia score on Ghazoui signature*

HR values for the Pan-cancer Cox-proportional hazards model (a) using the GSVA score with Ghazoui signature. The summary table reporting coefficients, standard error, z and p-values are shown in (b).

Appendix 6

Supplementary material including Python, R and Bash scripts used to produce the results and the figures available in this work can be found at the following link upon request:

https://github.com/Matteodigg/Thesis_repo_MDG

References

1. Ahmad, A. S., Ormiston-Smith, N. & Sasieni, P. D. Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960. Br. J. Cancer 112, 943–947 (2015).

2. Robert, C. et al. Ipilimumab plus Dacarbazine for Previously Untreated Metastatic Melanoma. N. Engl. J. Med. 364, 2517–2526 (2011).

3. Hodi, F. S. et al. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. N. Engl. J. Med. 363, 711–723 (2010).

4. Chapman, P. B. et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N. Engl. J. Med. 364, 2507–2516 (2011).

5. Tutzauer, J. et al. Breast cancer hypoxia in relation to prognosis and benefit from radiotherapy after breast-conserving surgery in a large, randomised trial with long-term follow-up. Br. J. Cancer 126, 1145–1156 (2022).

6. Theodoropoulos, V. Hypoxia-Inducible Factor 1? Expression Correlates with Angiogenesis and Unfavorable Prognosis in Bladder Cancer. Eur. Urology 46, 200–208 (2004).

7. Griffiths, E. A., Pritchard, S. A., Welch, I. M., Price, P. M. & West, C. M. Is the hypoxia-inducible factor pathway important in gastric cancer? Eur. J. Cancer 41, 2792–2805 (2005).

8. Brizel, D. M., Sibley, G. S., Prosnitz, L. R., Scher, R. L. & Dewhirst, M. W. Tumor hypoxia adversely affects the prognosis of carcinoma of the head and neck. Int. J. Radiat. Oncol. Biol. Phys. 38, 285–289 (1997).

9. Xiang, Z.-L. et al. Gene expression profiling of fixed tissues identified hypoxia-inducible factor-1α, VEGF, and matrix metalloproteinase-2 as biomarkers of lymph node metastasis in hepatocellular carcinoma. Clin. Cancer Res. 17, 5463–5472 (2011).

10. Hung, J.-J. et al. Prognostic significance of hypoxia-inducible factor-1 , TWIST1 and Snail expression in resectable non-small cell lung cancer. Thorax 64, 1082–1089 (2009).

11. Tanaka, N. et al. Expression of carbonic anhydrase 9, a potential intrinsic marker of hypoxia, is associated with poor prognosis in oesophageal squamous cell carcinoma. Br. J. Cancer 99, 1468–1475 (2008).

12. Milosevic, M. et al. Tumor hypoxia predicts biochemical failure following radiotherapy for clinically localized prostate cancer. Clin. Cancer Res. 18, (2012).

13. 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 (2015).

14. Pugh, C. W. & Ratcliffe, P. J. Regulation of angiogenesis by hypoxia: role of the HIF system. Nat. Med. 9, 677–684 (2003).

15. Semenza, G. L. & Wang, G. L. A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. Mol. Cell. Biol. 12, 5447–5454 (1992).

16. Maxwell, P. H., Pugh, C. W. & Ratcliffe, P. J. Inducible operation of the erythropoietin 3’ enhancer in multiple cell lines: evidence for a widespread oxygen-sensing mechanism. Proc. Natl. Acad. Sci. U. S. A. 90, 2423–2427 (1993).

17. Makino, Y. et al. Inhibitory PAS domain protein is a negative regulator of hypoxia-inducible gene expression. Nature 414, 550–554 (2001).

18. Moskwa, P. et al. miR-182-mediated downregulation of BRCA1 impacts DNA repair and sensitivity to PARP inhibitors. Mol. Cell 41, 210–220 (2011).

19. Camps, C. et al. hsa-miR-210 Is induced by hypoxia and is an independent prognostic factor in breast cancer. Clin. Cancer Res. 14, 1340–1348 (2008).

20. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. Cell 144, (2011).

21. Semenza, G. L. HIF-1: using two hands to flip the angiogenic switch. Cancer Metastasis Rev. 19, (2000).

22. Al Tameemi, W., Dale, T. P., Al-Jumaily, R. M. K. & Forsyth, N. R. Hypoxia-Modified Cancer Cell Metabolism. Front. Cell Dev. Biol. 0, (2019).

23. Stüben, G. et al. Erythropoietin restores the anemia-induced reduction in radiosensitivity of experimental human tumors in nude mice. Int. J. Radiat. Oncol. Biol. Phys. 55, 1358–1362 (2003).

24. Henke, M. et al. Erythropoietin to treat head and neck cancer patients with anaemia undergoing radiotherapy: randomised, double-blind, placebo-controlled trial. Lancet 362, 1255–1260 (2003).

25. Wardman, P. Nitroimidazoles as hypoxic cell radiosensitizers and hypoxia probes: misonidazole, myths and mistakes. Br. J. Radiol. 92, 20170915 (2019).

26. 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, (2010).

27. Cade, I. S. et al. Hyperbaric oxygen and radiotherapy: a Medical Research Council trial in carcinoma of the bladder. Br. J. Radiol. 51, 876–878 (1978).

28. Binley, K. et al. Hypoxia-mediated tumour targeting. Gene Ther. 10, 540–549 (2003).

29. Hicks, K. O. et al. Oxygen dependence and extravascular transport of hypoxia-activated prodrugs: comparison of the dinitrobenzamide mustard PR-104A and tirapazamine. Int. J. Radiat. Oncol. Biol. Phys. 69, 560–571 (2007).

30. Liu, S. K. et al. A novel poly(ADP-ribose) polymerase inhibitor, ABT-888, radiosensitizes malignant human cell lines under hypoxia. Radiother. Oncol. 88, 258–268 (2008).

31. Ashton, T. M. et al. The anti-malarial atovaquone increases radiosensitivity by alleviating tumour hypoxia. Nat. Commun. 7, 12308 (2016).

32. Skwarski, M. et al. Mitochondrial Inhibitor Atovaquone Increases Tumor Oxygenation and Inhibits Hypoxic Gene Expression in Patients with Non-Small Cell Lung Cancer. Clin. Cancer Res. 27, 2459–2469 (2021).

33. O'Connor, J. P. B., Robinson, S. P. & Waterton, J. C. Imaging tumour hypoxia with oxygen-enhanced MRI and BOLD MRI. Br. J. Radiol. 92, 20180642 (2019).

34. Hallac, R. R. et al. Correlations of noninvasive BOLD and TOLD MRI with pO2 and relevance to tumor radiation response. Magn. Reson. Med. 71, 1863 (2014).

35. Dence, C. S., Ponde, D. E., Welch, M. J. & Lewis, J. S. Autoradiographic and small-animal PET comparisons between 18F-FMISO, 18F-FDG, 18F-FLT and the hypoxic selective 64Cu-ATSM in a rodent model of cancer. Nucl. Med. Biol. 35, 713–720 (2008).

36. Wohlleben, G. et al. Influence of hypoxia and irradiation on osteopontin expression in head and neck cancer and glioblastoma cell lines. Radiat. Oncol. 10, (2015).

37. Hui, E. P. et al. Plasma Osteopontin, Hypoxia, and Response to Radiotherapy in Nasopharyngeal Cancer. Clin. Cancer Res. 14, 7080–7087 (2008).

38. Bache, M. et al. Elevated tumor and serum levels of the hypoxia-associated protein osteopontin are associated with prognosis for soft tissue sarcoma patients. BMC Cancer 10, 132 (2010).

39. Le, Q.-T. et al. Prognostic and Predictive Significance of Plasma HGF and IL8 in a Phase III trial of Chemoradiation with or without tirapazamine in locoregionally advanced head and neck cancer. Clin. Cancer Res. 18, 1798 (2012).

40. Shannon, A. M. et al. The mitogen-activated protein/extracellular signal-regulated kinase kinase 1/2 inhibitor AZD6244 (ARRY-142886) enhances the radiation responsiveness of lung and colorectal tumor xenografts. Clin. Cancer Res. 15, 6619–6629 (2009).

41. Winter, S. C. et al. Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. Cancer Res. 67, (2007).

42. 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, (2011).

43. 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, (2013).

44. Ringnér, M. What is principal component analysis? Nat. Biotechnol. 26, 303–304 (2008).

45. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 14, 7 (2013).

46. Bhandari, V. et al. Molecular landmarks of tumor hypoxia across cancer types. Nature Genetics 51, 308–318 (2019).

47. Yang, L. et al. A Gene Signature for Selecting Benefit from Hypoxia Modification of Radiotherapy for High-Risk Bladder Cancer Patients. Clin. Cancer Res. 23, 4761–4768 (2017).

48. Koong, A. C. et al. Candidate genes for the hypoxic tumor phenotype. Cancer Res. 60, 883–887 (2000).

49. Denko, N. C. et al. Investigating hypoxic tumor physiology through gene expression patterns. Oncogene 22, 5907–5914 (2003).

50. Jögi, A. et al. Human neuroblastoma cells exposed to hypoxia: induction of genes associated with growth, survival, and aggressive behavior. Exp. Cell Res. 295, 469–487 (2004).

51. Ning, W., Chu, T. J., Li, C. J., Choi, A. M. K. & Peters, D. G. Genome-wide analysis of the endothelial transcriptome under short-term chronic hypoxia. Physiol. Genomics 18, 70–78 (2004).

52. Manalo, D. J. et al. Transcriptional regulation of vascular endothelial cell responses to hypoxia by HIF-1. Blood 105, 659–669 (2005).

53. Wang, V., Davis, D. A., Haque, M., Huang, L. E. & Yarchoan, R. Differential gene up-regulation by hypoxia-inducible factor-1alpha and hypoxia-inducible factor-2alpha in HEK293T cells. Cancer Res. 65, 3299–3306 (2005).

54. Detwiller, K. Y. et al. Analysis of hypoxia-related gene expression in sarcomas and effect of hypoxia on RNA interference of vascular endothelial cell growth factor A. Cancer Res. 65, 5881–5889 (2005).

55. Bosco, M. C. et al. Hypoxia modifies the transcriptome of primary human monocytes: modulation of novel immune-related genes and identification of CC-chemokine ligand 20 as a new hypoxia-inducible gene. J. Immunol. 177, 1941–1955 (2006).

56. Mense, S. M. et al. Gene expression profiling reveals the profound upregulation of hypoxia-responsive genes in primary human astrocytes. Physiol. Genomics 25, 435–449 (2006).

57. 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, 5641–5647 (2006).

58. 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).

59. Elvidge, G. P. et al. Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition: the role of HIF-1alpha, HIF-2alpha, and other pathways. J. Biol. Chem. 281, 15215–15226 (2006).

60. Peters, D. G., Ning, W., Chu, T. J., Li, C. J. & Choi, A. M. K. Comparative SAGE analysis of the response to hypoxia in human pulmonary and aortic endothelial cells. Physiol. Genomics 26, 99–108 (2006).

61. Seigneuric, R. et al. Impact of supervised gene signatures of early hypoxia on patient survival. Radiother. Oncol. 83, 374–382 (2007).

62. Shi, Y.-F. et al. Hypoxia induces the activation of human hepatic stellate cells LX-2 through TGF-beta signaling pathway. FEBS Lett. 581, 203–210 (2007).

63. 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).

64. Beyer, S., Kristensen, M. M., Jensen, K. S., Johansen, J. V. & Staller, P. The histone demethylases JMJD1A and JMJD2B are transcriptional targets of hypoxia-inducible factor HIF. J. Biol. Chem. 283, 36542–36552 (2008).

65. van Malenstein, H. et al. A seven-gene set associated with chronic hypoxia of prognostic importance in hepatocellular carcinoma. Clin. Cancer Res. 16, 4278–4288 (2010).

66. 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).

67. Fardin, P. et al. The l1-l2 regularization framework unmasks the hypoxia signature hidden in the transcriptome of a set of heterogeneous neuroblastoma cell lines. BMC Genomics 10, 474 (2009).

68. Hu, Z. et al. A compact VEGF signature associated with distant metastases and poor outcomes. BMC Med. 7, 9 (2009).

69. 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).

70. 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. 49, 895–905 (2010).

71. 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).

72. Ghorbel, M. T. et al. Transcriptomic analysis of patients with tetralogy of Fallot reveals the effect of chronic hypoxia on myocardial gene expression. J. Thorac. Cardiovasc. Surg. 140, 337–345.e26 (2010).

73. Ghazoui, Z. et al. Close and Stable Relationship between Proliferation and a Hypoxia Metagene in Aromatase Inhibitor–Treated ER-Positive Breast Cancer. Clin. Cancer Res. 17, 3005–3012 (2011).

74. Starmans, M. H. W. et al. The prognostic value of temporal in vitro and in vivo derived hypoxia gene-expression signatures in breast cancer. Radiother. Oncol. 102, 436–443 (2012).

75. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191 (2009).

76. Multi-symbol checker. https://www.genenames.org/tools/multi-symbol-checker/.

77. Website. GeneCards – the human gene database www.genecards.org Stelzer G, Rosen R, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Iny Stein T, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary.

78. Zyla, J., Marczyk, M., Weiner, J. & Polanska, J. Ranking metrics in gene set enrichment analysis: do they matter? BMC Bioinformatics 18, 256 (2017).

79. Foroutan, M. et al. Single sample scoring of molecular phenotypes. BMC Bioinformatics 19, 404 (2018).

80. Tomfohr, J., Lu, J. & Kepler, T. B. Pathway level analysis of gene expression using singular value decomposition. BMC Bioinformatics 6, 225 (2005).

81. Dhawan, A. et al. Guidelines for using sigQC for systematic evaluation of gene signatures. Nat. Protoc. 14, 1377–1400 (2019).

82. Clough, E. & Barrett, T. The Gene Expression Omnibus Database. Methods Mol. Biol. 1418, 93–110 (2016).

83. Weinstein, J. N. et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. 45, 1113–1120 (2013).

84. Venet, D., Dumont, J. E. & Detours, V. Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput. Biol. 7, e1002240 (2011).

85. Aran, D. et al. Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat. Commun. 8, (2017).

86. Jing, X. et al. Role of hypoxia in cancer therapy by regulating the tumor microenvironment. Molecular Cancer 18, (2019).

87. Walsh, J. C. et al. The clinical importance of assessing tumor hypoxia: relationship of tumor hypoxia to prognosis and therapeutic opportunities. Antioxid. Redox Signal. 21, 1516–1554 (2014).

88. Muz, B., de la Puente, P., Azab, F. & Azab, A. K. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. Hypoxia 83, (2015).

89. Semenza, G. L. Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics. Oncogene 29, 625–634 (2010).

90. Mo, Z. et al. Identification of a Hypoxia-Associated Signature for Lung Adenocarcinoma. Front. Genet. 11, (2020).

91. Holmes, D. I. R. & Zachary, I. The vascular endothelial growth factor (VEGF) family: angiogenic factors in health and disease. Genome Biol. 6, 209 (2005).

92. Geindreau, M., Ghiringhelli, F. & Bruchard, M. Vascular Endothelial Growth Factor, a Key Modulator of the Anti-Tumor Immune Response. Int. J. Mol. Sci. 22, (2021).

93. Ellen, T. P., Ke, Q., Zhang, P. & Costa, M. NDRG1, a growth and cancer related gene: regulation of gene expression and function in normal and disease states. Carcinogenesis 29, 2–8 (2008).

94. Li, A. et al. Upregulation of NDRG1 predicts poor outcome and facilitates disease progression by influencing the EMT process in bladder cancer. Sci. Rep. 9, 5166 (2019).

95. Kappler, M. et al. P4HA1: A single-gene surrogate of hypoxia signatures in oral squamous cell carcinoma patients. Clin. Transl. Radiat. Oncol. 5, 6–11 (2017).

96. Xiong, G. et al. Collagen prolyl 4-hydroxylase 1 is essential for HIF-1α stabilization and TNBC chemoresistance. Nat. Commun. 9, 4456 (2018).

97. Tirado-Hurtado, I., Fajardo, W. & Pinto, J. A. DNA Damage Inducible Transcript 4 Gene: The Switch of the Metabolism as Potential Target in Cancer. Front. Oncol. 8, (2018).

98. Scandurro, A. B., Weldon, C. W., Figueroa, Y. G., Alam, J. & Beckman, B. S. Gene microarray analysis reveals a novel hypoxia signal transduction pathway in human hepatocellular carcinoma cells. Int. J. Oncol. 19, 129–135 (2001).

99. Wykoff, C. C., Pugh, C. W., Maxwell, P. H., Harris, A. L. & Ratcliffe, P. J. Identification of novel hypoxia dependent and independent target genes of the von Hippel-Lindau (VHL) tumour suppressor by mRNA differential expression profiling. Oncogene 19, (2000).

100. Denko, N. et al. Epigenetic Regulation of Gene Expression in Cervical Cancer Cells by the Tumor Microenvironment. Clin. Cancer Res. 6, 480–487 (2000).

101. Lendahl, U., Lee, K. L., Yang, H. & Poellinger, L. Generating specificity and diversity in the transcriptional response to hypoxia. Nat. Rev. Genet. 10, (2009).

102. Halle, C. et al. Hypoxia-induced gene expression in chemoradioresistant cervical cancer revealed by dynamic contrast-enhanced MRI. Cancer Res. 72, 5285–5295 (2012).

103. Boidot, R. et al. A generic cycling hypoxia-derived prognostic gene signature: application to breast cancer profiling. Oncotarget 5, 6947 (2014).

104. 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).

105. 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, 4067–4076 (2016).

106. Suh, Y.-E. et al. Association between hypoxic volume and underlying hypoxia-induced gene expression in oropharyngeal squamous cell carcinoma. Br. J. Cancer 116, 1057–1064 (2017).

107. Ye, I. C. et al. Molecular Portrait of Hypoxia in Breast Cancer: A Prognostic Signature and Novel HIF-Regulated Genes. Mol. Cancer Res. 16, 1889–1901 (2018).

108. Yang, L. et al. Development and Validation of a 28-gene Hypoxia-related Prognostic Signature for Localized Prostate Cancer. EBioMedicine 31, (2018).

109. Yang, L. et al. Validation of a hypoxia related gene signature in multiple soft tissue sarcoma cohorts. Oncotarget 9, (2017).

110. Dao, T. P. et al. Identification of a Prognostic Hypoxia-Associated Gene Set in IDH-Mutant Glioma. Int. J. Mol. Sci. 19, (2018).

111. Chen, Y. L. et al. A 17 gene panel for non-small-cell lung cancer prognosis identified through integrative epigenomic-transcriptomic analyses of hypoxia-induced epithelial-mesenchymal transition. Mol. Oncol. 13, (2019).

112. Zou, Y. F. et al. A signature of hypoxia-related factors reveals functional dysregulation and robustly predicts clinical outcomes in stage I/II colorectal cancer patients. Cancer Cell Int. 19, (2019).

113. Zhang, B. et al. A hypoxia-related signature for clinically predicting diagnosis, prognosis and immune microenvironment of hepatocellular carcinoma patients. J. Transl. Med. 18, 342 (2020).

114. Wang, J. et al. Development and validation of a hypoxia-related prognostic signature for breast cancer. Oncol. Lett. 20, (2020).

115. Shou, Y. et al. Identification of Signatures of Prognosis Prediction for Melanoma Using a Hypoxia Score. Front. Genet. 11, (2020).

116. Lin, W. et al. Characterization of Hypoxia Signature to Evaluate the Tumor Immune Microenvironment and Predict Prognosis in Glioma Groups. Front. Oncol. 10, (2020).

117. Sun, J. et al. Development and validation of a hypoxia-related gene signature to predict overall survival in early-stage lung adenocarcinoma patients. Ther. Adv. Med. Oncol. 12, (2020).

118. Calvo, T. M. et al. An Experimentally Defined Hypoxia Gene Signature in Glioblastoma and Its Modulation by Metformin. Biology 9, (2020).

119. Kim, S.-H. et al. HIG2 promotes colorectal cancer progression via hypoxia-dependent and independent pathways. Cancer Lett. 341, 159–165 (2013).

120. Liu, C., Zhou, X., Zeng, H., Wu, D. & Liu, L. HILPDA Is a Prognostic Biomarker and Correlates With Macrophage Infiltration in Pan-Cancer. Front. Oncol. 0, (2021).

121. van Dierendonck, X. A. M. H. et al. HILPDA Uncouples Lipid Droplet Accumulation in Adipose Tissue Macrophages from Inflammation and Metabolic Dysregulation. Cell Rep. 30, 1811–1822.e6 (2020).

122. Govindarajan, R., Duraiyan, J., Kaliyappan, K. & Palanisamy, M. Microarray and its applications. J. Pharm. Bioallied Sci. 4, S310 (2012).

123. Yi, M., Nissley, D. V., McCormick, F. & Stephens, R. M. ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. Sci. Rep. 10, 1–16 (2020).

124. Wu, X. et al. Regulation of cellular sterol homeostasis by the oxygen responsive noncoding RNA lincNORS. Nat. Commun. 11, 4755 (2020).

125. Leithner, K. et al. Hypoxia increases membrane metallo-endopeptidase expression in a novel lung cancer ex vivo model - role of tumor stroma cells. BMC Cancer 14, (2014).

126. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. Br. J. Cancer 83, (2000).

127. McClain, M. R., Palomaki, G. E., Nathanson, K. L. & Haddow, J. E. Adjusting the estimated proportion of breast cancer cases associated with BRCA1 and BRCA2 mutations: public health implications. Genet. Med. 7, 28–33 (2005).

128. Hartge, P., Struewing, J. P., Wacholder, S., Brody, L. C. & Tucker, M. A. The prevalence of common BRCA1 and BRCA2 mutations among Ashkenazi Jews. Am. J. Hum. Genet. 64, (1999).

129. Roa, B. B., Boyd, A. A., Volcik, K. & Richards, C. S. Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. Nat. Genet. 14, 185–187 (1996).

130. Abul-Husn, N. S. et al. Exome sequencing reveals a high prevalence of BRCA1 and BRCA2 founder variants in a diverse population-based biobank. Genome Med. 12, (2019).

131. Ferrone, C. R. et al. BRCA germline mutations in Jewish patients with pancreatic adenocarcinoma. J. Clin. Oncol. 27, 433–438 (2009).

132. van Asperen, C. J. et al. Cancer risks in BRCA2 families: estimates for sites other than breast and ovary. J. Med. Genet. 42, (2005).

133. Mallipatna, A., Marino, M. & Singh, A. D. Genetics of Retinoblastoma. Asia Pac J Ophthalmol (Phila) 5, 260–264 (2016).

134. Dyson, N. J. RB1: a prototype tumor suppressor and an enigma. Genes Dev. 30, 1492–1502 (2016).

135. Website: What is LFS? Li-Fraumeni Syndrome Association http://www.lfsassociation.org/what-is-lfs/ (2012).

136. Stacey, S. N. et al. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. Nat. Genet. 43, (2011).

137. Di Giovannantonio, M. et al. Heritable genetic variants in key cancer genes link cancer risk with anthropometric traits. J. Med. Genet. 58, (2021).

138. Wang, P. Y. et al. Increased oxidative metabolism in the Li-Fraumeni syndrome. N. Engl. J. Med. 368, (2013).

139. Bossé, Y. & Amos, C. I. A Decade of GWAS Results in Lung Cancer. Cancer Epidemiol. Biomarkers Prev. 27, 363–379 (2018).

140. Fachal, L. & Dunning, A. M. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. Curr Opin Solid State Mater Sci 30, 32–41 (2015).

141. Thorgeirsson, T. E. et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 452, 638–642 (2008).

142. Brücher, B. L. D. M., Brücher, B. L. D. & Jamall, I. S. Somatic Mutation Theory - Why it's Wrong for Most Cancers. Cell. Physiol. Biochem. 38, 1663–1680 (2016).

143. Greenman, C. et al. Patterns of somatic mutation in human cancer genomes. Nature 446, (2007).

144. Forbes, S. A. et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 43, D805–D811 (2014).

145. Nigro, J. M. et al. Mutations in the p53 gene occur in diverse human tumour types. Nature 342, 705–708 (1989).

146. Ozaki, T. & Nakagawara, A. Role of p53 in Cell Death and Human Cancers. Cancers 3, 994–1013 (2011).

147. Levine, A. J. p53, the cellular gatekeeper for growth and division. Cell 88, 323–331 (1997).

148. Levine, A. J. & Oren, M. The first 30 years of p53: growing ever more complex. Nat. Rev. Cancer 9, 749–758 (2009).

149. Hock, A. K. & Vousden, K. H. The role of ubiquitin modification in the regulation of p53. Biochim Biophys Acta Mol Cell Res 1843, 137–149 (2014).

150. Cheng, Q., Chen, L., Li, Z., Lane, W. S. & Chen, J. ATM activates p53 by regulating MDM2 oligomerization and E3 processivity. EMBO J. 28, 3857–3867 (2009).

151. Bhandari, V., Li, C. H., Bristow, R. G. & Boutros, P. C. Divergent mutational processes distinguish hypoxic and normoxic tumours. Nat. Commun. 11, 1–10 (2020).

152. Brodaczewska, K. K., Szczylik, C., Fiedorowicz, M., Porta, C. & Czarnecka, A. M. Choosing the right cell line for renal cell cancer research. Mol. Cancer 15, (2016).

153. Ashida, S., Nishimori, I., Tanimura, M., Onishi, S. & Shuin, T. Effects of von Hippel-Lindau gene mutation and methylation status on expression of transmembrane

carbonic anhydrases in renal cell carcinoma. J. Cancer Res. Clin. Oncol. 128, 561–568 (2002).

154. Leisz, S. et al. Distinct von Hippel-Lindau gene and hypoxia-regulated alterations in gene and protein expression patterns of renal cell carcinoma and their effects on metabolism. Oncotarget 6, (2015).

155. Chen, Y. et al. ZMYND8 acetylation mediates HIF-dependent breast cancer progression and metastasis. J. Clin. Invest. 128, (2018).

156. Rana, N. K., Singh, P. & Koch, B. CoCl2 simulated hypoxia induce cell proliferation and alter the expression pattern of hypoxia associated genes involved in angiogenesis and apoptosis. Biol. Res. 52, 1–13 (2019).

157. Ayrapetov, M. K. et al. Activation of Hif1α by the prolylhydroxylase inhibitor dimethyoxalyglycine decreases radiosensitivity. PLoS One 6, (2011).

158. Sethuraman, A. et al. BHLHE40 confers a pro-survival and pro-metastatic phenotype to breast cancer cells by modulating HBEGF secretion. Breast Cancer Res. 20, 1–17 (2018).

159. Sørensen, B. S. et al. Influence of oxygen concentration and pH on expression of hypoxia induced genes. Radiother. Oncol. 76, 187–193 (2005).

160. Sørensen, B. S., Alsner, J., Overgaard, J. & Horsman, M. R. Hypoxia induced expression of endogenous markers in vitro is highly influenced by pH. Radiother. Oncol. 83, 362–366 (2007).