RadiationGeneSigDB: A comprehensive database of oxic and hypoxic radiation response gene expression signatures

Venkata SK. Manem[1*], Andrew Dhawan[2]

[1]Institut Universitaire de Cardiologie et de Pneumologie de Québec, Université Laval, Quebec City, QC, Canada.
[2]Neurological Institute, Cleveland Clinic, Cleveland, OH, USA

***Corresponding Author:**
Venkata SK Manem
Research Professional
mail2mvskumar@gmail.com

Abstract

Radiation therapy is among the most effective and widely used modalities of cancer therapy in current clinical practice. With the advent of new high throughput genomic technologies and the continuous inflow of transcriptomic data, there has been a paradigm shift in the landscape of radiation oncology. In this era of personalized radiation medicine, genomic datasets hold great promise to investigate novel biomarkers predictive of radiation response. These biomarkers may be used to investigate radiation-induced biological events, and to predict clinical outcomes. In this regard, the number of available gene expression based signatures built under oxic and hypoxic conditions is getting larger. This poses two main questions in the field, namely, *i)* how reliable are these signatures when applied across a compendium of datasets in different model systems; and *ii)* is there redundancy of gene signatures. To address these fundamental radiobiologic questions, we curated a database of gene expression signatures predictive of radiation response under oxic and hypoxic conditions. *RadiationGeneSigDB* has a collection of 11 oxic and 24 hypoxic signatures with the standardized gene list as a gene symbol, Entrez gene id along with its function. We present the utility of this database through two case studies: *i)* comparing breast cancer oxic signatures in cell line data *vs.* patient data; and *ii)* comparing the similarity of head and neck cancer hypoxia signatures in clinical tumor data. This valuable, curated repertoire of published gene expression signatures provides a motivating example for how to search for similarities in radiation response for tumors arising from different tissues across model systems under oxic and hypoxic conditions.

Availability and implementation: *RadiationGeneSigDB* is implemented in R. The source code of this package and signatures can be downloaded from GitHub: https://github.com/vmsatya/RadiationGeneSigDB.

1 INTRODUCTION

The aim of ionizing radiation is to eliminate, with as much certainty as possible, all tumor cells in the primary tumor along with the regional lymph nodes, while at the same time limiting damage to the surrounding non-cancerous tissues. Over the last few decades, improvements to radiation therapy delivery have substantially increased the survival rate of cancer patients (Bernier *et al.*, 2004). Many of these advances have been due to technological improvements in radiation delivery and dose conformity, which include particle therapies coupled with advanced image guidance techniques (Baumann *et al.*, 2016; Verellen *et al.*, 2007). These advanced techniques are more precise in targeting the primary tumor volume with relatively lower radiation doses to surrounding healthy tissues, resulting in improved patient survival (Tree *et al.*, 2013). Moreover,, the decision-making underlying the stratification of patients into different chemo-radiotherapeutic protocols to date has been based on clinical variables such as tumor stage, histology, age of the patient, gender, etc. derived from interrogation of large population cohorts. However, it is well-supported from several studies that intra-tumor heterogeneity and inter-patient variability are key features that can dictate the tumor response to a certain kind of treatment regimen, be it radiation therapy or chemotherapy, and importantly, this may affect the results and interpretation of these cohort-based studies (Bentzen and Overgaard, 1994; Kozin *et al.*, 2008; Krause *et al.*, 2009).

The two pillars driving the field of personalized radiation oncology are: *i)* treatment delivery and dose conformity arising from technological improvements, which include particle therapies and advanced image guidance techniques; *ii)* novel biomarker-guided tools, integrating concomitant chemotherapy (Baumann *et al.*, 2016). To tailor radiation therapy, it is crucial to build predictive assays that are more confidently able to stratify patients, and concomitantly have associated impactful radiotherapeutic regimens (Peters, 1990). This could augment the existing radiobiological treatment strategies to more biologically-driven personalized radiation treatment to individual patients. The evolution of new high throughput technologies and the continuous inflow of transcriptomic data have created new avenues to understand complex biological events induced by ionizing radiation, through data driven analysis, at a level beyond the gross clinical variables of an individual, and instead, at the individual tumor level. The ever-expanding arsenal of transcriptomic data holds great promise to investigate novel biomarkers that are predictive of radiation response. In the literature, several studies have attempted to associate radiosensitivity with molecular/genomic features (Scott *et al.*, 2017), and this number is getting larger. But, there have been no systematic efforts to build a database of *radiation response* gene expression signatures validated or designed for clinical use.

In the literature, many groups conducted comprehensive genome-wide assessments of gene expression profiling and built gene expression based signatures for radiation response under oxic and hypoxic conditions. Broadly, two methods have been used to identify these gene signatures, namely, data-driven (bottom-up) and hypothesis-based (top-down) approaches. The performances of these transcriptomic signatures have been evaluated on various datasets with limited to no independent validation. Moreover, there is only minimal overlap between the genes of these different signatures. This may be attributed to the different transcriptomic platforms,

such as microarray or RNA-sequencing, training sets, and statistical tools used to build these signatures. Moreover, in order to develop biomarkers reproducible and appropriate for clinical translation, a database should be built to house these predictive models.

At present, there is no radiation response signature database that could potentially address these fundamental questions. A wealth of molecular data (Curtis *et al.*, 2012; Cancer Genome Atlas Research Network *et al.*, 2013) and gene expression based signatures (Subramanian *et al.*, 2005) are publicly available through diverse online resources. In this study, we manually curated a number of radiation response gene signatures from the literature, known as the *RadiationGeneSigDB*, and implemented this as an *R* package. This database will facilitate users, *i)* to compare radiation response signatures across pan-cancer datasets; *ii)* to investigate the prognostic value of these signatures on a compendium of clinical datasets using meta-analysis; and, *iii)* to investigate the tissue specificity of radiation response using these signatures across different transcriptomic platforms. Ultimately, the goal of this work is to improve the development, and spur a greater understanding of how transcriptomic signatures can be used to augment precision radiation oncology, for improved patient outcomes, and more effectively designed clinical trials.

2 METHODS

2.1 Curation of RadiationGeneSigDB

In most online resources, gene signatures are often included in tables or figures embedded in publications. These signatures often use non-standard gene identifiers, making comparison to other gene signatures, or even to the original data a significant challenge. To be of maximal value, these gene signatures should be available through an easily accessible database resource that provides gene sets in a more standard format. Radiation response gene expression based signatures are identified in Pubmed. In our first release, we collected 11, 24 oxic and hypoxic signatures respectively, and these are described in Supplementary Tables 1 and 2. We manually annotated them with gene symbol, Entrez Gene ID, and gene function.

2.2 Datasets

Cancer cell lines were profiled at the genomic level and the processed data are available for download from a public database, The Cancer Cell Line Encyclopedia (CCLE) (Barretina *et al.*, 2012). For this study, we used all breast cancer lines (26 in total) from the CCLE dataset. The METABRIC dataset was used for patient data (Curtis *et al.*, 2012) and selected only those patients treated with radiation therapy. Out of 1992 patients, 232 patients were reportedly treated with radiation. We retrieved head and neck squamous cell carcinoma (HNSCC) primary tumor transcriptomic data from the Tumor Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research Network *et al.*, 2013), and selected only the patients treated with radiation therapy (99 in total).

2.3 Molecular subtyping and signature score

We used the SCMOD2 model (Haibe-Kains *et al.*, 2012) to assign each tumor sample into the four established molecular subtypes of breast cancer: Basal-like (TNBC), Her2-enriched,

Luminal A (ER-high) and Luminal B (ER-low). We used the SCMOD2 implementation available in the *genefu* package (Gendoo *et al.*, 2016). We computed a signed average (the sign being determined by the sign of the gene coefficient or direction) using the sig.score function in *genefu*. We computed the correlation between the the published gene signatures using the Spearman correlation coefficient.

3 CASE STUDIES

We present here two case studies utilizing the radiation response gene signature database, highlighting its utility, and the associated *R* code.

3.1 Comparison of *oxic* breast gene signatures in cell lines and patients

The objective of this case study is to conduct an unbiased comparison of two different prognostic breast cancer signatures (namely, Piening et al. (Piening *et al.*, 2009), Speers et al. (Speers *et al.*, 2015)) predictive of radiation response under oxic conditions. We compared the signature scores in cell line data as well as in patients, and assessed the similarities and differences in their behavior across these datasets. Importantly, the overlap of genes between these two signatures is minimal, attributable to the different platforms, training sets, and statistical methods used in their generation. We computed signature scores across the 26 breast cancer cell lines from the the CCLE RNA-seq database, as well as for all patient samples from the METABRIC dataset. Before proceeding further with this analysis, we first assayed the quality of gene signature application on each of these datasets using *sigQC,* as developed in Dhawan et al (Dhawan *et al.*, 2017). Interestingly, this precursor analysis, with summary radar plots in Supplementary Figure 1, showed that there are significant variations in quality between the signatures on different datasets. That is, the signatures appeared to have strongest quality when applied on CCLE data or TNBC data, but not the ER-low subtype. Next, the signature scores for both the Piening and Speers signatures were then compared directly with the Spearman correlation coefficient (Figure 1). We observed a moderate Spearman correlation (ρ ~0.69) between the signature scores in the cell line data. However, we observed significant differences when comparing signature scores in patient data. We observed that for clinical data, when stratified by molecular subtype, the Spearman correlation has moderate value for triple negative breast cancer (TNBC), HER2-positive and ER-high cancers. For the ER-low subtype, the Spearman correlation between the signatures is weak (ρ~0.25). The moderate correlation between the two breast cancer signatures can be attributed to: *i)* different statistical tools implemented on different platforms, with a small overlap in genes; *ii)* different experimental assays and protocols used to generate dose-response data; *iii)* inconsistent experimental methodologies to generate data across labs; *iv)* signature development being underpowered due to lack of sample size along with enough replicates. This could also have implications with regard to the evaluation of radiation response in an *in-vitro* setting and translating it to the clinical practice. Hence, there is a dire need to build robust radiation signatures that are independent of platforms and technology.

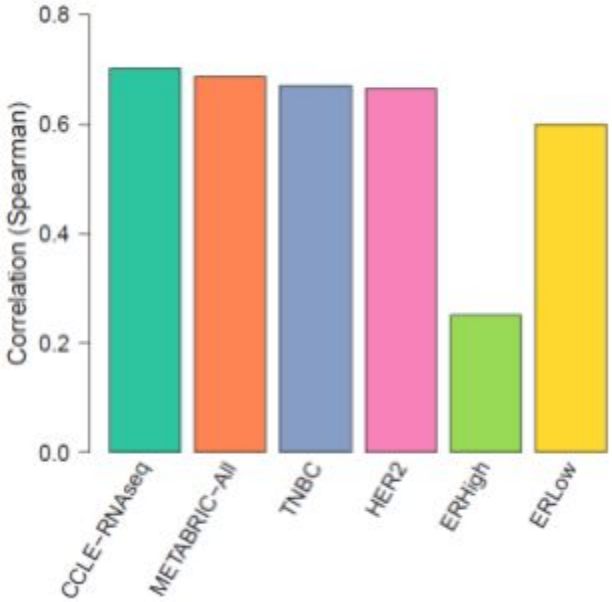


Figure 1: Cell line vs. Patient model systems. Comparison of breast cancer radiation response signature scores in CCLE breast cancer cell lines and in METABRIC patient cohort.

3.2 Comparison of HNSCC *hypoxia* gene signatures in patients

In this case study, we aimed to compare three HNSCC hypoxia gene expression signatures in the TCGA cohort. The first hypoxia signature is identified by Toustrup et al. (Toustrup *et al.*, 2011), which was validated in a HNSCC cohort. The second hypoxia signature is developed by Eustace et al. (A. Eustace *et al.*, 2013) on laryngeal tumors, which was validated in HNSCC cohort. Lastly, the third signature was built by Lendahl et al. (Lendahl *et al.*, 2009), which was derived to determine a pan-cancer signature for radiation response in hypoxia. Firstly, comparing the composition of each of these signatures, we noted that there is just a single common gene between the three signatures. Next, we performed a quality analysis, to ensure legitimacy when applying these signatures on the TCGA dataset with *sigQC* and noted that all three signatures show good quality on this dataset and are consistent between each other, as shown in the summary radar plots in Supplementary Figure 2. We found that the Toustrup and Lendahl, and the Toustrup and Eustace hypoxia signatures were strongly correlated (ρ=0.8), whereas the Lendahl and Eustace signatures were weakly correlated (ρ=0.46) (Figure 2-Left panel), which is consistent with (Tawk *et al.*, 2016). Having identified differences in their genetic composition, but relative similarity in the behavior, we next asked whether similar biological processes could be defining these signatures. To achieve this, we performed pathway analysis using the GO terms from the MSigDB (Subramanian *et al.*, 2005). For an FDR < 10%, 13, 37, and 51 transcriptional pathways were found to be enriched using Toustrup, Eustace, Lendahl signatures respectively. We found only 3 pathways that were commonly enriched between all the 3 signatures (Figure 2- Right panel), namely, GO:Oxidation_Reduction_Process", "GO:Glucose_Metabolic_Process," GO:Monosaccharide _Metabolic_Process. We found the pathways, "GO: Response_To_Oxygen_levels" and "GO:CellularResponse_Oxygen_Levels" to

be enriched related to the oxygen levels using Toustrup and Lendahl signatures, but not enriched using the Eustace signature, as summarized by the Venn diagram in Figure 2 (right).

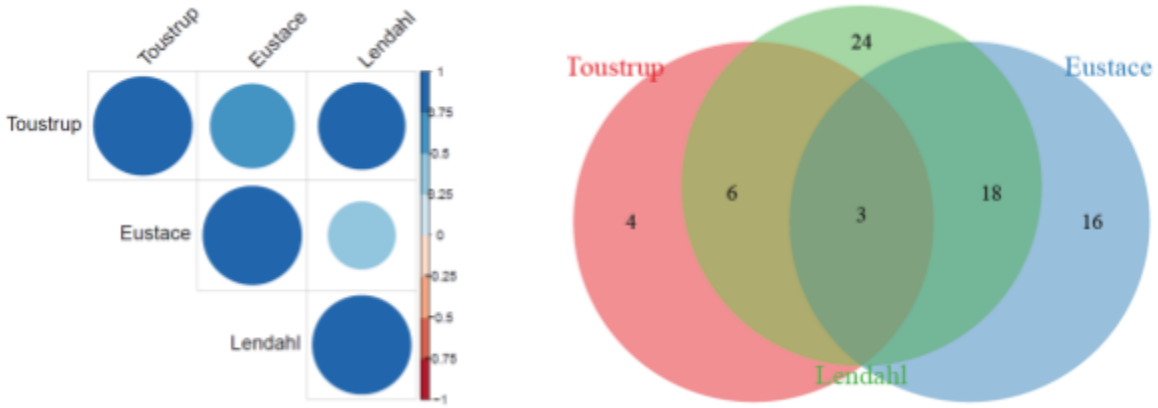


Figure 2: Comparison of hypoxia gene signatures in TCGA HNSCC. Left panel: Correlation of signature scores. **Right panel:** Venn diagram illustrating the transcriptional pathways enriched using the three hypoxia signatures (FDR<10%).

4. CONCLUSION

In this era of personalized medicine, an area of current excitement in the field of radiation oncology is the study of changes in the transcriptome induced by radiation therapy, termed radiogenomics. With the evolution of transcriptomic sequencing technologies, radiogenomics has emerged as a new field to help decode changes in the genetic events induced by radiation therapy, and also identify the biomarkers that are predictive of radiation response. Due to the continuous growth of genome-wide association studies and transcriptomic studies across multiple cancers, the number of biomarkers predictive of radiation response is ever-increasing. The lack of genomic indicators of radiation response has impeded the administration of radiation therapy to individual patients. Moreover, if a predictive biomarker were to be integrated into the clinical practice, it would require a reproducible assay that is agnostic to platforms and technologies. This study is an effort to build a database of radiation response gene signatures under *oxic* and *hypoxic* conditions. This repertoire of gene expression based signatures are publicly available as an open source package in *R*. This will facilitate comparison of radiation response signatures across cancer types, and also enable us to investigate the prognostic value of these signatures using meta-analysis approaches. We hope and envision that this package will help users to compare their own signature to those in the *RadiationGeneSigDB* database, and help build better biomarkers by using multiple datasets in the discovery, or pre-clinical phase. Furthermore, *RadiationGeneSigDB* coupled with *RadioGx, a* novel computational platform of radiogenomics datasets (Manem *et al.*, 2018), will enable us to build reliable and clinically-verifiable genomic predictors of radiation response.

REFERENCES

Abazeed,M.E. *et al.* (2013) Integrative radiogenomic profiling of squamous cell lung cancer. *Cancer Res.*, **73**, 6289–6298.

Amundson,S.A. *et al.* (2008) Integrating global gene expression and radiation survival

parameters across the 60 cell lines of the National Cancer Institute Anticancer Drug Screen. *Cancer Res.*, **68**, 415–424.
Barretina,J. *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
Baumann,M. *et al.* (2016) Radiation oncology in the era of precision medicine. *Nat. Rev. Cancer*, **16**, 234–249.
Bentzen,S.M. and Overgaard,J. (1994) Patient-to-Patient Variability in the Expression of Radiation-Induced Normal Tissue Injury. *Semin. Radiat. Oncol.*, **4**, 68–80.
Bernier,J. *et al.* (2004) Radiation oncology: a century of achievements. *Nat. Rev. Cancer*, **4**, 737–747.
Buffa,F.M. *et al.* (2010) Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br. J. Cancer*, **102**, 428–435.
Cancer Genome Atlas Research Network *et al.* (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
Curtis,C. *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**, 346–352.
Dhawan,A. *et al.* (2017) sigQC: A procedural approach for standardising the evaluation of gene signatures. *bioRxiv*, 203729.
Eustace,A. *et al.* (2013) A 26-Gene Hypoxia Signature Predicts Benefit from Hypoxia-Modifying Therapy in Laryngeal Cancer but Not Bladder Cancer. *Clin. Cancer Res.*, **19**, 4879–4888.
Eustace,A. *et al.* (2013) A 26-gene hypoxia signature predicts benefit from hypoxia-modifying therapy in laryngeal cancer but not bladder cancer. *Clin. Cancer Res.*, **19**, 4879–4888.
Fjeldbo,C.S. *et al.* (2016) 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.
Gendoo,D.M.A. *et al.* (2016) Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics*, **32**, 1097–1099.
Ghazoui,Z. *et al.* (2011) Close and Stable Relationship between Proliferation and a Hypoxia Metagene in Aromatase Inhibitor-Treated ER-Positive Breast Cancer. *Clin. Cancer Res.*, **17**, 3005–3012.
Haibe-Kains,B. *et al.* (2012) A three-gene model to robustly identify breast cancer molecular subtypes. *J. Natl. Cancer Inst.*, **104**, 311–325.
Hu,Z. *et al.* (2009) A compact VEGF signature associated with distant metastases and poor outcomes. *BMC Med.*, **7**, 9.
de Jong,M.C. *et al.* (2015) Pretreatment microRNA Expression Impacting on Epithelial-to-Mesenchymal Transition Predicts Intrinsic Radiosensitivity in Head and Neck Cancer Cell Lines and Patients. *Clin. Cancer Res.*, **21**, 5630–5638.
Kim,H.S. *et al.* (2012) Identification of a radiosensitivity signature using integrative metaanalysis of published microarray data for NCI-60 cancer cells. *BMC Genomics*, **13**, 348.
Kozin,S.V. *et al.* (2008) Inter- and intramouse heterogeneity of radiation response for a growing paired organ. *Radiat. Res.*, **170**, 264–267.
Krause,M. *et al.* (2009) Heterogeneity of tumour response to combined radiotherapy and EGFR inhibitors: differences between antibodies and TK inhibitors. *Int. J. Radiat. Biol.*, **85**, 943–954.
Lendahl,U. *et al.* (2009) Generating specificity and diversity in the transcriptional response to hypoxia. *Nat. Rev. Genet.*, **10**, 821–832.
van Malenstein,H. *et al.* (2010) A seven-gene set associated with chronic hypoxia of prognostic importance in hepatocellular carcinoma. *Clin. Cancer Res.*, **16**, 4278–4288.

Manem,V.S.K. *et al.* (2018) Modeling cellular response in large-scale radiogenomic databases to advance precision radiotherapy.
Nuyten,D.S.A. *et al.* (2006) Predicting a local recurrence after breast-conserving therapy by gene expression profiling. *Breast Cancer Res.*, **8**, R62.
Peters,L.J. (1990) The ESTRO Regaud lecture. Inherent radiosensitivity of tumor and normal tissue cells as a predictor of human tumor response. *Radiother. Oncol.*, **17**, 177–190.
Piening,B.D. *et al.* (2009) A radiation-derived gene expression signature predicts clinical outcome for breast cancer patients. *Radiat. Res.*, **171**, 141–154.
Pitroda,S.P. *et al.* (2014) DNA repair pathway gene expression score correlates with repair proficiency and tumor sensitivity to chemotherapy. *Sci. Transl. Med.*, **6**, 229ra42.
Ragnum,H.B. *et al.* (2015) The tumour hypoxia marker pimonidazole reflects a transcriptional programme associated with aggressive prostate cancer. *Br. J. Cancer*, **112**, 382–390.
Scott,J.G. *et al.* (2017) A genome-based model for adjusting radiotherapy dose (GARD): a retrospective, cohort-based study. *Lancet Oncol.*, **18**, 202–211.
Seigneuric,R. *et al.* (2007) Impact of supervised gene signatures of early hypoxia on patient survival. *Radiother. Oncol.*, **83**, 374–382.
Sørensen,B.S. *et al.* (2010) Identifying pH independent hypoxia induced genes in human squamous cell carcinomas in vitro. *Acta Oncol.*, **49**, 895–905.
Speers,C. *et al.* (2015) Development and Validation of a Novel Radiosensitivity Signature in Human Breast Cancer. *Clin. Cancer Res.*, **21**, 3667–3677.
Starmans,M.H.W. *et al.* (2012) The prognostic value of temporal in vitro and in vivo derived hypoxia gene-expression signatures in breast cancer. *Radiother. Oncol.*, **102**, 436–443.
Subramanian,A. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 15545–15550.
Tang,Z. *et al.* (2017) Predicting radiotherapy response for patients with soft tissue sarcoma by developing a molecular signature. *Oncol. Rep.*, **38**, 2814–2824.
Tawk,B. *et al.* (2016) Comparative analysis of transcriptomics based hypoxia signatures in head- and neck squamous cell carcinoma. *Radiother. Oncol.*, **118**, 350–358.
Torres-Roca,J.F. *et al.* (2005) Prediction of radiation sensitivity using a gene expression classifier. *Cancer Res.*, **65**, 7169–7176.
Toustrup,K. *et al.* (2011) Development of a hypoxia gene expression classifier with predictive impact for hypoxic modification of radiotherapy in head and neck cancer. *Cancer Res.*, **71**, 5923–5931.
Tree,A.C. *et al.* (2013) Stereotactic body radiotherapy for oligometastases. *Lancet Oncol.*, **14**, e28–37.
Verellen,D. *et al.* (2007) Innovations in image-guided radiotherapy. *Nat. Rev. Cancer*, 71–71.
Weichselbaum,R.R. *et al.* (2008) An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 18490–18495.
Winter,S.C. *et al.* (2007) Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res.*, **67**, 3441–3449.
Yang,L. *et al.* (2017) A Gene Signature for Selecting Benefit from Hypoxia Modification of Radiotherapy for High-Risk Bladder Cancer Patients. *Clin. Cancer Res.*, **23**, 4761–4768.
Yang,L. *et al.* (2018) Development and Validation of a 28-gene Hypoxia-related Prognostic Signature for Localized Prostate Cancer. *EBioMedicine*, **31**, 182–189.

Supplementary Material

Supplementary Table 1 presents the original study, number of genes in the signature and the histology for the radiation response gene expression signatures under **oxic conditions** (fully oxygenated).

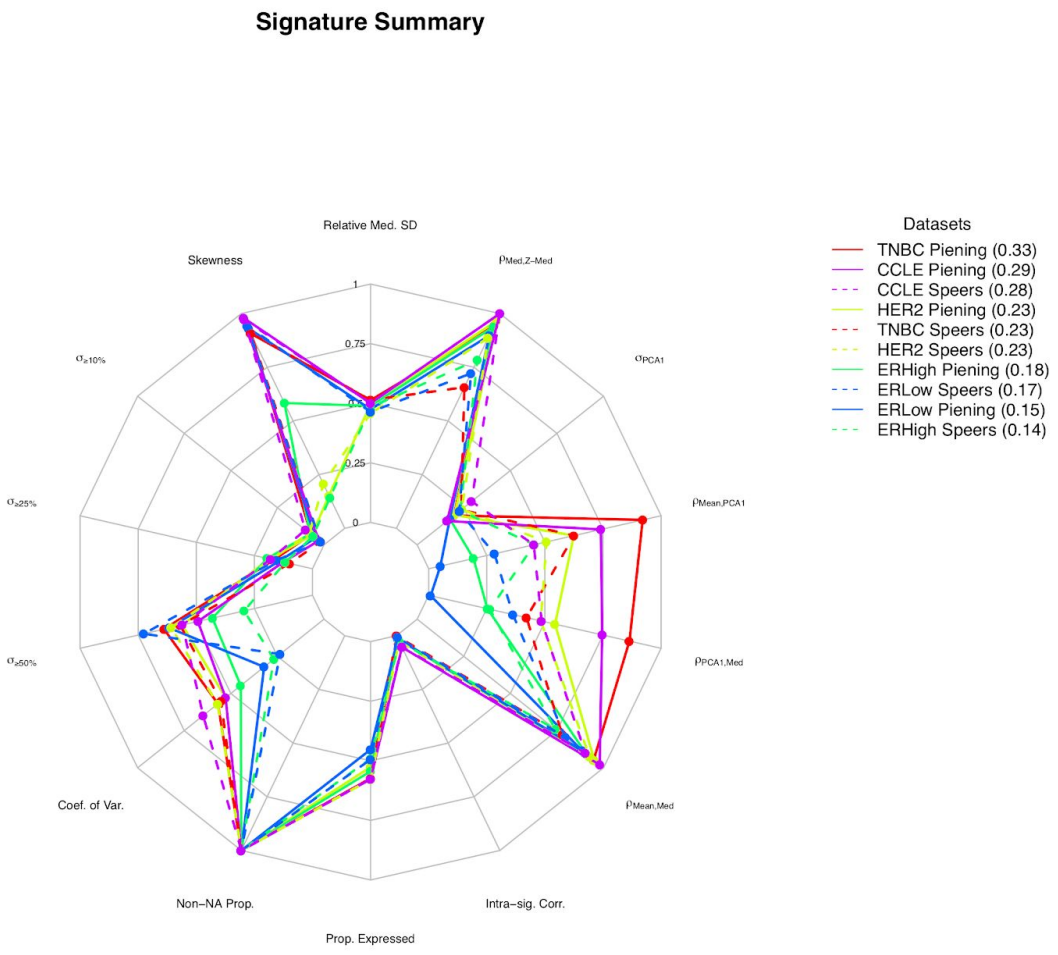
Signature	Number of genes	Histology
Piening et al. (Piening *et al.*, 2009)	281	Breast Cancer
Nuyten et al. (Nuyten *et al.*, 2006)	335	Breast Cancer
Torres-Roca et al. (Torres-Roca *et al.*, 2005)	10	Pan-cancer
Speers et al. (Speers *et al.*, 2015)	51	Breast Cancer
Pitroda et al. (Pitroda *et al.*, 2014)	4	Carcinoma
Kim et al. (Kim *et al.*, 2012)	31	Pan-cancer
De-Jong et al. (de Jong *et al.*, 2015)	1177	Head and Neck Squamous Cell Carcinoma
Amundson et al. (Amundson *et al.*, 2008)	21	Pan-cancer
Abazeed et al. (Abazeed *et al.*, 2013)	2248	Squamous Cell Lung Cancer
Weichselbaum et al. (Weichselbaum *et al.*, 2008)	49	Squamous Cell Carcinoma
Tang et al. (Tang *et al.*, 2017)	26	Soft Tissue Sarcoma

Supplementary Table 2 presents the original study, number of genes in the signature and the histology for the radiation response gene expression signatures under **hypoxic conditions** (absence of oxygen).

Signature	Number of genes	Histology
Toustrup et al. (Toustrup *et al.*, 2011)	15	Head and Neck Cancer
Eustace et al. (Eustace *et al.*, 2013)	26	Head and Neck Cancer
Lendahl et al. (Lendahl *et al.*, 2009)	30	Head and Neck Cancer
Winter et al. (Winter *et al.*, 2007)	99	Pan-Cancer
Buffa et al. (Buffa *et al.*, 2010)	51	Pan-Cancer
Hu et al. (Buffa *et al.*, 2010; Hu *et al.*, 2009)	13	Breast Cancer
Sorensen et al. (Sørensen *et al.*, 2010)	28	Squamous Cell Carcinoma
Yang et al. (Yang *et al.*, 2018)	28	Prostate Cancer
Fjeldbo CS et al. (Fjeldbo *et al.*, 2016)	6	Cervical Cancer
Lang et al. (Yang *et al.*, 2017)	24	Bladder Cancer
Ragnum et al. (Ragnum *et al.*, 2015)	32	Prostate Cancer
Seigneuric et al. (Seigneuric *et al.*, 2007) (Early 0% hypoxia)	72	Mammary Epithelial
Seigneuric et al. (Seigneuric *et al.*, 2007) (Late 0% hypoxia)	71	Mammary Epithelial
Seigneuric et al. (Seigneuric *et al.*, 2007)	34	Mammary Epithelial

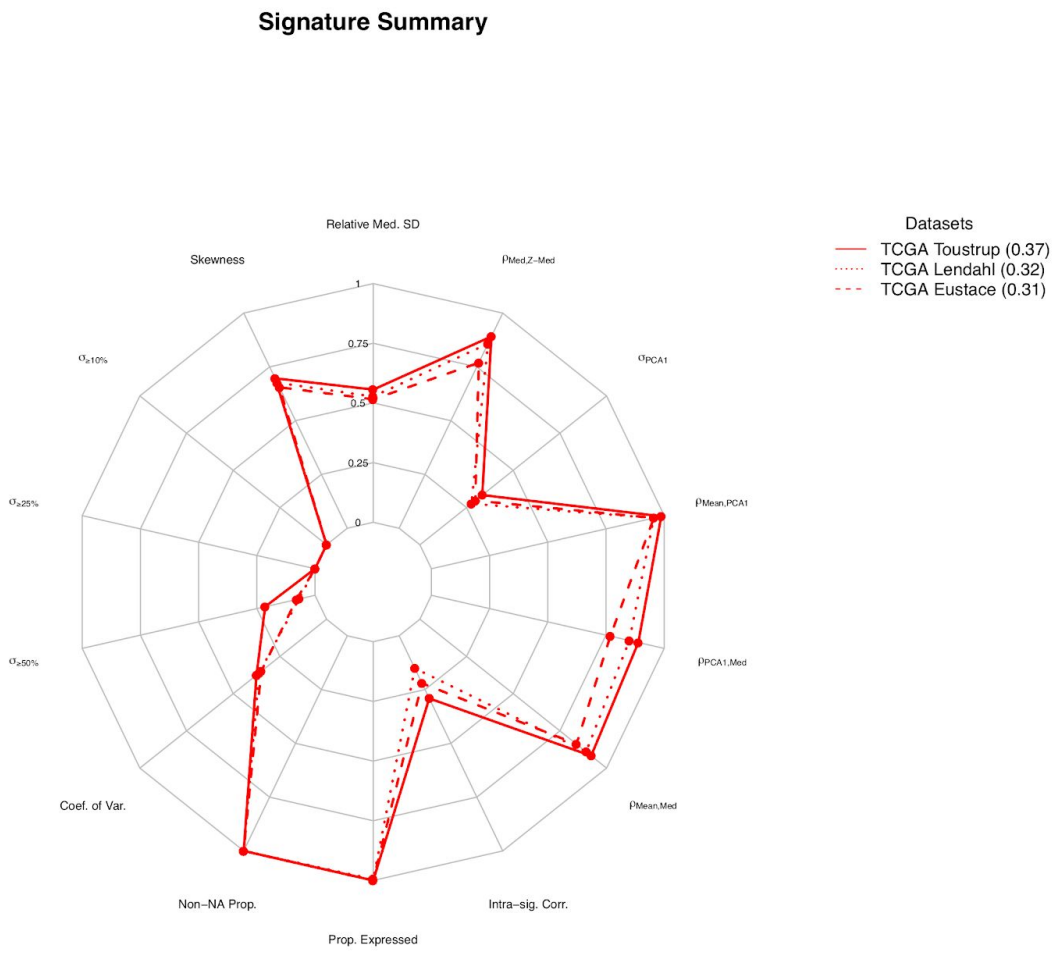
(Early 2% hypoxia)		
Seigneuric et al. (Seigneuric *et al.*, 2007) (Late 2% hypoxia)	32	Mammary Epithelial
Starmans et al. (Starmans *et al.*, 2012) (Cluster 1)	69	Breast Cancer
Starmans et al. (Starmans *et al.*, 2012) (Cluster 2)	246	Breast Cancer
Starmans et al. (Starmans *et al.*, 2012) (Cluster 3)	157	Breast Cancer
Starmans et al. (Starmans *et al.*, 2012) (Cluster 4)	95	Breast Cancer
Starmans et al. (Starmans *et al.*, 2012) (Cluster 5)	162	Breast Cancer
Starmans et al. (Starmans *et al.*, 2012) (Cluster 6)	14	Breast Cancer
Starmans et al. (Starmans *et al.*, 2012) (Cluster 7)	28	Breast Cancer
Ghazoui et al. (Ghazoui *et al.*, 2011)	70	ER+ Breast Cancer
Van Malenstein et al. (van Malenstein *et al.*, 2010)	7	Hepatocellular Carcinoma

Supplementary Figure 1



Supplementary Figure 1: Radar plots produced for *sigQC* summary metric evaluation of Piening and Speers signatures on CCLE and METABRIC datasets. Each ray of the radar plot evaluates one of the summary metrics checked by *sigQC* for gene signature quality prior to application on a dataset. Full explanation of each metric can be found in Dhawan et al., 2017. Briefly, this plot reveals that the Piening signature on the TNBC dataset (solid red line) is among the highest-quality application, along with the Piening signature on the CCLE dataset (solid purple line). Among the lowest performing signature-dataset combinations are the Speers and Piening signatures on the ER-High and ER-Low datasets, as poorer performance is observed on nearly every metric considered. Signature-dataset combinations are ranked in the legend by quality, with the numeric value in brackets representing the area contained within the radar chart lines for each signature-dataset combination.

Supplementary Figure 2



Supplementary Figure 2: Radar plots produced for *sigQC* summary metric evaluation of Toustrup, Lendahl, and Eustace signatures on TCGA HNSCC dataset. Each ray of the radar plot evaluates one of the summary metrics checked by *sigQC* for gene signature quality prior to application on a dataset. Full explanation of each metric can be found in Dhawan et al., 2017. Briefly, this plot reveals that each of the three signatures has approximately similar quality when applied on this dataset, with the Eustace signature showing a slight reduction in quality, owing primarily to differences in signature gene score correlation, likely owing to model-specificity in its use. Signature-dataset combinations are ranked in the legend by quality, with the numeric value in brackets representing the area contained within the radar chart lines for each signature-dataset combination.

REFERENCES

Abazeed,M.E. *et al.* (2013) Integrative radiogenomic profiling of squamous cell lung cancer. *Cancer Res.*, **73**, 6289–6298.

Amundson,S.A. *et al.* (2008) Integrating global gene expression and radiation survival parameters across the 60 cell lines of the National Cancer Institute Anticancer Drug Screen. *Cancer Res.*, **68**, 415–424.

Buffa,F.M. *et al.* (2010) Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br. J. Cancer*, **102**, 428–435.

Eustace,A. *et al.* (2013) A 26-gene hypoxia signature predicts benefit from hypoxia-modifying therapy in laryngeal cancer but not bladder cancer. *Clin. Cancer Res.*, **19**, 4879–4888.

Fjeldbo,C.S. *et al.* (2016) 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.

Ghazoui,Z. *et al.* (2011) Close and Stable Relationship between Proliferation and a Hypoxia Metagene in Aromatase Inhibitor-Treated ER-Positive Breast Cancer. *Clin. Cancer Res.*, **17**, 3005–3012.

Hu,Z. *et al.* (2009) A compact VEGF signature associated with distant metastases and poor outcomes. *BMC Med.*, **7**, 9.

de Jong,M.C. *et al.* (2015) Pretreatment microRNA Expression Impacting on Epithelial-to-Mesenchymal Transition Predicts Intrinsic Radiosensitivity in Head and Neck Cancer Cell Lines and Patients. *Clin. Cancer Res.*, **21**, 5630–5638.

Kim,H.S. *et al.* (2012) Identification of a radiosensitivity signature using integrative metaanalysis of published microarray data for NCI-60 cancer cells. *BMC Genomics*, **13**, 348.

Lendahl,U. *et al.* (2009) Generating specificity and diversity in the transcriptional response to hypoxia. *Nat. Rev. Genet.*, **10**, 821–832.

van Malenstein,H. *et al.* (2010) A seven-gene set associated with chronic hypoxia of prognostic importance in hepatocellular carcinoma. *Clin. Cancer Res.*, **16**, 4278–4288.

Nuyten,D.S.A. *et al.* (2006) Predicting a local recurrence after breast-conserving therapy by gene expression profiling. *Breast Cancer Res.*, **8**, R62.

Piening,B.D. *et al.* (2009) A radiation-derived gene expression signature predicts clinical outcome for breast cancer patients. *Radiat. Res.*, **171**, 141–154.

Pitroda,S.P. *et al.* (2014) DNA repair pathway gene expression score correlates with repair proficiency and tumor sensitivity to chemotherapy. *Sci. Transl. Med.*, **6**, 229ra42.

Ragnum,H.B. *et al.* (2015) The tumour hypoxia marker pimonidazole reflects a transcriptional programme associated with aggressive prostate cancer. *Br. J. Cancer*, **112**, 382–390.

Seigneuric,R. *et al.* (2007) Impact of supervised gene signatures of early hypoxia on patient survival. *Radiother. Oncol.*, **83**, 374–382.

Sørensen,B.S. *et al.* (2010) Identifying pH independent hypoxia induced genes in human squamous cell carcinomas in vitro. *Acta Oncol.*, **49**, 895–905.

Speers,C. *et al.* (2015) Development and Validation of a Novel Radiosensitivity Signature in Human Breast Cancer. *Clin. Cancer Res.*, **21**, 3667–3677.

Starmans,M.H.W. *et al.* (2012) The prognostic value of temporal in vitro and in vivo derived hypoxia gene-expression signatures in breast cancer. *Radiother. Oncol.*, **102**, 436–443.

Tang,Z. *et al.* (2017) Predicting radiotherapy response for patients with soft tissue sarcoma by developing a molecular signature. *Oncol. Rep.*, **38**, 2814–2824.

Torres-Roca,J.F. *et al.* (2005) Prediction of radiation sensitivity using a gene expression

classifier. *Cancer Res.*, **65**, 7169–7176.
Toustrup,K. *et al.* (2011) Development of a hypoxia gene expression classifier with predictive impact for hypoxic modification of radiotherapy in head and neck cancer. *Cancer Res.*, **71**, 5923–5931.
Weichselbaum,R.R. *et al.* (2008) An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 18490–18495.
Winter,S.C. *et al.* (2007) Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res.*, **67**, 3441–3449.
Yang,L. *et al.* (2017) A Gene Signature for Selecting Benefit from Hypoxia Modification of Radiotherapy for High-Risk Bladder Cancer Patients. *Clin. Cancer Res.*, **23**, 4761–4768.
Yang,L. *et al.* (2018) Development and Validation of a 28-gene Hypoxia-related Prognostic Signature for Localized Prostate Cancer. *EBioMedicine*, **31**, 182–189.