Technische Universität München
Fakultät für Medizin

Radiogenomics in Ewing Sarcoma – Deciphering the Interphase of Functional Genomics and Functional Imaging

Carolin Marie Prexler

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Wolfgang A. Weber

Prüfer*innen der Dissertation:

1. Prof. Dr. Stefan Burdach
2. Prof. Dr. Hans-Werner Mewes

Die Dissertation wurde am 20.10.2022 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 21.02.2023 angenommen.

Contents

1 Introduction

1.1 Ewing Sarcoma

1.1.1 Epidemiology and Prognosis

Ewing sarcoma (EwS) is the second-most common primary bone malignancy in children and adolescents, which can also affect soft tissue [1-3]. It occurs in adults as well, where incidence and prognosis decrease with age. Survival rates in EwS have strongly increased from only 10% in the 1970s to currently 70-80% for localized disease [1, 4, 5]. This has been due to the introduction of a multimodal therapeutic approach combining chemotherapy, surgery and/ or radiation [6]. However, the 5-year survival rate for patients with metastases or recurrences is still dismal at only 10-30% [1, 5]. Furthermore, little therapeutic progress to improve outcome has been made in the last two decades [7-10], even for localized disease.

Therapeutic intensity stratification is of utmost importance in children because of severe long-term effects. Therapies have acute as well as chronic adverse effects leading to organ toxicities, mental disabilities or even secondary cancer [1, 11].

Both aspects, toxicity and poor survival rates, require treatment improvement, for example by the discovery of new methods, and individualization of therapy [12].

1.1.2 Genetic Background

EwS was first described by James Ewing in 1921 and has been characterized in great genetic detail in recent decades. The genetic characteristic of EwS is a translocation event, most commonly t(11;22)(q24;12) [2, 3, 13]. This translocation fuses the 5' RNA binding domain of a member of the FET gene family with the 3' domain of a member of the ETS gene family, which are transcription factors (TFs) [6]. The most common fusion is EWS-FLI1 occurring in 85-90% of cases [1-3], followed by EWS-ERG, and rarely fusions of EWS to ETV1, ETV4, or FEV [2]. The chimeric fusion protein is an oncoprotein that activates or represses transcription of many target genes [2, 13, 14]. The fusion is necessary and sufficient for tumorigenesis as no additional mutation is required [15]. However, EWS-FLI1 expression is usually toxic for cells [16]. It can only be expressed at stable levels in cells that are capable of mitigating replicative stress [16]. This is the case in cells that tolerate high proliferation rate and transcriptional activity at the same time, which are cells in developmental stages [16]. This could be a reason that EwS shows a stem-cell like phenotype with poor differentiation [15, 17]. However, the exact cell of origin is still under debate: candidates are mesenchymal stem cells, neural crest cells, or even endothelial cells [17-24].

Apart from the EWS-ETS fusion, there are few somatic mutations in EwS tumors. 4 broad genetic studies on patients with EwS found recurrent mutations in STAG2 (mostly loss of function mutations in 21.5%, 17% or 15-20% of EwS tumors, respectively), in TP53 (mostly truncating mutations in 6.2%, 7.1%, or 5-20%, respectively) and in CDKN2A (eg homozygous deletions in 13.8%) [2, 3, 15, 25].

The frequency of these mutations indicates that large cohorts with 100s of patients are required for genetic studies of EwS.

However, despite its homogeneous genetic background, EwS shows a variable phenotype. Thus, epigenetics and transcriptional regulation are thought to play an important role [1-3, 17, 25, 26]. However, it is unclear how they contribute to the malignant phenotype and whether they are related to imaging data.

1.1.3 Necessity of New Targets for Treatment

Since EwS is so clearly characterized by the fusion transcript, the obvious question is whether this can be exploited for treatment?

The problem is that the fusion transcript itself cannot be targeted [15, 25]. It has been observed that EWS-FLI1 expression levels vary in EwS tumors, which has opposed consequences: high levels promote proliferation, whereas low levels foster invasiveness [16]. Hence, downregulation of EWS-FLI1 leads to metastasis [6].

In addition, targeted immunotherapies are difficult in EwS. First, the oncoprotein EWS-FLI1 does not bind well to human leukocyte antigen (HLA), and therefore is not presented on tumor cells [27]. Second, there are hardly any additional mutations, which is why no other tumor specific antigens are presented.

In conclusion, new targets for treatment have to be identified.

1.2 General Characteristics of Childhood Cancer

1.2.1 Pediatric Versus Adult Cancer: Genetic Differences in Oncogenesis

As said, EwS is genetically very homogeneous. In general, genetic homogeneity is a typical pattern in childhood cancer and fundamentally distinguishes it from cancer in adults. There are four basic differences of pediatric and adult cancer on genetic level [28].

First, pediatric cancer accumulates 14 times less mutations than adult cancer (0.13 versus 1.8 mutations per megabase, respectively) [11, 28]. 3 large genetic studies on patients with EwS found a low mutational burden of 0.15 and 0.38 mutations per megabase, respectively [3, 15], with on average 319 mutations and 10 coding mutations per genome [2]. The number of mutations increases for patients with relapse by the factor of 2 to 3 [11, 15, 25]. This may be due to treatment effects as Crompton et al. [25] observed 3 times more somatic mutations after treatment in pediatric cancer patients. The number of mutations in EwS seems to be independent of the age of the patient [11, 15]. The low mutational burden in EwS may be explained by the gene fusion, as for example fusion driven pediatric cancers like fusion-positive rhabdomyosarcoma have less mutations than fusion-negative rhabdomyosarcoma [3].

The second difference between pediatric and adult cancer is the number of tumor drivers. Pediatric cancers are in most cases characterized by a single event, while there are multiple drivers in adult cancer [28].

Third, the mutations in pediatric and adult cancer affect different genes [28]. Gröbner et al. [11] found that 30% of genes mutated in pediatric cancer overlap with genes mutated in adult cancer, Ma et al. [29] found 45%, which means that only about one third of the genes mutated in pediatric cancer are genes that are typically known from adult cancer types [28]. Gröbner et al. [11] found 77 genes mutated in pediatric cancer, so called significantly mutated genes (SMGs). SMGs are found in 93% of adult tumors but only in 47% of pediatric tumors, even rarer in some entities like EwS (<15%) [11]. What pediatric and adult SMGs have in common, is that most genes have “epigenetic” or “transcriptional regulation” function [11]. In contrast, in adult cancer “DNA repair” genes are frequently muted, which only play a minor role in pediatric cancer [11]. This shows that DNA repair is often disrupted in adult cancer types and contributes to tumorigenesis, but is functional in pediatric cancer. This suits the fact that

pediatric cancer do not accumulate mutations [17]. The only gene in “DNA repair” that is also mutated in pediatric cancer is TP53 [11].
The fourth difference is that the tumor drivers are entity specific in pediatric cancer [28]. Each entity is defined by its characteristic genetic event, e.g. which genes are fused by a chromosomal translocation. The timing of the genetic event in developmental context is important for tumorigenesis [28]. In contrast, tumor drivers are often shared across adult cancer types, where usually multiple hits are needed for cancer development [28].

In conclusion, these differences between adult and pediatric cancer may reflect two distinct ways of cancer development [28].

1.2.2 Examination of Systemic Diseases Requires Whole-Body Approaches Like Imaging

Despite the differences in oncogenesis between children and adults, there are difficulties in cancer treatment that are in common.
The major challenge for targeted therapies is resistance, which is related to heterogeneity and plasticity of tumors [30]. When cancer becomes a systemic disease, the treatment of patients is complex. Lesions are often heterogeneous among themselves, and in addition there is intratumoral heterogeneity between different clones. This may give rise to both resistance and recurrence.
When performing tissue studies to characterize the disease, biopsy of a single lesion is clinical standard. It is challenging to assess multiple or even all lesions. However, the complexity of a systemic disease cannot be mapped with such a locally restricted procedure. Whole-body molecular imaging might aid to overcome these limitations [31, 32]. Radiomics offers new opportunities in analyzing these large imaging datasets. The combination of advanced image analysis with tissue based genomic data – called radiogenomics – allows an in-depth characterization of the disease.

1.3 PET Tumor Glucose Uptake Indicating Malignancy

1.3.1 Functional Imaging

A common method in clinical practice for diagnosing tumors and assessing therapy response is 18-F-fluorodeoxyglucose positron emission tomography (18F-FDG-PET). For 18F-FDG-PET, radiolabeled glucose is used, which is taken up like normal glucose and accumulates in the cells. Thus, one can measure the glucose uptake quantitatively by measuring the radioactivity concentration: cells with increased uptake correspond to a stronger PET signal, which also refers to increased metabolic activity. This is why PET is also referred to as functional molecular imaging, because it does not give insight into the anatomy like computed tomography (CT) or magnetic resonance (MR), but provides information about functional processes in the cells.
The signals in PET are measured quantitatively per volume unit (voxel) as standardized uptake value (SUV). As the name indicates, SUV is a relative measure of radioactivity concentration standardized by the weight of the patient and the amount of injected radiolabeled glucose, which also incorporates the decay in the meantime between injection and imaging. Due to the standardization, SUVs can be compared between patients.

In short, a tumor in PET consists of a collection of 3D voxels with corresponding signal strength given as SUVs. The SUV values can be summarized in different ways to characterize the tumor based on image features (IFs). One way is to use the maximum SUV in the tumor region (SUVmax) describing the highest rate of FDG uptake in the tumor tissue. Advantageously, SUVmax is invariant to the exact definition of the tumor boundaries, called segmentation [33]. Segmentation is a nontrivial problem, as boundaries cannot be clearly defined in the peripheral regions of a tumor with lower SUV values. Thus, segmentation varies depending on the delineation created by algorithms or different experts [31, 34-36]. This variability affects most IFs [36, 37] but does not bias SUVmax [38, 39].

1.3.2 Prognostic Value of SUV

Quantitative FDG-PET parameters and in particular SUVmax have been shown to be prognostic in several cancers and at different points during the course of the disease.

Higher SUVmax correlates with tumor growth, worse survival, poor prognosis, advanced stage, and worse course of disease in various cancer entities (thyroid cancer, cervical cancer, breast cancer, glioma, ovarian cancer, head and neck squamous cell cancer, renal cell carcinoma, lung cancer, esophageal cancer, and pancreatic cancer amongst others) [40-54]. SUVmax is prognostic in primary cancer [41, 42, 46, 48], as well as in recurrence [44, 45, 52-54]. A lung cancer study found that SUVmax signals show similar pattern for primary and relapsed disease [52]. Furthermore, SUVmax is found to be prognostic in studies on pre-treatment PETs [42, 47, 52], and PETs after neoadjuvant therapy [51] in different types of cancer.

The prognostic value of SUVmax applies similarly to EwS, where PET signals vary significantly between individual patients. High SUVmax is associated with worse survival, poor histopathological response and stage of the disease in primary EwS [55-60]. So far, there are no studies examining the prognostic value in relapsed EwS. However, SUVmax values were studied in primary and recurrence: a comprehensive study on almost 500 bone and soft tissue sarcoma patients showed a similar SUVmax distribution for primary and relapsed cases, especially for EwS [61]. Furthermore, like in other entities, SUVmax is prognostic pre-treatment [55-57, 59] and after neoadjuvant therapy [57-60]. Given these previously published data, it is reasonable to assume that SUVmax may also prove prognostic in both primary and relapsed EwS.

However, the biological basis for the generally elevated but variable glucose metabolism of Ewing sarcoma has not been well explored so far.

1.4 Objectives of Our Study

1.4.1 Objective 1: To Investigate Information Content of Transcriptomic and Radiomic Data

To our knowledge, there is no radiogenomic study in EwS using large-scale transcriptomic data with molecular, functional imaging. In the present study, we connect these data in order to investigate their relation and whether they complement each other. We attempt to infer tumor entity or genomic features from molecular imaging parameters, and predict survival based on the different data types.

1.4.2 Objective 2: To Elucidate the Transcriptomic Basis for Varying Glucose Uptake

As stated, EwS has a homogeneous genetic background but shows a variable phenotype, e.g. regarding glucose uptake or survival.
We investigate SUVmax referring to the glycolytic activity of the tumor. As the biological basis for increased glucose uptake is not well understood, this is correlated with transcriptomic data to search for differences in gene expression levels regarding SUVmax.

Our assumption is that gene expression changes will also be visible in altered PET signals.
Several studies in entities other than EwS (lung cancer, breast cancer, hepatocellular carcinoma, and head and neck cancer) state that gene expression affects PET signals. This is not limited to glycolytic pathways, but involves multiple cellular processes as alterations in pathways reflecting a broad spectrum of oncogenic changes were found [62-69]. These include pathways associated with the hallmarks of cancer [70, 71], such as the cell cycle, proliferation, apoptosis, metabolism, (oncogenic) signaling and signal transduction, cell-to-cell adhesion and extracellular matrix organization, DNA repair, and immune system [63, 64, 66-68].
Thus, we aim at characterizing EwS with variant SUVmax on the transcriptomic level, and hypothesize about the biological basis for elevated glucose uptake. We do not perform causational studies, but we provide observations of correlations and assume that the PET signal is driven by transcriptomics and not the other way around.

1.4.3 Objective 3: To Identify Prognostic Biomarker Candidates

As was mentioned before, new treatment options are urgently needed. A first step to accomplish this is to identify genes that can serve as biomarkers for risk stratification, i.e., correlate with survival or malignancy features.
Due to the prognostic value of SUVmax, we use it as quantitative phenotype in the analysis of the transcriptomic data. The results indicate which genes and pathways may be more or less active with regard to increased levels of malignancy. We hypothesize which genes or pathways are biomarker candidates for risk stratification and qualify for future studies of novel treatment options.

2 Materials and Methods

An overview of all analysis steps is provided in Figure 1, which depicts the whole workflow.
In general, we used R Statistical Software (v4.0.2) [72] for data analysis, statistical testing and plot generation.
For all machine learning tasks, we used random forest modeling as this method outperformed other machine learning methods in several studies across different omics datasets [73-76].

2.1 Inclusion Criteria and Clinical Data

We included all patients suffering from Ewing sarcoma (EwS) in the „Kinderklinik Schwabing" and "Klinikum rechts der Isar der Technischen Universität München" in the years 2011 to 2019 who fulfilled our inclusion criteria. The inclusion criteria were as follows: all patients suffering from primary or relapsed EwS, aged up to 40 years, with image data and tissue sampled from the same lesion. Patients without any prior therapy or untreated relapse were classified as "untreated", otherwise "treated". The time interval between tissue sampling and imaging had to be short: maximum 6 weeks for untreated patients, and maximum 2 weeks for treated patients, thus ensuring that tissue sample and PET reflected the same biological characteristics of the tumor. This yielded 19 samples from 17 patients.
The recorded clinical data (see
Table 1 and Suppl. Table 1) included sex (male, female), disease state (primary disease, relapse), sample type (tumor, metastasis), therapy (untreated, treated), age at PET imaging, imaging modality (PET-CT, PET-MR), and overall survival (begin: diagnosis of primary disease or relapse; end: death or last follow up).
The data were analyzed retrospectively. The registry study was approved by the local ethics committee (reference number 223/16S). All patients gave their written consent for data evaluation.

2.2 Gene Expression Data

2.2.1 Tissue and Microarray Preparation

Frozen tumor samples for expression analysis were obtained from biopsies or resection specimens, and passed the quality control of experienced pathologists. Sample preparation followed the Affymetrix protocol and was previously described [19]. Gene expression was measured using microarray chips [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version].
The advantage of using microarray chips for expression measurements was that our lab had experience in that method. We had collected a data basis comprising 144 samples: 31 normal tissue, 48 EwS, 20 osteosarcoma, 13 rhabdomyosarcoma, 11 synovial sarcoma, and 21 other entities. These could be used for normalization and quality control. This way, we identified 11 arrays of poor quality, which were excluded in a first step. The remaining 133 samples were clustered hierarchically. Based on this, we identified 29 EwS samples that could be used for analyses. They either showed a typical EwS expression profile or were tested positively for the EWS gene fusion and had good tissue quality. We could include 19/29 for the radiogenomic analysis as these also had PET data available.

2.2.2 Expression Data Preprocessing

The microarray data was processed using Robust Multichip Average (RMA) preprocessing [77-79] (R package oligo [80]). This method computed log2 expression values of genes by applying background subtraction, quantile normalization and median-polish summarization of probe sets to genes. For summarization, Brainarray (version 24) was used to obtain entrez gene ids. Brainarray supplies up-to-date definitions of probe sets, which are more accurate than the original Affymetrix definitions [81-83]. The entrez gene ids were further annotated by gene symbols using Brainarray (20722 genes).
Additional quality control was applied using hierarchical clustering of samples (Euclidean distance and average linkage using R package dendextend [84]) and boxplots of log2 expression values.

2.2.3 Expression Data Filtering

We applied two filtering steps on the gene expression data.
In a first step, we excluded genes with low expression, i.e., we removed genes with average expression below 10 as microarray chips do not function accurately at low expression levels. 198 genes were removed, 20524 genes remained.

The second filtering step was based on a machine learning approach applied to external datasets in order to obtain genes associated with EwS survival. We collected all EwS datasets from the GEO database [85] (as of March 2018) of gene expression plus survival data, which yielded 3 datasets: GSE63155 [73], GSE17618 [86], and GSE63156 [73]. The survival data was used to split the patients into long-term survival (overall survival, OS>5 years) or short-term survival (OS<5 years and dead). Ambiguous patients were excluded, which yielded n=31 for GSE63155 (12 short OS, 19 long OS), n=40 for GSE17618 (22 short OS, 18 long OS), and n=31 for GSE63156 (9 short OS, 22 long OS). Each dataset was analyzed separately to avoid bias of different data sources.
The external expression data was preprocessed the same way as our expression data described above. To reduce dimensionality, 50% of the genes with low coefficient of variation were removed [87] as these are constant in the dataset and have low information content for distinguishing subgroups of patients.
Random forest classification was applied to the external datasets to predict binary OS in repeated 10-fold cross-validation (R package caret [88], method “rf” using 1000 trees). Due to class imbalance in all datasets, downsampling was performed to obtain the same number of samples with long- and short-term OS. To increase the number of subgroups analyzed, we repeated this analysis 10 times. For each dataset, the models selected a set of genes that were informative for predicting survival time (see scheme in Figure 9). We called the overlap of the 3 gene sets “potential survival genes” as these were independently associated with survival in all 3 external datasets (1491 genes). Functional annotation of these genes was analyzed using DAVID Bioinformatics Resources (v6.8) [89, 90], category UP_KEYWORDS. Benjamini-adjusted q-values<0.01 were considered significant.
The “potential survival genes” were used to filter the expression data in our cohort (1376/1491 genes). Thus, we focused on genes that may be prognostic in EwS for our subsequent analyses.

2.3 Image Data

2.3.1 Image Feature Calculation

We analyzed positron emission tomography computed tomography (PET-CT) / positron emission tomography magnetic resonance (PET-MR) to obtain PET image features (IFs) for those lesions that were used for gene expression measurements. All patients underwent 18-F-fluorodeoxyglucose PET (18F-FDG-PET) imaging for diagnostic or staging purposes in our institution.

The PET series were analyzed and controlled for quality using OsiriX DICOM viewer [91] and tumors were delineated using cuboids by experienced nuclear medicine physicians. Exact, voxel-wise delineations (so-called masks) were obtained based on a standardized uptake value (SUV) threshold of 40% of the maximal SUV for each lesion, which is a common standard [56, 63, 92-94]. Based on the segmented tumors, 3D IFs were calculated according to the image biomarker standardization initiative (IBSI) [95] using PyRadiomics [96] v3.0.1 with standard settings (Python version 3.8 as of October 2020). Spatial resampling to 4x4x4mm was applied to obtain isotropic spacing for both PET-CT and PET-MR. This is needed as a non-uniform resolution affects some IFs [63]. For image discretization a fixed bin size of 0.5 was used, which has advantages compared to fixed bin number [97].

IFs from different feature classes were calculated, namely shape, first, and second-order features. Filters were applied to the original image series to obtain derived images, which were used to obtain higher-order IFs. Altogether, 1502 IFs were calculated. For a detailed description of the IFs and formulas, see [37, 95, 96].

SUVmax was obtained from the original image, i.e., no filters were applied (PyRadiomics feature “original_firstorder_Maximum”).

2.3.2 Image Feature Analysis

2.3.2.1 Robustness

To examine the robustness of IFs, the intraclass correlation coefficient (ICC) was calculated, which is a measure of inter-rater reliability. We analyzed robustness with regard to two aspects: imaging modality and lesion delineation.

For the analysis of robustness towards imaging modality, we used 21 double series (21 PET-CT plus 21 PET-MR) from 15 sarcoma patients, including Ewing sarcoma (n=13 double series), osteosarcoma (n=5), rhabdomyosarcoma (n=1), and synovial sarcoma (n=2). The lesions were segmented using 40% of SUVmax. The intraclass correlation for absolute inter-rater agreement was calculated for each IF by comparing PET-CT and PET-MR using the R package irr [98], function icc (settings: twoway model, type agreement). To define robust IFs, we used a cutoff of ICC≥0.75 [99, 100], resulting in 260/1502 features.

To study the sensitivity towards delineation variability, different segmentations were generated per lesion, based on 20%, 30%, and 40% of SUVmax. ICC for consistency was calculated for each IF using the R package irr [98], function icc (settings: twoway model, type consistency). This was analyzed separately for each imaging modality: we used PET-CT series (n=106 sarcoma patients, referred to as analysis a) and PET-MR series (n=54, analysis b). In addition, we examined a mixed dataset of all available PET series (n=160, analysis c on PET-CT and PET-MR).

To define IFs robust to delineation variation, we used a cutoff of ICC≥0.8 [96]. This resulted in 1157/1502 stable IFs in PET-CT series (analysis a), 1071/1502 stable IFs in PET-MR series

(analysis b), and 1141/1502 stable IFs in the mixed dataset (analysis c). The sets of stable IFs from analysis a, b, and c overlapped in 1009/1502 IFs. Venn diagrams were generated using the tool jvenn [101].
Finally, we combined both analyses. This resulted in 184 IFs that were robust towards imaging modality and lesion delineation.

2.3.2.2 Correlation and Redundancy

We investigated whether the IFs were redundant or complementary descriptors for a tumor. For this analysis, we used all PET series of sarcoma patients (PET-CT and PET-MR, n=160) based on 40% SUVmax segmentation. The pairwise Pearson correlation was calculated for all 1502 IFs. Plots were generated using the R package psych [102] (function corPlot) and the R package ggcorrplot [103] (function ggcorrplot).
We obtained a set of non-redundant IFs by applying a correlation cutoff of 0.9 [104]. If two IFs had abs(r)>0.9, the IF with higher mean absolute correlation to all IFs in the dataset was removed (R function caret::findCorrelation [88]). This resulted in 321/1502 non-redundant IFs.

2.4 Image Based Diagnosis Prediction

We analyzed whether it is possible to distinguish different sarcoma entities or molecular subtypes in EwS based on PET radiomic data (see Figure 1, the panel in the middle on the left). For this prediction study, we used 184 IFs that were robust towards imaging modality and lesion delineation. To predict the sarcoma entity, the dataset included 138 PET series: 75 EwS and 63 non-EwS (35 osteosarcoma, 15 rhabdomyosarcoma, 10 synovial sarcoma, and 3 others). To predict the fusion type in EwS, the dataset included 67 PET series of EwS patients: 48/67 positive for EWS-FLI1, and 19/67 negative for EWS-FLI1. The 19 EWS-FLI1 negative tumors included tumors with approved EWS-ERG or EWS-ETV1 translocations, tumors with EWS fusions with a unknown translocation partner other than FLI1, and tumors diagnosed as EwS but with negative test for the EWS-FLI1 fusion.
For each task, a random forest classification model was built (R package caret [88], method “rf” using 500 trees). Before modeling, each dataset was split into a training cohort (80% of samples) and a validation cohort (20% of samples). The training cohort was downsampled to obtain equal group sizes. Then it was used for parameter tuning in 5-fold repeated cross-validation and model generation. The model with highest area under the receiver operating characteristic curve (ROC AUC) was chosen as final model, and tested on the validation cohort.

2.5 Statistical Analysis of SUVmax and Clinical Data

2.5.1 SUVmax Distribution Regarding Clinical Variables

We tested for equal distribution of SUVmax values regarding the following clinical variables: sex (male or female), disease state (primary disease or relapse), sample type (tumor or metastasis), therapy (treated or untreated), and age (<=15 years or >15 years). A two-sided Welch two-sample t-test was applied (R function stats::t.test). P-values<0.05 were considered significant.

2.5.2 Survival Analysis

We applied univariate Kaplan-Meier analyses for overall survival with log-rank tests (R package survival [105, 106]). We tested for sex (male vs. female), disease state (primary disease vs. relapse), age (<=15 years vs. >15 years), and SUV categories (samples split by median SUVmax into low SUV vs. high SUV). P-values<0.05 were considered significant.
Furthermore, we built multivariate Cox proportional hazards models (R package survival [105, 106]) including SUV and disease state (primary disease vs. relapse). We fitted two models: the first model used binary SUV categories (low SUV vs. high SUV defined by median SUVmax), the second model used continuous SUVmax values. Forest plots were generated using R package survminer [107]). P-values<0.05 were considered significant.

2.5.3 Proliferation and SUVmax

We used linear regression to correlate SUVmax and the expression of 4 common proliferation markers: PCNA (proliferating cell nuclear antigen), MKI67 (marker of proliferation Ki-67), TOP2A (DNA topoisomerase II alpha), and MCM2 (minichromosome maintenance complex component 2). To determine whether there was a link between SUVmax and the proliferation rate, we examined the slope of the regression lines and r^2, i.e., the squared Pearson correlation coefficient.

2.6 Correlation Analysis With Radiomic Data

2.6.1 Outcome Prediction Using Radiomic and Transcriptomic Data

We examined whether overall survival time could be inferred from transcriptomic data, from radiomic data, or from transcriptomic plus radiomic data (see Figure 1, the panel at the bottom left). For this, we predicted overall survival of the patients using random forest classification models with leave-one-out cross-validation (R package caret [88], method “rf” using 1000 trees).
The patients were divided into "long survival" or "short survival" based on their overall survival time using a cutoff of 2 years. Patients, who had died within 2 years had “short survival” (n=11), and patients with overall survival of at least 2 years had “long survival” (n=6). As 2 patients were excluded who were lost to follow up within 2 years, we examined 17/19 patients of our cohort (Suppl. Table 1).
We computed 3 models: model 1 based on transcriptomic data, model 2 based on radiomic data, and model 3 based on transcriptomic plus radiomic data. The transcriptomic input data comprised 20524 genes with average expression above 10 and the radiomic input data comprised 184 IFs robust to modality and delineation. Before model generation, the input data were filtered to remove features that were constant in the dataset. Features with coefficient of variation <0.7 were removed, respectively [87]. This resulted in 1774/20524 genes, and 112/184 IFs. Hence, model 1 was based on the expression of 1774 genes, model 2 was based on 112 IFs, and model 3 was based on 1886 mixed features (1774 genes + 112 IFs). The performance of the 3 models was compared using R package MLeval [108].

2.7 Correlation Analysis With SUVmax

2.7.1 Correlation of SUVmax and Immune Cell Infiltration

We investigated whether there was a relation between immune cell infiltration and SUVmax (see Figure 1, the panel at the bottom right, top row).

The infiltration rates were obtained from the deconvolution algorithm Cibersortx [109], which estimated absolute infiltration of 22 immune cell types based on bulk expression data (587 genes in pre-calculated reference profiles). The estimated cell types were: B cells (naive and memory), plasma cells, T cells (CD8, CD4 naive, CD4 memory resting, CD4 memory activated, follicular helper, regulatory (Tregs), and gamma delta), NK cells (resting and activated), monocytes, macrophages (M0, M1, and M2), dendritic cells (resting and activated), mast cells (resting and activated), eosinophils, and neutrophils. For each cell type, an absolute score was returned, which allowed intra- and inter-sample comparison. In addition, Cibersortx calculated empirical p-values for each sample, which represented the credibility of the deconvolution results. However, we did not consider the p-values in our analysis.

We analyzed the immune cell signature for all EwS samples with microarray data, which were 29 samples (referred to as the Schwabing dataset). We compared the signature to 4 public datasets of EwS tumors (Delattre GSE34620 [110], Savola GSE17618 [86], Dirksen GSE63157 [73], and Lawlor GSE68776 [111]) and a dataset of EwS cell lines (Savola GSE17618 cell lines [86]).

For 19/29 samples in the Schwabing dataset, we had SUVmax values available. For these, we correlated absolute CD8 T cell and M2 macrophage infiltration with SUVmax in two ways. First, we used continuous SUVmax values in linear regression. Second, we examined whether CD8 and M2 infiltration differed among SUV groups of tumors (tumors with high SUVmax compared to tumors with low SUVmax) using a Wilcoxon rank sum test.

2.7.2 Correlation of SUVmax and Gene Expression Using Linear Regression

After preprocessing and filtering of the expression data, 1376 genes remained which were correlated with SUVmax using linear regression (see Figure 1, the panel at the bottom right, bottom row). “Least squares” was used as fitting method. Moderated t-statistics were calculated based on an empirical Bayes method in the R package limma [112], which is an acronym for “linear models for microarray data”. Limma increases statistical power to determine differentially expressed genes [112, 113] and shows best performance and broad applicability in a comparison of different statistical testing methods for gene expression data [114]. We applied the limma-trend method to fit a trend to the prior variances, which is beneficial for microarray data, if the data has a mean-variance relationship [113]. Limma was run on all genes with average expression above 10 (20524 genes). We did not apply further filtering as we wanted to use as many genes as possible to obtain a stable estimate of the expression variance for modeling. Afterwards, the model results were filtered for the “potential survival genes”, which resulted from the machine learning approach described previously (1376 genes). We corrected for multiple testing using the Benjamini-Hochberg procedure, i.e., false discovery rate (FDR) adjustment [115]. Adjusted p-values (q-values)<0.05 were considered significant. For comparability, slopes were normalized using Z-score standardization.

Furthermore, we defined relations of high effect size by transferring the standard abs(logFC) cutoff of 1 – which corresponds to doubling or halving of gene expression in classical two group comparison – to regression modeling. When splitting our cohort by median SUVmax

into two groups, the group of low SUV lesions had a mean SUVmax of 3.2, whereas the group of high SUV lesions had a mean SUVmax of 10. When transferring the comparison of groups to regression, a doubling or halving of gene expression over 6.8 SUV units, which was the difference in means of the groups, corresponded to abs(slope) of 0.146. Therefore, we considered all relations of abs(slope)>0.146 to be of high effect size (23 genes).
The heatmap depicting Z-scaled gene expression in all samples with clinical data in the side bars was generated using R function GMD::heatmap.3 [116]. The dendrograms are based on Euclidean distance and average linkage.

2.7.3 Enrichment Analyses

The results of linear regression of gene expression and SUVmax were further tested in enrichment analyses for functional annotation by applying 2 tools: Enrichr [117-119] and gene set enrichment analysis (GSEA v4.1.0) [120, 121]. The set of 23 genes of high effect size (abs(slope)>0.146) was analyzed using Enrichr (applied in September 2021). We focused on pathways in “Reactome 2016” and the Gene Ontology (GO) knowledgebase including “GO biological process 2021”, “GO molecular function 2021” and “GO cellular component 2021”. Adjusted p-values (q-values)<0.01 were considered significant.
Additionally, we performed GSEA based on all 1376 genes ranked by their correlation with SUVmax. As input, the genes were ranked by decreasing slope. On this ranked gene list, we tested annotations of these categories: H (hallmark gene sets), C2cp (curated gene sets: canonical pathway), C3tft (regulatory target gene sets: transcription factor targets), and C5 (ontology gene sets). For each gene set, a normalized enrichment score (NES) is calculated, which corresponds to the maximum or minimum value of the curve in the enrichment plot (enrichment score = ES) normalized by gene set size. If GSEA is used for hypothesis generation, the developers suggest a less stringent cutoff for significance. Thus, FDR<0.1 was considered significant.

3 Results

An overview of all analysis steps is provided in Figure 1, which depicts the whole workflow. The 3 panels at the top refer to the data that was used in this analysis: the PET image data, the clinical data, and the gene expression data. The 2 panels in the middle refer to analyses that combine clinical data and PET radiomic data or SUVmax, respectively. The 2 panels at the bottom refer to analyses that integrate transcriptomic data and PET radiomic data or SUVmax, respectively.

3.1 Description of Patient's Characteristics

We defined inclusion criteria to obtain the cohort for our retrospective analysis: all patients suffering from EwS, age up to 40 years, with image data and tissue sampled from the same lesion. Patients without any prior therapy or untreated relapse were classified as "untreated", otherwise "treated". We only included patients with short time interval between tissue sampling and imaging: maximal 6 weeks for untreated patients, maximal 2 weeks for treated patients. Thereby we ensured that tissue sample and PET reflect the same biological characteristics of the tumor.
Out of 75 patients referred to our institution during the duration of the study, only 17 met all of our quality standards and PET and tissue samples were available in a timely manner. We included 19 samples (primary disease n=5 and recurrence n=14) from 17 EwS patients (female n=10 and male n=7) aged 3 to 31 years. Tissue samples for expression analysis were obtained from one lesion each (tumor n=5 and metastasis n=14) before (n=12) or during (n=7) treatment. From this tissue, expression of all genes (20722) was assessed by microarray chips (HuGene-1_0-st Affymetrix Human Gene 1.0 ST Array, transcript (gene) version). Of the same lesion, we also analyzed positron emission tomography computed tomography (PET-CT, n=15) / positron emission tomography magnetic resonance (PET-MR, n=4) to obtain SUVmax, which measures the maximal glucose uptake in the tumor. The glucose uptake varied in our cohort (Figure 2): SUVmax showed a spectrum of 1.9 to 21.3, with a mean of 6.8 and median of 5.4. Based on the median, the samples were split into two groups: lesions with low SUV (n=9, SUVmax 1.9 to 5.1, mean 3.2, median 2.8) and lesions with high SUV (n=10, SUVmax 5.4 to 21.3, mean 10.1, median 9.0).

A summary of all patient characteristics is provided in Table 1 and Suppl. Table 1.

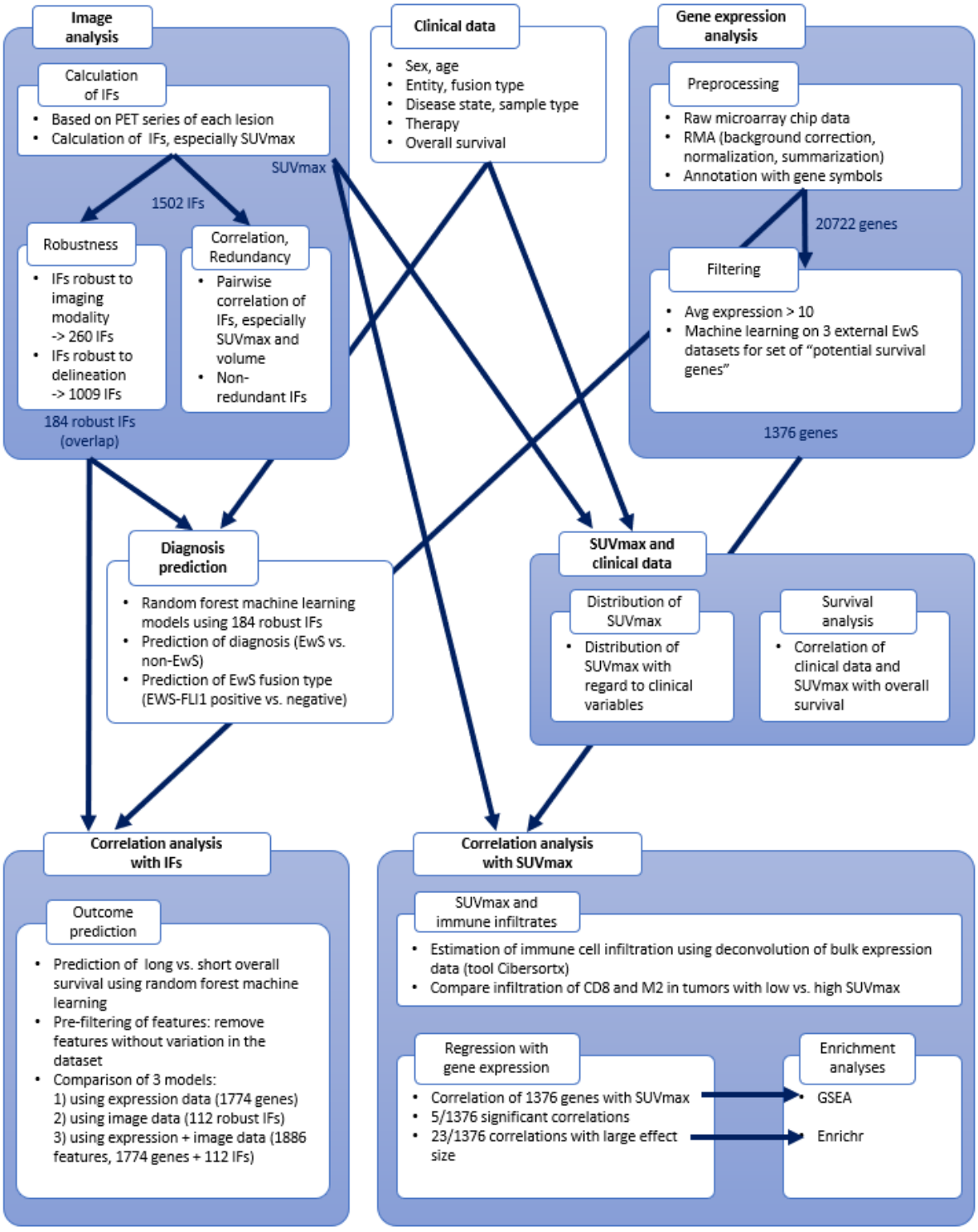


Figure 1: Workflow scheme.*The 3 panels at the top refer to the data that is used in this analysis: the PET image data, the clinical data, and the gene expression data. PET images are used to calculate IFs, especially SUVmax values. The IFs are analyzed regarding their robustness and redundancy. Gene expression data is preprocessed and filtered. The 2 panels in the middle refer to analyses that combine clinical data and PET radiomic data or SUVmax, respectively. Diagnosis is predicted based on PET radiomic data. The SUVmax distribution is analyzed with regard to clinical data, and a survival analysis is performed. The 2 panels at the bottom refer to analyses that integrate transcriptomic data and PET radiomic data or SUVmax, respectively. Transcriptomic and PET radiomic data are combined for outcome prediction. Transcriptomic data and SUVmax are combined to investigate associations of glucose uptake and immune cell infiltration or gene expression in general. The results of the correlation analysis of gene expression and SUVmax are further annotated using enrichment analyses.*

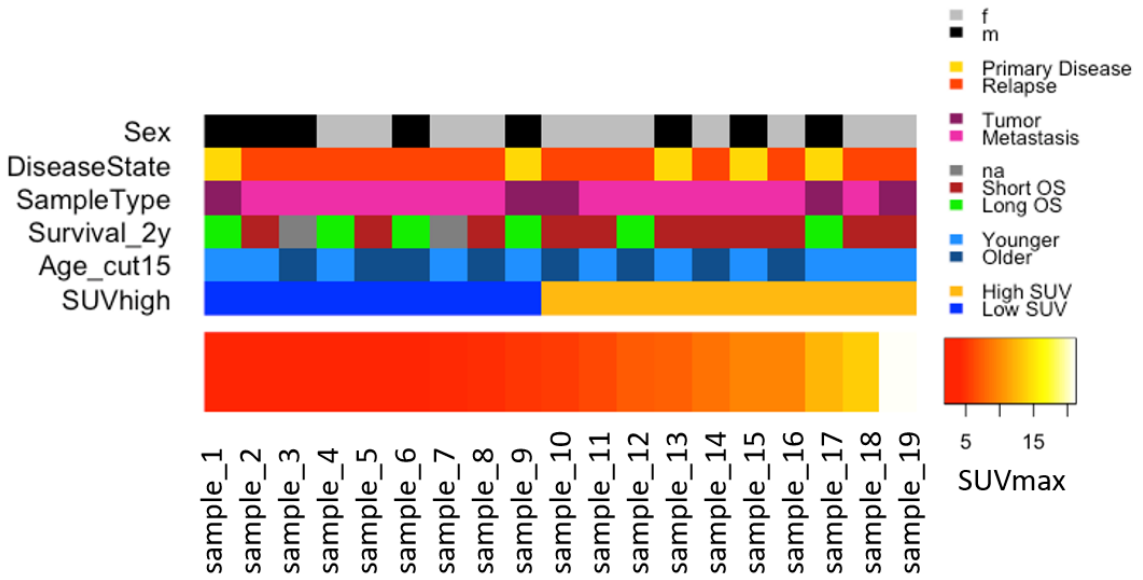


Figure 2: 19 EwS samples ordered by increasing SUVmax having a range of 1.9 to 21.3.*On top, additional clinical information is provided about sex (female or male), disease state (primary disease or relapse), sample type (tumor or metastasis), 2-year overall survival, age (≤15 years or >15 years), and categorical partitioning of the samples into high SUV or low SUV (split by median SUVmax).*

		Number	Fraction
Total		19	1
Sex	Female	11	0.58
	Male	8	0.42
Disease state	Primary disease	5	0.26
	Relapse	14	0.74
Sample type	Tumor	5	0.26
	Metastasis	14	0.74
Therapy	Untreated	12	0.63
	Treated	7	0.37
Age at PET	All: number	19	1
	Range	3 - 31	
	Median	14	
	Mean	14.8	
	≤15: number	11	0.58
	Range	3 - 15	
	Median	10	
	Mean	9.2	
	>15: number	8	0.42
	Range	17 - 31	
	Median	21	
	Mean	22.5	
Imaging modality	PET-CT	15	0.79
	PET-MR	4	0.21
SUVmax	All: number	19	1
	Range	1.898 - 21.269	
	Median	5.387	
	Mean	6.819	
	SUVmax low: number	9	0.47
	Range	1.898 - 5.084	
	Median	2.756	
	Mean	3.207	
	SUVmax high: number	10	0.53
	Range	5.387 - 21.269	
	Median	9.035	
	Mean	10.07	

Table 1: Patients' characteristics of our EwS cohort.

3.2 Image Analysis

3.2.1 Image Feature Description

For each PET series containing one lesion, we calculated 3D image features (see Figure 1, the panel at the top left). To improve the comparability of radiogenomic studies, the use of standardized IFs is enormously important [31, 63, 96, 122]. One approach to achieve this is automatic feature extraction instead of subjective, manual extraction [123]. We used an established tool called PyRadiomics [96], which follows the image biomarker standardization initiative (IBSI) [95].

We calculated features from different classes, namely shape features, first-order statistics, and second-order statistics (glcm, gldm, glrlm, glszm, and ngtdm). Shape features include geometric characteristics of a lesion describing its 3-dimensional extension. First-order features focus on the histogram of intensity values. Second-order features additionally include the spatial relationship of intensities describing textural heterogeneity.

Additionally, different filters were applied to the raw PET data to obtain derived images. Dependent on filter type, certain patterns are emphasized such as contours, fine-grained or coarse-grained structures. On the derived images, first- and second-order features were calculated yielding features of higher order.

Suppl. Table 2 shows an overview of the IFs including a description of feature classes, filter types, and the number of features, respectively. For a more detailed description of the IFs and formulas, see [37, 95].

Finally, we obtained 1502 IFs describing each lesion.

3.2.2 Robustness of Image Features

Some PET IFs have higher variability than others [63]. Thus, it is important to examine their robustness and only use reproducible IFs for analysis.

3.2.2.1 Robustness Towards Imaging Modality

The first question was: how comparable are the IFs between PET-CT and PET-MR?

Since we have a mixed dataset with PETs from both modalities, this question is relevant for our study. Only features that show agreement between PET-CT and PET-MR can be used in a mixed analysis to draw meaningful conclusions. Moreover, they are valid across studies. Therefore, we examined the features for robustness with respect to imaging modality and identified stable features.

This analysis was based on 21 double series (21 PET-CT plus 21 PET-MR) from 15 sarcoma patients, including Ewing sarcoma (n=13 double series), osteosarcoma (n=5), rhabdomyosarcoma (n=1), and synovial sarcoma (n=2). The lesions were segmented using 40% of SUVmax. To determine agreement between PET-CT and PET-MR, we examined the intraclass correlation for absolute inter-rater agreement for each IF. The intraclass correlation coefficient (ICC) is a measure of inter-rater reliability, with a maximum value of 1 indicating perfect agreement.

The ICC distribution split by filter type can be seen in Suppl. Figure 1. The LoG filters and a wavelet filter show higher ICC values than the other filter types: there were significantly more IFs with ICC≥0.75 than expected (empirical q-value 0.00008 for log.sigma.4.0.mm.3D, 0 for log.sigma.5.0.mm.3D, and 0.0152 for wavelet.LHL). Wavelet filters focus on contours and de-noise the image by neglecting details. LoG filters are used for edge detection. Since these 3

filter types include IFs with high ICC, they seem to focus on structures that are similarly depicted in both modalities. Note that not all edge or contour detection filters contain IFs achieving high ICCs.

There are two guidelines for interpreting ICC values, which are summarized in Table 2. To define robust IFs, we used a cutoff of ICC≥0.75, since this includes the categories "excellent" and "good" according to Koo and Li [99] and "excellent" according to Cicchetti [100]. By this definition, 260/1502 features agreed between PET-CT and PET-MR.
SUVmax, which we used in the subsequent correlation analyses, had an ICC of 0.723 indicating "moderate" or "good" accordance between PET-CT and PET-MR.

(A)

Guideline of Cicchetti		
Inter-rater agreement	**Cutoff**	**Number of IFs (%)**
excellent	0.75 ≤ ICC ≤ 1.00	260 (17%)
good	0.60 ≤ ICC < 0.75	437 (29%)
fair	0.40 ≤ ICC < 0.60	491 (33%)
poor	ICC < 0.40	314 (21%)
		1502

(B)

Guideline of Koo and Li		
Inter-rater agreement	**Cutoff**	**Number of IFs (%)**
excellent	0.90 ≤ ICC ≤ 1.00	6 (0.4%)
good	0.75 ≤ ICC < 0.90	254 (17%)
moderate	0.50 ≤ ICC < 0.75	707 (47%)
poor	ICC < 0.50	535 (36%)
		1502

Table 2: Two alternative guidelines for ICC interpretation as excellent, good, fair/ moderate, or poor inter-rater agreement. *For each guideline, the cutoffs for the 4 categories are listed. The last column indicates the number of IFs that fall into each category in our analysis of robustness to imaging modality (PET-CT versus PET-MR). (A) The guideline according to Cicchetti. (B) The guideline according to Koo and Li.*

3.2.2.2 Robustness Towards Delineation Variability

Another aspect was: how sensitive are the IFs towards delineation variation of a lesion?
This is an important aspect as tumor segmentation is very subjective and difficult to reproduce. Therefore, it is important to identify IFs that are robust to changes of tumor boundaries.

This analysis was based on different segmentations with varying degrees of strictness. For each lesion, 3 thresholds were used to define tumor boundaries: 20%, 30%, and 40% of SUVmax. The values of each IF were compared between these 3 segmentations and evaluated using the ICC for consistency. This was examined separately for each imaging modality. We used PET-CT series (n=106 sarcoma patients, referred to as analysis a) and PET-MR series (n=54, referred to as analysis b). With this, we aimed to identify IFs that were stable in PET-CT and in PET-MR, respectively. In addition, we examined a mixed dataset of all available PET series (n=160 PET-CT or PET-MR, referred to as analysis c).

We examined which feature classes and which filter types were least affected when tumor boundaries varied. The ICC distribution split by filter type is shown for analysis a on PET-CT series in Suppl. Figure 2A. For all filter types, ICC distributions were in a similar range. This shows that the filter type had minor impact on the delineation sensitivity. Median ICC was over 0.8 for all filter types, around 0.9 for all filters except the logarithm filter, indicating that many features had high consistency independent of variation in tumor boundaries.

The ICC distribution split by feature class showed slightly more differences between classes (Suppl. Figure 2B) but consistently achieved high ICC values as well (median ICC>0.85).
Note that analysis b on PET-MR and analysis c on the mixed dataset showed the same pattern as analysis a on PET-CT.

To define IFs robust to delineation variation, we used a cutoff of ICC≥0.8. Since the 3 segmentations were based on the same raw PET data and the series differed only in their lesion boundaries, many IFs achieved higher ICC values in this analysis than in the analysis investigating the sensitivity to imaging modalities. 0.8 is commonly used when determining sensitivity to delineation variation according to the European Association of Nuclear Medicine (EANM) annual conference 2019, and is for example used by van Griethuysen et al. [96].
Applying ICC≥0.8, we obtained 1157/1502 stable IFs in PET-CT series (analysis a), 1071/1502 stable IFs in PET-MR series (analysis b), and 1141/1502 stable IFs in the mixed dataset (analysis c).

Next, we compared the results of analyses a, b, and c. The aim was to identify the IFs that were stable in the mixed dataset as well as in the PET-CT and PET-MR series, respectively. The sets of stable IFs from analysis a, b, and c overlapped in 1009/1502 IFs (Suppl. Figure 3). Overall, the sets of stable features showed a large overlap: 1009/1157 = 87% for analysis a on PET-CT series, 1009/1071 = 94% for analysis b on PET-MR series, and 1009/1141 = 88% for analysis c on mixed PET series. This suggests that the use of a mixed dataset, a PET-CT-only or PET-MR-only dataset had minor effect on determining delineation-insensitive IFs.
We examined whether any feature class or filter type was particularly prevalent among the 1009 robust IFs. No filter type was enriched among the robust IFs (empirical q-values for all filter types >0.05). In terms of feature classes, first-order features (empirical q-value=0), glcm features (empirical q-value =0.015), and glrlm features (empirical q-value=0) were enriched. First-order features are intensity histogram features. Glcm and glrlm features assess the spatial distribution of SUVs by analyzing the neighborhood of a voxel (glcm) or by analyzing stretches of the same SUV value in an image (glrlm). These 3 feature classes are particularly robust to variation of tumor boundaries.
Note that SUVmax is invariant to delineation variability (ICC=1 in each analysis).

3.2.2.3 Combination Into a Set of Robust Image Features

After analyzing the IFs for their robustness to imaging modality and delineation, we compared and combined the results. The goal was to determine those IFs satisfying both properties.
There are four times more features that are robust to delineation than to modality (1009 versus 260 IFs). The overlap comprises 184 IFs (Figure 3), which is 18% (184/1009) of the delineation-robust IFs, and 71% (184/260) of the modality-robust IFs. Thus, if a feature is robust to modality, it is likely to be robust to delineation as well.
In summary, these 184 IFs are independent of tumor boundaries and imaging modality (PET-CT or PET-MR) and thus suitable for further analysis.

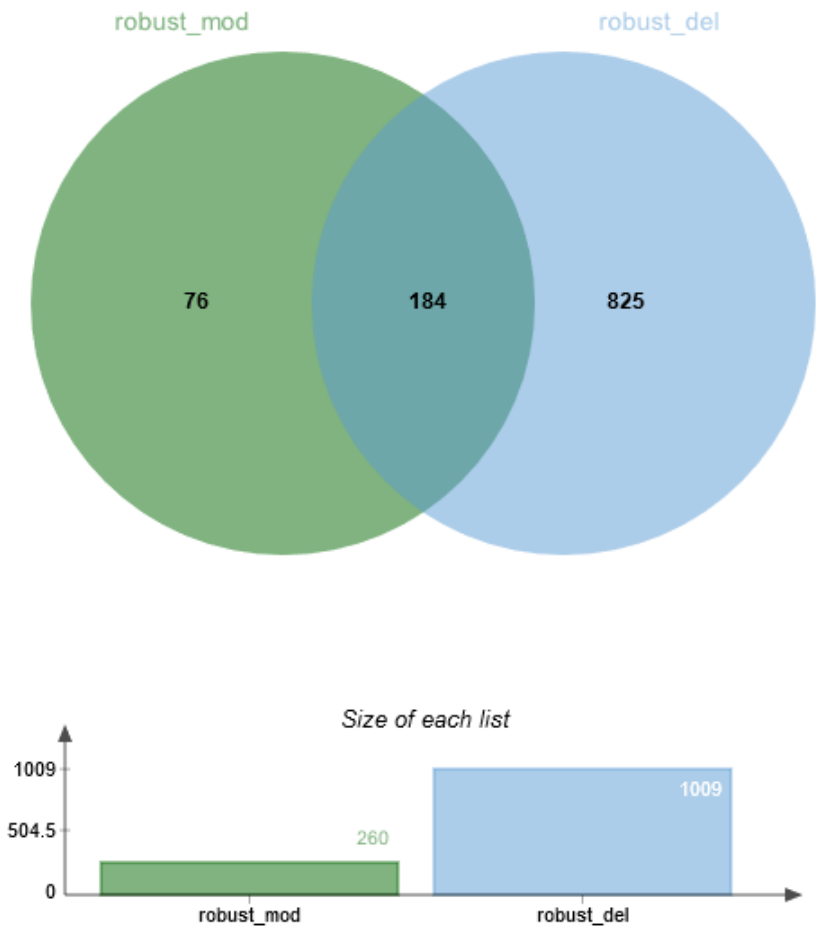


Figure 3: Venn diagram depicting the IFs robust to imaging modality (robust_mod, green) and delineation (robust_del, blue).*The overlap of both lists comprises 184 IFs. Below, the barplot shows the number of IFs in each list: 260 IFs are robust to modality, 1009 are robust to delineation.*

3.2.3 Pairwise Correlation and Redundancy of Image Features

Another aspect when examining the IFs is their redundancy. We wanted to know whether the IFs represent the same or different characteristics of a lesion, which provides an overview of their information content. Moreover, one can reduce redundancy in the dataset by focusing on complementary IFs.

3.2.3.1 Pairwise Correlation

For this analysis, we used all PET series of sarcoma patients (PET-CT and PET-MR, n=160) based on 40% SUVmax segmentation. To determine redundancy, pairwise Pearson correlation was calculated for all 1502 IFs. The resulting distribution of correlation coefficients (Pearson's r) is shown in Suppl. Figure 4. In general, there were more positive correlations than negative ones. Most of the correlations were close to 0, indicating complementary information content of the IFs. However, there was also a portion of IFs with high pairwise correlation between 0.8 to 1.

Next, we examined whether the IFs within one feature class were more strongly correlated among each other than with the IFs of other classes. In this case, one would expect some regular pattern in the correlations respecting the feature classes.

We examined all 107 IFs on the original image using a correlogram (Suppl. Figure 5). Visual inspection showed that the shape features were clearly separated from the other feature classes: they were only correlated among themselves. Note that shape features are the only feature class computed on the mask files without including the SUV values of the voxels. This fundamental difference could explain why the shape features were complementary to the other feature classes, capturing different properties of a lesion.
Apart from that, there were some blocks of correlated features (e.g., the first-order features), but there was no obvious pattern reflecting feature classes. Thus, it seems that intraclass correlations are in general not stronger than interclass correlations, especially for second-order IFs.

Since shape and first-order features were strongly correlated among themselves, we examined these feature classes in more detail (Figure 4).
For the shape features, the red square in the correlogram corresponds to a large cluster of 10 strongly correlated features: MinorAxisLength, Maximum2DDiameterSlice, Maximum3DDiameter, Maximum2DDiameterColumn, Maximum2DDiameterRow, MajorAxisLength, SurfaceArea, LeastAxisLength, VoxelVolume, and MeshVolume. All these IFs describe the tumor's dimensions, which explains their high redundancy. Sphericity and SurfaceVolumeRatio were (slightly) negatively correlated with these IFs. Flatness and Elongation were correlated with each other but not with the other shape features.
Among the first-order features, a group of 13 strongly correlated IFs stood out: Entropy, RobustMeanAbsoluteDeviation, InterquartileRange, MeanAbsoluteDeviation, Mean, 90Percentile, Median, RootMeanSquared, Maximum, 10Percentile, Minimum, Range, and Variance. TotalEnergy and Energy were complementary to these 13 IFs. The 3 IFs Skewness, Kurtosis, and Uniformity were redundant among each other but complementary to the remaining first-order features.

3.2.3.2 Relation of Lesion Volume and SUVmax

In this analysis, two IFs are of particular interest, namely lesion volume (original_shape_VoxelVolume) and SUVmax (original_firstorder_Maximum). It has been observed before that there is a relationship: the larger the volume, the higher the SUVmax in different entities, e.g., lung cancer [124-126] or breast cancer (moderate correlation) [127, 128]. Furthermore, these important features can be expected to affect other features as well [63]. Thus, the question arises whether some IFs are just surrogates for volume or SUVmax in our dataset.

Volume (VoxelVolume) is a shape feature, belonging to the cluster of 10 redundant shape features (Figure 4A). Likewise, SUVmax (Maximum) as a first-order feature belonged to the large cluster of redundant first-order features (Figure 4B). Thus, a large fraction within their feature class was a surrogate for volume and SUVmax, respectively.
Regarding all IFs, a low proportion correlated strongly with lesion volume, as only 48/1502 IFs achieved abs(r)>0.9 and 77/1502 achieved abs(r)>0.8. This was probably due to the fundamental differences between shape features and the other feature classes. In conclusion, only a low number of IFs except shape features were indirect surrogates for lesion volume.
For SUVmax, a total of 157/1502 achieved abs(r)>0.9 and 415/1502 achieved abs(r)>0.8. Thus, more IFs correlated strongly with SUVmax than with lesion volume. This was presumably influenced by the fact that all feature classes except shape features were calculated for each filter type. For example, the feature firstorder_maximum was computed 16 times, once on

the original image and on 15 derived images. As expected, these duplicated features correlated strongly with each other: 12/16 firstorder_maximum had abs(r)>0.9 to original_firstorder_maximum.

Next, we examined the relation of volume and SUVmax to each other.
Based on all 160 PET series, we obtained no correlation (r=0.07).
In addition, we focused on the cohort of 19 EwS samples we used for linear regression analysis of gene expression and SUVmax. Again, no correlation was found (r=-0.02, p=0.95; Figure 5). 17/19 lesions in this cohort had a size of 0.8 cm^3 to 109 cm^3. 2 lesions were much larger, namely 509 cm^3 and 725 cm^3. When omitting these 2 very large lesions and testing the correlation on lesions with maximal 110 cm^3, a moderate positive correlation between volume and SUVmax was detected (r=0.57, p=0.02).
In short, SUVmax only correlates with volume in a subgroup of our cohort.

3.2.3.3 Non-Redundant Image Features

Finally, we aimed to obtain a set of non-redundant IFs by applying a correlation cutoff of 0.9 [104]. If two IFs had abs(r)>0.9, the IF with higher mean absolute correlation to all IFs in the dataset was removed. This resulted in 321/1502 non-redundant IFs. See Suppl. Table 3 for a listing, split up by feature class and filter type.
Percentage-wise, many of the glrlm and first-order features were redundant, as only 7% and 16% of the IFs remained, respectively. In contrast, many of the glszm (36.7%) and shape (35.7%) features were non-redundant. The latter may be unexpected at first glance as it was noted before that the shape features were highly correlated among each other. However, the non-redundant shape features were Elongation, Flatness, Sphericity and SurfaceVolumeRatio, which showed low correlation with all shape features (Figure 4A). As a representative for the large group of strongly correlated features, only MajorAxisLength was retained. Thus, the result was as expected.
In terms of filter type, many wavelet-LLL (12.9%), square root (14%), and original features (15%) were redundant. In contrast, many IFs of gradient (29%) and wavelet-HHH filters (29%) were non-redundant.

In summary, the 321 non-redundant IFs provide complementary information about tumor shape and structure in PET imaging.

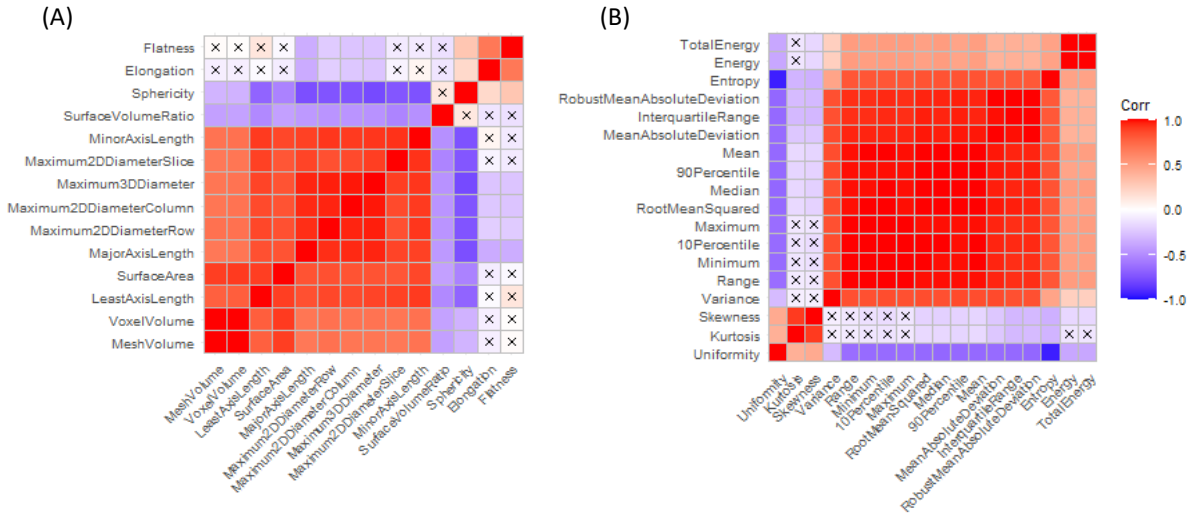
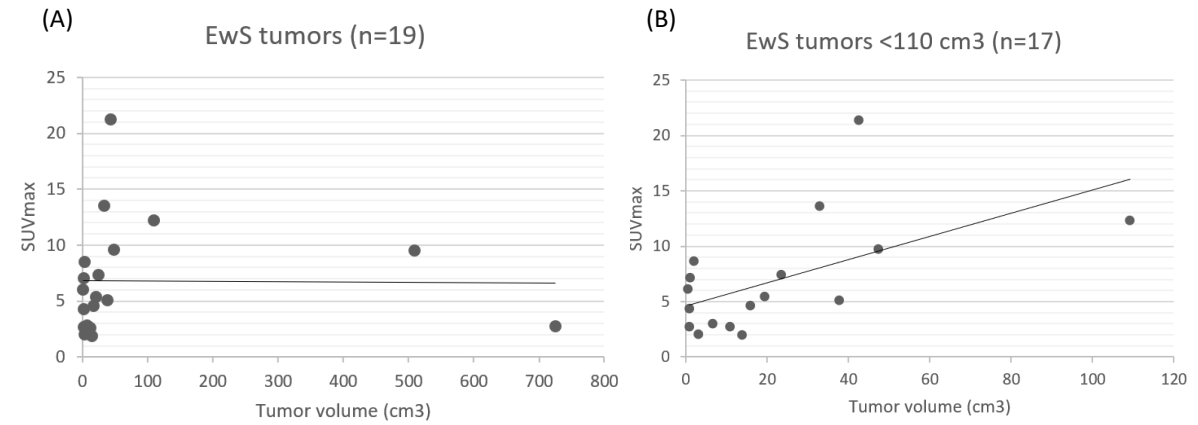


Figure 4: Correlogram depicting intraclass pairwise correlation of (A) all 14 shape features and (B) all 18 first-order features (original images, 160 PET series). *IFs are clustered hierarchically (complete linkage). A cross indicates a non-significant correlation.*



(C)

	EwS tumors (n=19)	EwS tumors <110 cm³ (n=17)
Regression line slope	-0.0004	0.105
Pearson's correlation r (95% CI)	-0.02 (-0.47; 0.44)	0.57 (0.12; 0.82)
p-value	0.95	0.02

Figure 5: Correlation of SUVmax (original_firstorder_Maximum) and tumor volume (original_shape_VoxelVolume). *(A) Analyzing all 19 samples in the cohort used for regression analysis, SUVmax does not correlate with tumor volume. 17/19 tumors have a volume of 0.8 cm³ to 109 cm³, 2/19 large tumors stand out with 509 cm³ and 725 cm³. (B) Focusing on the group of tumors with a maximal volume of 110 cm³ (n=17, excluding the large tumors), there is a mediate correlation of SUVmax and tumor volume. (C) Table providing details of the regression analysis.*

3.3 Correlation of Radiomic and Clinical Data

3.3.1 Image Based Diagnosis Prediction

We wanted to know whether it is possible to distinguish different sarcoma entities or molecular subtypes in EwS on the basis of PET radiomic data (see Figure 1, the panel in the middle on the left). Do the entities or gene fusion types have different structure – e.g., in terms of homogeneous or heterogeneous functional regions – that is reflected in the PET and thus in the IFs?

Background for this question is a study of Macpherson et al. [61], who compared the SUVmax values of 17 entities of bone and soft tissue sarcoma (n=493). They found that high-grade sarcomas were associated with higher SUVmax values. A limitation of this study was that only SUVmax was examined. More IFs need to be analyzed to fully capture the functional heterogeneity in the tumor.

In addition, we asked whether different gene fusions in EwS had a characteristic structure of PET signals that could be measured in the IFs. A study examining fusion characteristics was performed by Soffer et al. [129]. They found that EWS translocation positive tumors had significantly lower SUVmax than EWS translocation negative Ewing-like tumors. We complement this study by including more IFs in the analysis and focusing on distinguishing fusion types.

To answer these questions, we applied a machine learning approach: random forest classification with repeated cross-validation on a training set and subsequent testing of the model on a validation set. We predicted either entity or fusion type based on 184 PET IFs that were robust regarding modality and delineation.

For entity prediction, our dataset included 138 PET series: 75 EwS, 35 osteosarcoma, 15 rhabdomyosarcoma, 10 synovial sarcoma, and 3 others. Since the groups varied in size, we arranged them into EwS (n=75) and non-EwS (n=63). Table 3A provides an overview of the composition of the training and validation sets. Performance in cross-validation was slightly better than random guessing (ROC AUC=0.647, F1=0.605). However, performance on the validation set was not superior to random prediction as the ROC curve was close to the diagonal line (ROC AUC=0.539, F1=0.686; Figure 6A). Accuracy was 59% (16/27 correct predictions in the validation set). In summary, the PET image features did not contribute much more information than random prediction to distinguish EwS versus non-EwS (Cohen's kappa=0.139 on the validation set).

For the question regarding gene fusions, our dataset included 67 PET series of EwS patients. 48/67 were positive for EWS-FLI1, and 19/67 were negative for EWS-FLI1, which included tumors diagnosed as EwS but with a negative test for the EWS-FLI1 fusion indicating other translocation partners of EWS. For example, EWS-ERG or EWS-ETV1 translocation were approved in some samples. Table 3B provides an overview of the composition of the training and validation set. Performance of the model in cross-validation was comparable to random prediction (ROC AUC=0.434, F1=0.356). Performance of the validation set was better as the ROC curve was superior to the diagonal line (ROC AUC=0.704, F1=0.571; Figure 6B). However, this could be an artifact due to the coarse steps of the curve as only 3 samples were EWS-FLI1 negative in the validation set. Accuracy was 50% (6/12 correct predictions in the validation set). Taken together, the performance was comparable to random prediction (Cohen's kappa=0.077 on the validation set). It was not possible to distinguish fusion types in EwS based on the PET IFs in our dataset.

In conclusion, random forest modeling did not detect any characteristic differences in PET signal structures in EWS-FLI1 positive compared with EWS-FLI1 negative EwS tumors. That different fusion types have similar structuring in PET supports that we can analyze different fusion types together in subsequent analysis.
In terms of entity prediction, a reliable discrimination of EwS versus non-EwS was not possible based on the information obtained from PET images. This may indicate similar structures of PET signals in all entities; and if there were differences, they could not be detected reliably in our dataset. This can also be due to the fact that we had to aggregate different sarcoma entities (osteosarcoma, rhabdomyosarcoma, synovial sarcoma, and other entities) into one non-EwS group comparing them to EwS. Further studies comparing single entities may provide other results about PET signal structures in sarcoma entities.

(A)

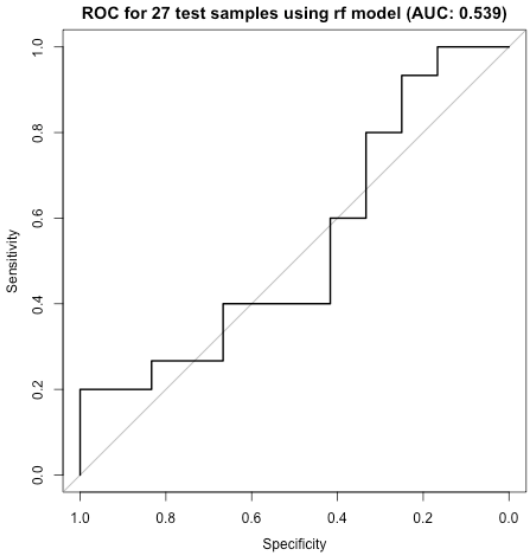
Group for modeling	**Entity**	**Dataset**	**Training**	**Validation**
EwS	EwS	75	60	15
Non-EwS	Sum	63	51	12
	Osteosarcoma	35	29	6
	Rhabdomyosarcoma	15	11	4
	Synovial sarcoma	10	8	2
	Other entity	3	3	0

(B)

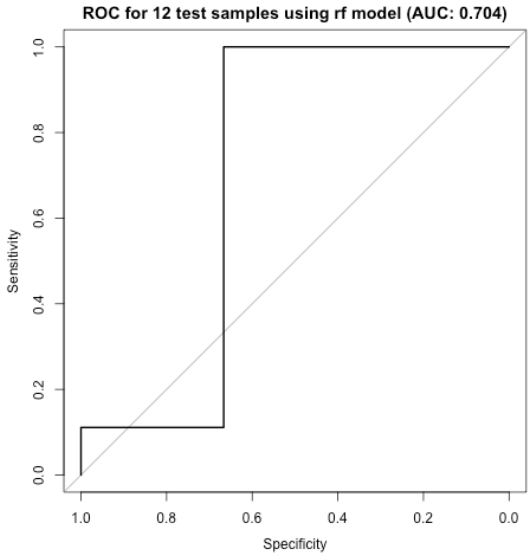
Group for modeling: EwS molecular subtype (fusion)	**Dataset**	**Training**	**Validation**
EWS-FLI1 positive	48	39	9
EWS-FLI1 negative	19	16	3

Table 3: Dataset composition for random forest prediction of (A) entity (EwS vs. non-EwS) or (B) EwS fusion type (EWS-FLI1 positive vs. negative).*For both questions, the whole dataset was split into a training cohort (80%) and a validation cohort (20%) before modeling. The training cohort – after downsampling to obtain equal group sizes – was used for parameter tuning in 5-fold repeated cross-validation and model generation. The final model that achieved the highest ROC AUC in cross-validation was tested on the validation cohort afterwards.*

(A) EwS versus non-EwS



(B) EWS-FLI1 positive versus negative



(C)

		Reference	
	Entity	EwS	Non-EwS
Prediction	EwS	12	8
	Non-EwS	3	4
	Sum	15	12
	Accuracy	0.593	

	Reference	
EWS-FLI1	positive	negative
positive	4	1
negative	5	2
Sum	9	3
Accuracy	0.5	

Figure 6: ROC curve indicating performance of final random forest model on validation set.*(A) Predict entity EwS versus non-EwS. The ROC curve is close to the grey diagonal line indicating random prediction (ROC AUC 0.539). (B) Predict EwS fusion type EWS-FLI1 positive versus negative. The ROC curve is superior to random prediction (ROC AUC 0.704). However, this can be an artifact due to the coarse steps of the curve as only 3 samples are EWS-FLI1 negative in the validation set. (C) Confusion matrices for both prediction tasks. 16/27 and 6/12 samples are predicted correctly, respectively, giving accuracies of 59% and 50%. Considering the ROC curves and the accuracies, prediction is not superior to random prediction for both tasks.*

3.3.2 Equal SUVmax Distribution Regarding Clinical Variables

We explored the correlation of SUVmax with clinical variables in our cohort (see Figure 1, the panel in the middle on the right). Potential correlations would render the clinical variables confounding factors and introduce bias in subsequent analyses when we correlated gene expression with SUVmax. To rule out confounding factors, we related all clinical variables in our dataset with SUVmax: sex (male vs. female), disease state (primary disease vs. relapse), sample type (tumor vs. metastasis), therapy (untreated vs. treated), and age (<=15 years vs. >15 years).

Each variable split the patients into two groups, which we compared for equal distribution of SUVmax values applying a Welch two-sample t test. We found that SUVmax was equally distributed with respect to all tested variables (Figure 7).

This demonstrates that there are no confounding factors among the clinical variables in our dataset. Thus, when we correlate gene expression with SUVmax, we mainly correlate with metabolic activity in the form of glucose uptake.

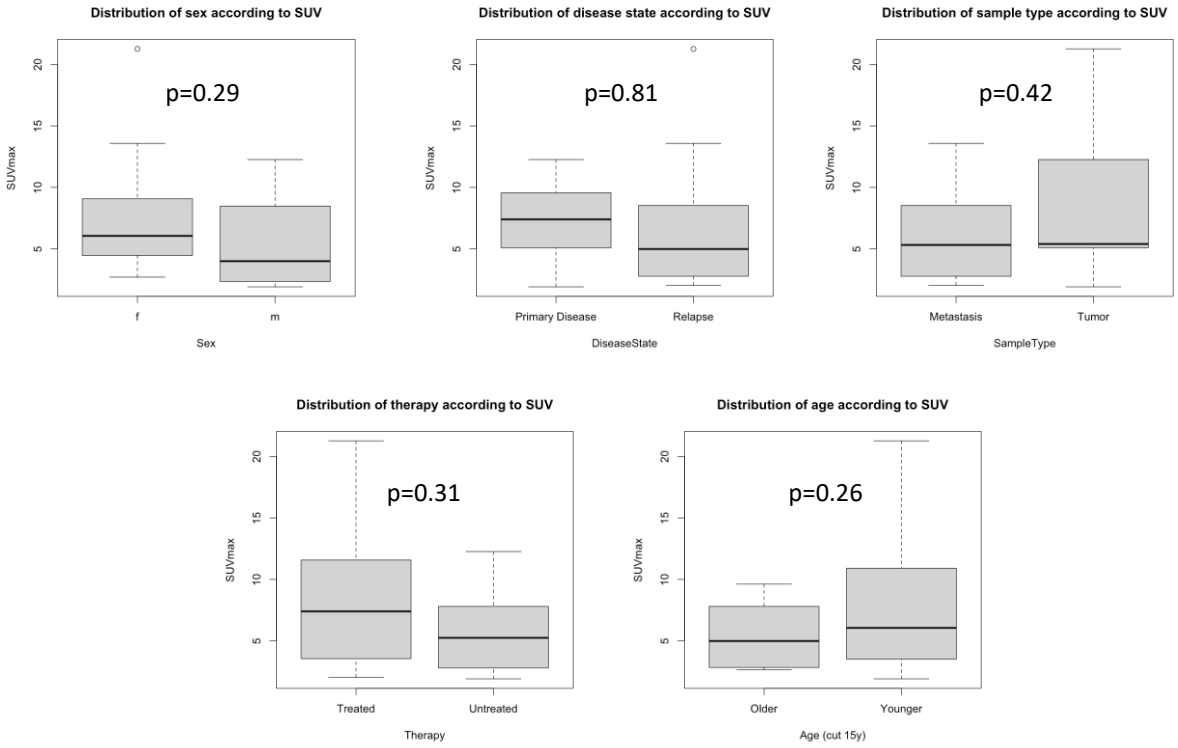


Figure 7: Boxplots showing equal SUVmax distribution with regard to clinical variables: *sex (female or male), disease state (primary disease or relapse), sample type (tumor or metastasis), therapy (treated or untreated), and age (≤15 years or >15 years). P-values from the Welch two-sample t-test.*

3.3.3 Survival Analysis

We investigated the impact of clinical variables and SUVmax on survival in our cohort (see Figure 1, the panel in the middle on the right). As measure for survival, overall survival starting from the day of initial diagnosis or relapse was used.

We tested the following variables for their correlation with overall survival: sex (male vs. female), disease state (primary disease vs. relapse), age (<=15 years vs. >15 years), and binary SUV categories (low vs. high SUV split by median SUVmax). To this end, in a first step, we performed a univariate Kaplan-Meier analysis for each variable. We found no significant correlations with overall survival (Figure 8A). The confidence intervals (CIs) of the curves were very wide and overlapped strongly, which was due to our small sample size.

This univariate analysis indicated that no variable had a strong, significant correlation with overall survival on its own.

As a next step, we combined several variables in a multivariate Cox proportional hazards model. For this, we used those variables in our dataset that have an impact on survival according to the literature, namely disease state, SUVmax, and age [130, 131]. However, since age did not show any correlation with overall survival in our dataset in univariate analysis (p=0.65), we did not include age into the model. Disease state and SUVmax showed some tendency in univariate analysis (p=0.087 and p=0.14, respectively). Thus, disease state and SUVmax remained for the multivariate analysis.

There were two ways to use SUV in the model: First, we compared two groups of tumors (with high vs. low SUVmax values split by median SUVmax), and second, we tested continuous SUVmax values.

The first model based on SUV groups achieved an overall p-value of 0.06 in the likelihood ratio test. Thus, there was a trend that relapse (p=0.10, Hazard Ratio HR [95 CI]=3.7 [0.79; 16.9]) and high SUV (p=0.12, HR [95 CI]=2.6 [0.78; 8.9]) were associated with higher risk of death (Figure 8B).

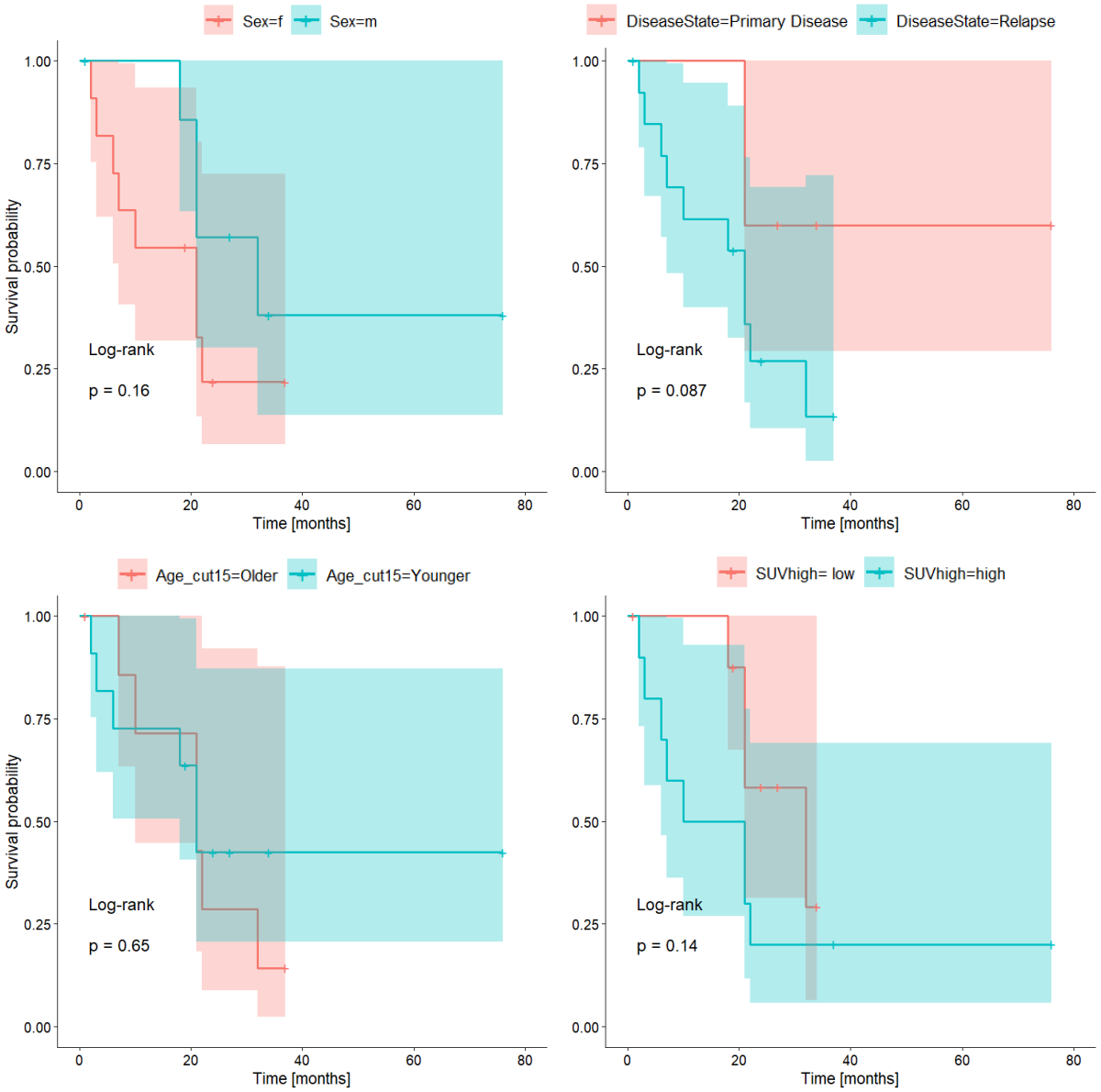
The second model based on continuous SUVmax obtained an overall p-value of 0.02. In combination, both variables showed significant correlation with overall survival (Figure 8C): higher SUVmax (p=0.02, HR [95 CI]=1.2 [1.0; 1.3]) and relapse (p=0.05, HR [95 CI]=5.0 [1.0; 24.9]) were associated with shorter overall survival.

In conclusion, the multivariate model obtained significant correlations when using the exact fine grading of SUVmax values instead of the coarse classification of tumors into "high SUV" versus "low SUV". Altogether, there was a trend that SUVmax together with disease state (primary disease vs. relapse) were prognostic factors in our cohort.

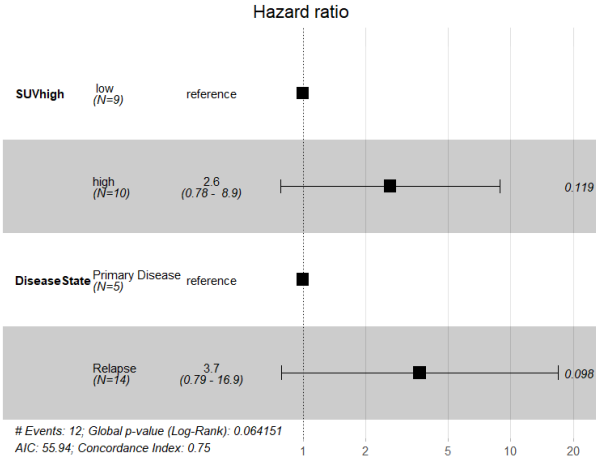
Figure 8: Survival analysis.*(A) Kaplan-Meier curves for univariate analyses. There is no significant difference in overall survival with regard to sex (male vs. female), disease state (primary disease vs. relapse), age (≤15 years vs. >15 years), or SUV categories (low SUV vs. high SUV). (B+C) Forest plots for multivariate Cox proportional hazards models including SUV and disease state (primary disease vs. relapse). (B) Model using binary SUV categories (samples split into low SUV vs. high SUV by median SUVmax) indicates a trend that high SUV and relapse decrease overall survival time (global p-value=0.06 in log-rank test). (C) Model using continuous SUVmax values shows that higher SUVmax and relapse significantly decrease overall survival time (global p-value=0.02 in log-rank test). For continuous SUVmax, the hazard ratio is 1.2 per increase of 1 SUV unit.*

(continued →)

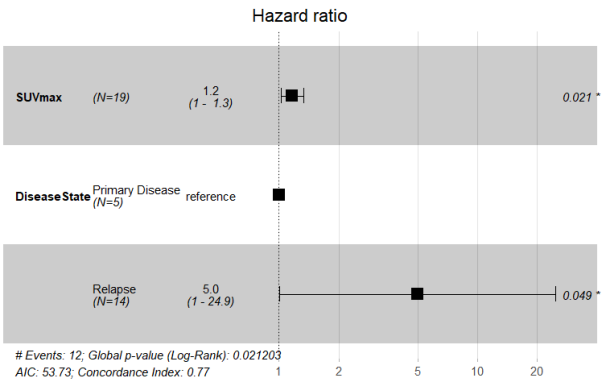
(A) Kaplan-Meier



(B) Cox model using disease state and binary SUV categories



(C) Cox model using disease state and continuous SUVmax values



3.3.4 No Correlation of SUVmax With Proliferation Associated Gene Expression

Next, we examined whether glucose uptake was merely a measure of proliferation. Hence, we tested for association between SUVmax and proliferation rate of cells in means of expression of 4 genes associated with proliferation: PCNA, MKI67, TOP2A, and MCM2. We correlated the expression of these genes with SUVmax using linear regression, which showed weak correlations (Table 4). For all 4 genes, we obtained positive correlations that were small in terms of effect size, as the slope of the regression line was close to 0 (0.03 to 0.07). Moreover, the samples varied a lot around the regression lines, as the variance of gene expression explained by SUVmax (r^2) was only 0.03 to 0.23. PCNA expression correlated strongest with SUVmax, with a slope of 0.07 and r^2 of 0.23, which corresponds to a doubling of gene expression over 14 SUV units. Hence, even the strongest expression change is unlikely to be of biological relevance for processes in a cell.

All in all, these results indicate that there was no relation of SUVmax with markers of proliferation in our dataset. This implies that proliferation rate was not a key factor in our data.

	PCNA	**MKI67**	**TOP2A**	**MCM2**
Regression line slope	0.071	0.037	0.027	0.027
Gene expression doubling over x SUV units	14.18	27.03	36.63	37.31
r^2	0.234	0.053	0.027	0.118

Table 4: Linear regression analysis of SUVmax and expression of 4 proliferation markers (PCNA, MKI67, TOP2A, MCM2) indicates no correlation of SUVmax and proliferation rate. *Details of the regression analysis are provided: the regression line slopes meaning the expression change over 1 SUV unit; and r^2 (coefficient of determination of the Pearson correlation) indicating the proportion of variance in expression levels that is explained by SUVmax.*

3.4 Gene Expression Analysis

3.4.1 Preprocessing

After analyzing our cohort with regard to radiomic and clinical data, the next step in our workflow was the gene expression analysis (see Figure 1, the panel at the top right). We first performed standard preprocessing steps consisting of RMA with background subtraction, quantile normalization, and summarization of probe sets to genes. For the summarization, we used the Brainarray annotation, which supplies up-to-date definitions of probe sets [81-83]. After summarization we obtained 20726 genes, which were further annotated with gene symbols, yielding 20722 genes. In addition, we performed quality control to ensure that there were no outliers in the dataset and that all samples could be included.
Finally, 20722 genes remained in our dataset for further analysis.

3.4.2 Gene Filtering

A problem in large-scale analyses is the high dimensionality of the data, e.g. the expression of 20722 genes [123, 132]. To overcome this, we performed two filtering steps to reduce the number of genes for further analyses.
In the first filtering step, we excluded genes with low expression as microarray chips do not perform accurately in this case. We kept genes with an average expression above 10. Altogether, 198 genes were removed, leaving 20524 for further analysis.
In the second filtering step, we focused on genes that may be prognostic in EwS. To obtain such survival-related genes, we used a machine learning approach on 3 external public EwS datasets: GSE63155, GSE17618, GSE63156 (Figure 9). We predicted long-term versus short-term survival based on gene expression separately for each external dataset by using random forest classification models. The models achieved an average performance of ROC AUC of 0.67 in GSE63156, 0.75 in GSE63155, and 0.87 in GSE17618 in repeated cross-validation. Furthermore, the models identified a group of genes that were predictive in each external dataset. These 3 gene sets overlapped in 1491 genes. Since these 1491 genes were related to survival in 3 independent datasets, we refer to them as "potential survival genes" for EwS. Functional annotation of these genes yielded phosphoprotein, alternative splicing, polymorphism, acetylation, cytoplasm, cell division, cell cycle, Golgi apparatus, DNA replication, disease mutation, mitosis, cell junction, and endoplasmic reticulum (Figure 10). The “potential survival genes” were kept in our dataset for further analysis.

After both filtering steps, we obtained a total of 1376 genes for further analysis. We reduced the dimensionality of our gene expression data by focusing on genes that contain information about survival in external, independent EwS cohorts.

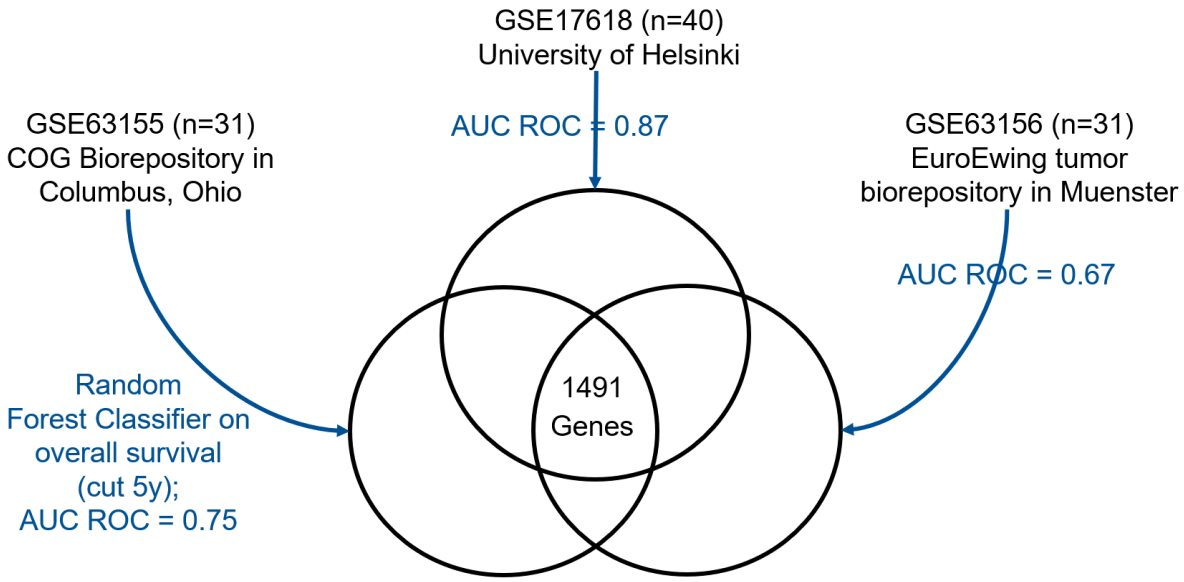


Figure 9: Diagram of machine learning analysis in order to obtain "potential survival genes" in EwS.*Random forest classifiers are applied to 3 public datasets (GSE63155, GSE17618, GSE63156) in 10-fold cross-validation, obtaining ROC AUC of 0.67 to 0.87. These models yield genes that are predictive for survival for each dataset. The intersection of these 3 gene sets contains 1491 genes, which we consider as "potential survival genes" in EwS.*

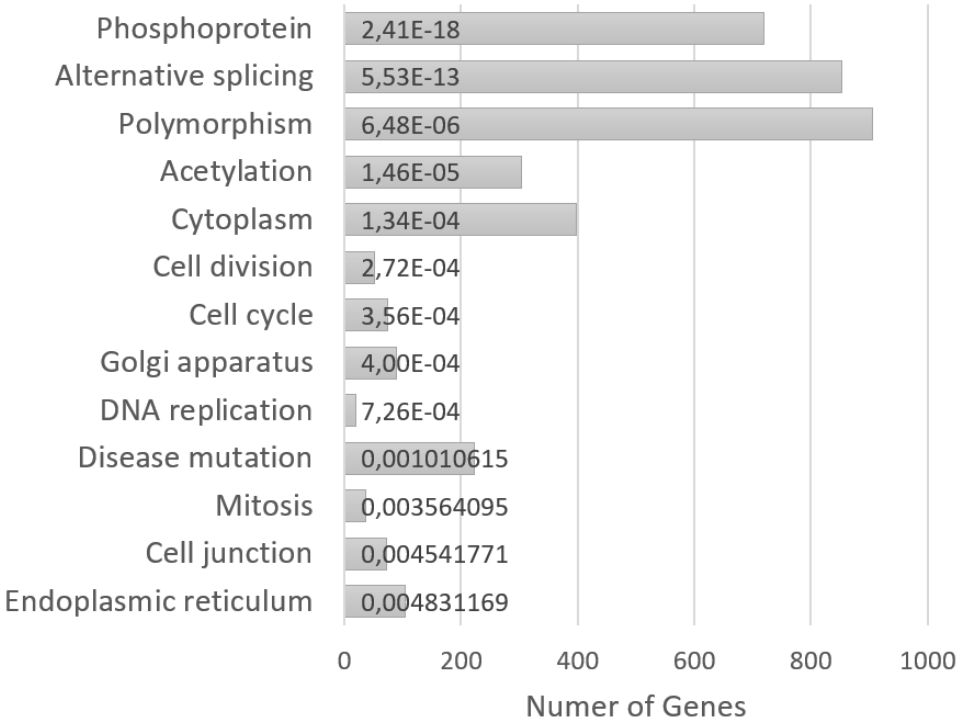


Figure 10: DAVID functional annotation of "potential survival genes" obtained from machine learning analysis on 3 external datasets.*13 terms in category UP-KEYWORDS obtained significant p-values (Benjamini adjusted p<0.01). For each term, the number of annotated genes among the "potential survival genes" is indicated by the bar and the adjusted p-value is given by the number in the bar.*

3.5 Correlation Analysis of Radiomic and Transcriptomic Data

3.5.1 Outcome Prediction using Radiomic and Transcriptomic Data

When integrating image and expression data into one analysis, we first aimed at predicting outcome (see Figure 1, the panel at the bottom left). As some PET image features and expression of genes are predictive for survival, we examined whether overall survival could be inferred in our cohort. This shows, how much information the PET image features and gene expression values in our dataset contain about patient survival, respectively. Next, we asked whether integrating expression and image data improves outcome prediction compared with expression or image data alone. This indicates whether the information from image and tissue is redundant or complementary.

To predict overall survival (binary classification, cut two years), we applied random forest machine learning modeling with leave-one-out cross-validation. We computed 3 models: model 1 using expression data (1774 genes), model 2 using image data (112 IFs, robust with respect to delineation and modality), and model 3 using expression and image data (1886 features, 1774 genes + 112 IFs). We compared the performance of the 3 models in predicting long-term overall survival (n=6 "long survival") versus short-term overall survival (n=11 "short survival") in 17 EwS patients (Suppl. Table 1).
The performance of the models was worse than baseline (Figure 11). Model 1 based on expression data achieved an area under the precision-recall curve (PR AUC)=0.47, model 2 based on PET data achieved PR AUC=0.50, and model 3 based on combined data achieved PR AUC=0.47. Matthew's correlation coefficient (MCC) was negative and close to zero, implying random prediction (MCC=-0.19, MCC=-0.12, and MCC=-0.12 based on an optimal cutoff for discrimination). These results indicate no information gain regarding overall survival based on expression data, image data, or the combination of both.
Possible reasons for this could be the limited sample size or the heterogeneous composition of the cohort. Examining the predictions of one model, e.g. the third model based on expression and image data (Table 5), gives more insight. 4/6 "long survival" patients scored very high for "short survival" in the model, so they were unambiguous false predictions. One explanation is as follows: 1/17 samples was removed before model generation, respectively (leave-one-out cross-validation). The model was computed on the remaining 16 samples (training dataset), which corresponds to a search for patterns (features) that distinguish the "long survival" patients from the "short survival" patients in training. The model was then applied to the removed test sample to check how well the learned pattern fitted a new patient. In our models, a characteristic pattern for "short survival" seemed to be identified in the training set that fitted 4/6 of the "long survival" patients very well, resulting in their high scoring for "short survival". This implies that in a cross-validation round with a “long survival” patient as test sample, the remaining 5 “long survival” patients in the training dataset did not exhibit a recognizable pattern of feature values, which could be learned by the classifier to correctly identify a new “long survival” patient. The resulting false scoring causes poor ranking of patients (measured in PR AUC) and poor classification rate (measured in MCC).

In conclusion, there is no consistent pattern of gene expression and/or IFs in our dataset that could be learned to predict outcome superior to random guessing. This analysis needs replication with a larger dataset to identify and model predictive patterns that may exist.

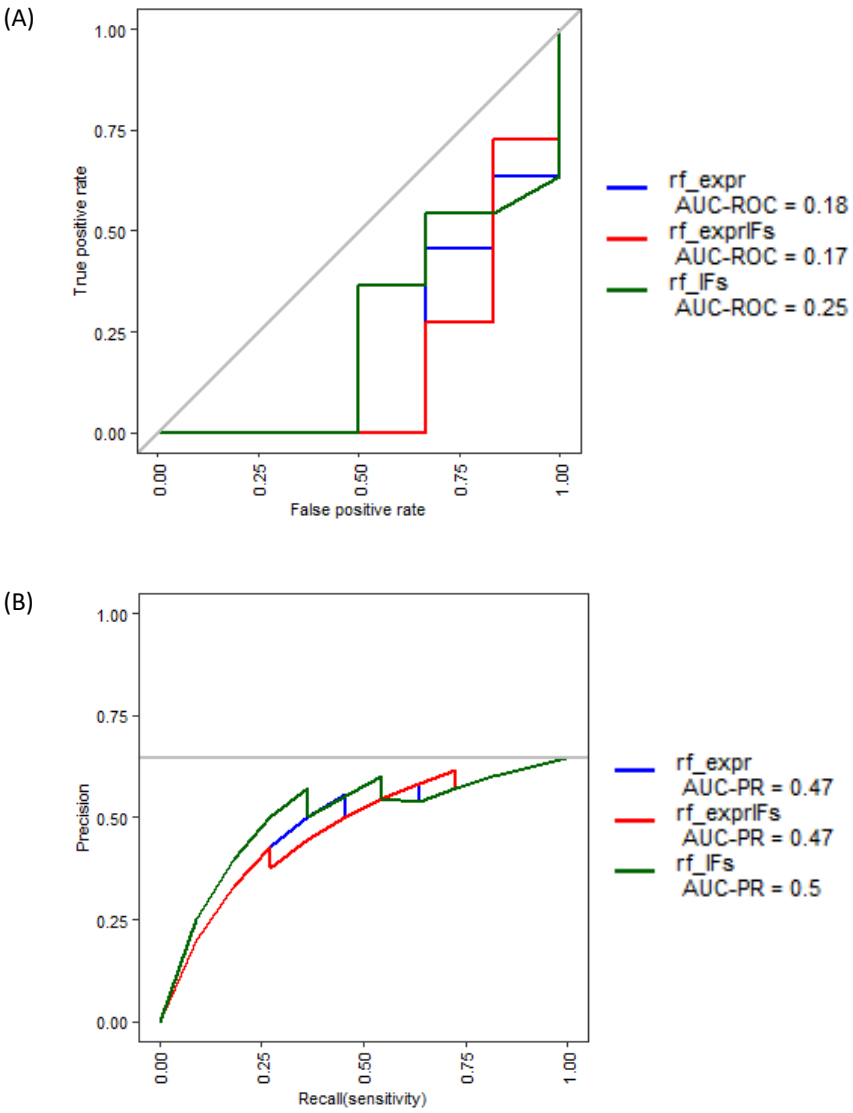


Figure 11: Performance comparison of models for survival time prediction.*The 3 models are based on expression data ("expr", blue), image data ("IFs", green), or expression plus image data ("exprIFs", red). The grey line indicates the baseline model: baseline ROC AUC=0.5 and baseline PR AUC=0.65. (A) ROC curve with ROC AUC. (B) Precision Recall (PR) curve with PR AUC.*

(A)

Patient's real survival	Predicted score for „short survival“	Predicted score for „long survival“	Prediction (default cutoff ≥0.5 for „short survival“)	Prediction (optimal cutoff ≥0.57 for „short survival“)
"short survival"	0.406	0.594	"long survival"	"long survival"
"short survival"	0.46	0.54	"long survival"	"long survival"
"short survival"	0.475	0.525	"long survival"	"long survival"
"long survival"	0.569	0.431	"short survival"	"long survival"
"short survival"	0.57	0.43	"short survival"	"short survival"
"short survival"	0.578	0.422	"short survival"	"short survival"
"short survival"	0.58	0.42	"short survival"	"short survival"
"short survival"	0.582	0.418	"short survival"	"short survival"
"short survival"	0.604	0.396	"short survival"	"short survival"
"long survival"	0.615	0.385	"short survival"	"short survival"
"short survival"	0.64	0.36	"short survival"	"short survival"
"short survival"	0.648	0.352	"short survival"	"short survival"
"short survival"	0.692	0.308	"short survival"	"short survival"
"long survival"	0.732	0.268	"short survival"	"short survival"
"long survival"	0.735	0.265	"short survival"	"short survival"
"long survival"	0.746	0.254	"short survival"	"short survival"
"long survival"	0.807	0.193	"short survival"	"short survival"

(B)

		Reference	
	Prediction (default cutoff ≥0.5 for „short survival“)	„short survival“	„long survival“
Prediction	„short survival“	8	6
	„long survival“	3	0
	Sum	11	6
	Accuracy	0.471	

	Reference	
Prediction (optimal cutoff ≥0.57 for „short survival“)	„short survival“	„long survival“
„short survival“	8	5
„long survival“	3	1
Sum	11	6
Accuracy	0.529	

Table 5: Survival time prediction ("long survival" vs. "short survival") of 17 EwS in leave-one-out cross-validation generated from the random forest model using expression plus image data (1886 features). *537/1886 features were selected in the model (509 genes and 28 IFs). (A) The patient's real survival time, and the predicted scores (probabilities) for "short survival" and "long survival" are listed. Based on the scores, a prediction was made using the default cutoff ≥0.5 for "short survival". This classified 3/17 as "long survival" patients and 14/17 as "short survival" patients (accuracy=0.47, Matthew's correlation coefficient MCC=-0.34). Based on MCC, an optimal cutoff for discrimination was identified: ≥0.57 for "short survival". This classified 4/17 as "long survival" patients and 13/17 as "short survival" patients (accuracy=0.53, MCC=-0.12). Samples are sorted by increasing score for "short survival". (B) Confusion matrices for default and optimized cutoff values.*

3.5.2 Correlation of SUVmax and Immune Cell Infiltration

The immune system plays an important role in cancer. Whether and which immune cells can infiltrate the tumor determines distinct immune responses, which are associated with outcome. For example, infiltration of CD8 T cells counteracts tumor progression [133], while infiltration of immunosuppressive macrophages correlates with poor patient survival [134, 135]. This was observed in several cancer types and in EwS [18]. Since both inflammation and PET are prognostic, we investigated whether there was an association in our cohort. Therefore, we correlated immune infiltration with SUVmax (see Figure 1, the panel at the bottom right, top row).

3.5.2.1 Immune Cell Profile In Ewing Sarcoma

The infiltrates of different immune cells were determined based on gene expression applying the deconvolution algorithm Cibersortx [109]. Cibersortx used the expression of 587 genes characteristic of immune cell types. Based on these cell type reference profiles, the absolute proportion of 22 immune cell types in bulk tumor tissue was estimated.

We analyzed all EwS samples with microarray data, which were 29 samples. Their immune cell profile was dominated by CD8 T cells (18% of total immune cell infiltrates), M0 macrophages (16%), monocytes (15%), and regulatory T cells (Tregs, 12%) (Suppl. Table 4). This profile differed from the profile described by Stahl et al. [18], who analyzed 197 EwS tumors using Cibersort [136], the predecessor of Cibersortx. This profile was dominated by M2 macrophages (about 25% of immune cell infiltrates) and M0 macrophages (about 13%).
In our dataset, there may be patients who were analyzed after allogeneic stem cell transplantation, which may well explain a T cell infiltrate. The clinical annotation is not complete but at least one patient was analyzed several months after stem cell transplantation, and we cannot exclude the possibility that there are more patients, or that some of them have received therapies that affect immune cell infiltration.

To further validate the immune cell profile in our dataset (Schwabing, n=29), we compared it to 4 public datasets of EwS tumors (Delattre GSE34620 [110], Savola GSE17618 [86], Dirksen GSE63157 [73], and Lawlor GSE68776 [111]) and a dataset of EwS cell lines (Savola GSE17618 cell lines [86]). Overall, two types of immune signatures arose (Table 6, Suppl. Figure 6). The datasets from Delattre and Savola showed an "M2 predominant" signature with high infiltration of M2 and M0 macrophages. In contrast, the datasets from Dirksen, Lawlor, and Schwabing showed high infiltration of Tregs, activated NK cells, CD8 cells, MO macrophages, and monocytes, which we summarized as a "T cell predominant" signature.
This raised the question of why there were different signatures and whether correlations with other factors might explain the differences among the datasets (cohort composition, treatment, affy platform; see dataset description in Table 6). In terms of tumor type, homogeneous cohorts comprising only primary tumors were "M2 predominant" (Delattre) or "T cell predominant" (Dirksen). Mixed cohorts of primary and relapsed tumors showed a "M2 predominant" (Savola) or "T cell predominant" signature (Schwabing). With respect to treatment, mixed cohorts of treated and untreated patients exhibited both signatures: "M2 predominant" in the Savola dataset and "T cell predominant" in the Schwabing dataset. Thus, these factors cannot explain the differences among the datasets. Strikingly, all datasets with "M2 predominant" signature were measured with HG-U133_Plus_2 microarray chips, whereas the datasets with "T cell predominant" signature were measured with HuEx-1_0-st or HuGene-1_0-st chips. This raised the question whether different array platforms introduced

systematic bias in the measurement of expression of the genes in the cell type reference profile.
However, one must consider that there are many factors that affect immune infiltration, such as the type of treatment. The composition of the tissue sample also affects the estimated infiltration of immune cells. Such data were not available, so the question regarding differences among the datasets could not be adequately answered here.

The EwS cell lines (dataset Savola cell lines) also showed immune infiltration but in much lower proportions than in the EwS tumor samples. M2 macrophages were the most abundant. This suggests that the EwS cells themselves have an expression profile similar to immune cell types, and therefore could increase the estimated proportion in the profiles.

To conclude, further studies are needed to address the question of a unique EwS immune profile but this is beyond the scope of this work. In summary, our dataset had a similar profile to some public datasets, from which we concluded that our dataset had, at least to some extent, a typical EwS immune profile.

Dataset (GEO Id)	**Array platform**	**n with signif. profile / n total**	**Tumor type**	**Treatment**	**Predominant immune infiltrates**	**Immune signature**
Delattre (GSE34620)	HG-U133 _Plus_2	71/117	primary tumor	no statement	M2, M0	M2 predominant
Savola (GSE17618)	HG-U133 _Plus_2	44/44	32 primary, 7 metastasis, 5 recurrence	29 untreated, 15 treated	M2, M0	M2 predominant
Savola cell lines (GSE17618)	HG-U133 _Plus_2	11/11	//	//	M2	M2 predominant
Dirksen (GSE63157)	HuEx-1_0-st	24/85	primary tumor	no statement	Tregs	T cell predominant
Lawlor (GSE68776)	HuEx-1_0-st	5/32	local tumor	untreated	activated NK cells, Tregs	T cell predominant
Schwabing (-)	HuGene-1_0-st	9/29	5 primary, 16 recurrence, 8 na	12 untreated, 8 treated, 9 na	CD8, M0, monocytes	T cell predominant

Table 6: Signature of immune cell infiltrates in 6 EwS datasets.*Each dataset is described with technical details (GEO ID; array platform) and composition (number of significant deconvolution results, meaning Cibersortx p<0.05, of total sample number; tumor type; patients' treatment). The predominant immune cell infiltrates are listed (cell types with mean absolute abundance>0.25% from Cibersortx). There are two main signatures: M2 predominant (Delattre, Savola, Savola cell lines) or T cell predominant (Schwabing, Dirksen, Lawlor).*

3.5.2.2 P-values Provided by Cibersortx

For further interpretation of the results, Cibersortx calculates empirical p-values for each sample, which represent the credibility of the deconvolution results. The p-values were significant for a subset of samples in each dataset (Table 6). This indicated that the absolute proportion of immune cell infiltrates had differing reliability.

The profiles of the samples with a p-value<0.05 (9/29 Schwabing, 71/117 Delattre, 24/85 Dirksen, 5/32 Lawlor, 44/44 Savola, and 11/11 Savola cell lines) differed slightly from the profiles of the full datasets. However, they retained the same characteristics and signatures: it had minor impact whether the profile was based on the complete datasets or the significant deconvolution results. In previous studies using Cibersort, such as that of Stahl et al. [18], a p-value has not yet been calculated because it is only available in the successor Cibersortx. For these reasons, we decided to use the profiles of all samples regardless of the p-values.

3.5.2.3 Correlation of SUVmax With CD8 and M2 Infiltration

To investigate relations between immune infiltration and PET signal, we correlated the immune profiles in our dataset with SUVmax, which was available for 19/29 of our EwS samples. The first approach was to split the samples into groups of immune subtypes analogous to Stahl et al. [18], who defined 3 subtypes: “T cell predominant", characterized by high T cell infiltration but low M2 and neutrophil infiltration; "M2 neutrophil predominant", characterized by low T cell infiltration but high M2 and neutrophil infiltration; "mixed" for other combinations. However, this definition was not applicable to our dataset. We had no neutrophil infiltration in our samples. Based on T cell and M2 infiltration, 6/19 samples were "T cell predominant" (high T cell and low M2 infiltration based on median infiltration, respectively), 5/19 were "M2 predominant" (low T cell and high M2 infiltration), and 8/19 were "mixed" (both T cell and M2 either high or low). This grouping seemed ambiguous. For this reason, and because of the fundamental differences in the profiles of the Schwabing cohort and the Stahl cohort described earlier, we decided that this grouping approach was not applicable to our cohort.

Instead of splitting the samples into groups, we analyzed the absolute frequencies of infiltrates and focused on immune cell types of particular interest: immune-activating CD8 T cells promoting anti-tumor responses [133, 137, 138]; and immune-suppressive, pro-tumoral M2 macrophages [18, 138].
We correlated absolute CD8 and M2 infiltration with SUVmax in two ways. First, we used continuous SUVmax values in linear regression. Second, we examined whether CD8 and M2 infiltration differed among SUV groups of tumors (tumors with high SUV compared to tumors with low SUV).
CD8 T cell abundance correlated negatively with SUVmax in linear regression modeling (p-value=0.14, slope -0.012) (Figure 12A). Comparison of SUV groups showed significantly lower CD8 infiltration in tumors with high SUV (p-value=0.01): the mean abundance of CD8 T cells was 0.28 in tumors with high SUV and 0.47 in tumors with low SUV (Figure 12B).
M2 macrophage abundance correlated positively with SUVmax in linear regression modeling (p-value=0.07, slope 0.008) (Figure 12C). Comparison of SUV groups showed more M2 infiltration in tumors with high SUV (p-value=0.08): the mean abundance of M2 macrophages was 0.15 in tumors with high SUV and 0.07 in tumors with low SUV (Figure 12D).

In summary, there is a tendency that M2 macrophages are more likely to infiltrate tumors with high SUV, whereas CD8 T cells are more likely to infiltrate tumors with low SUV. Thus, we conclude that there might be a relationship between immune cell infiltration and glucose uptake.

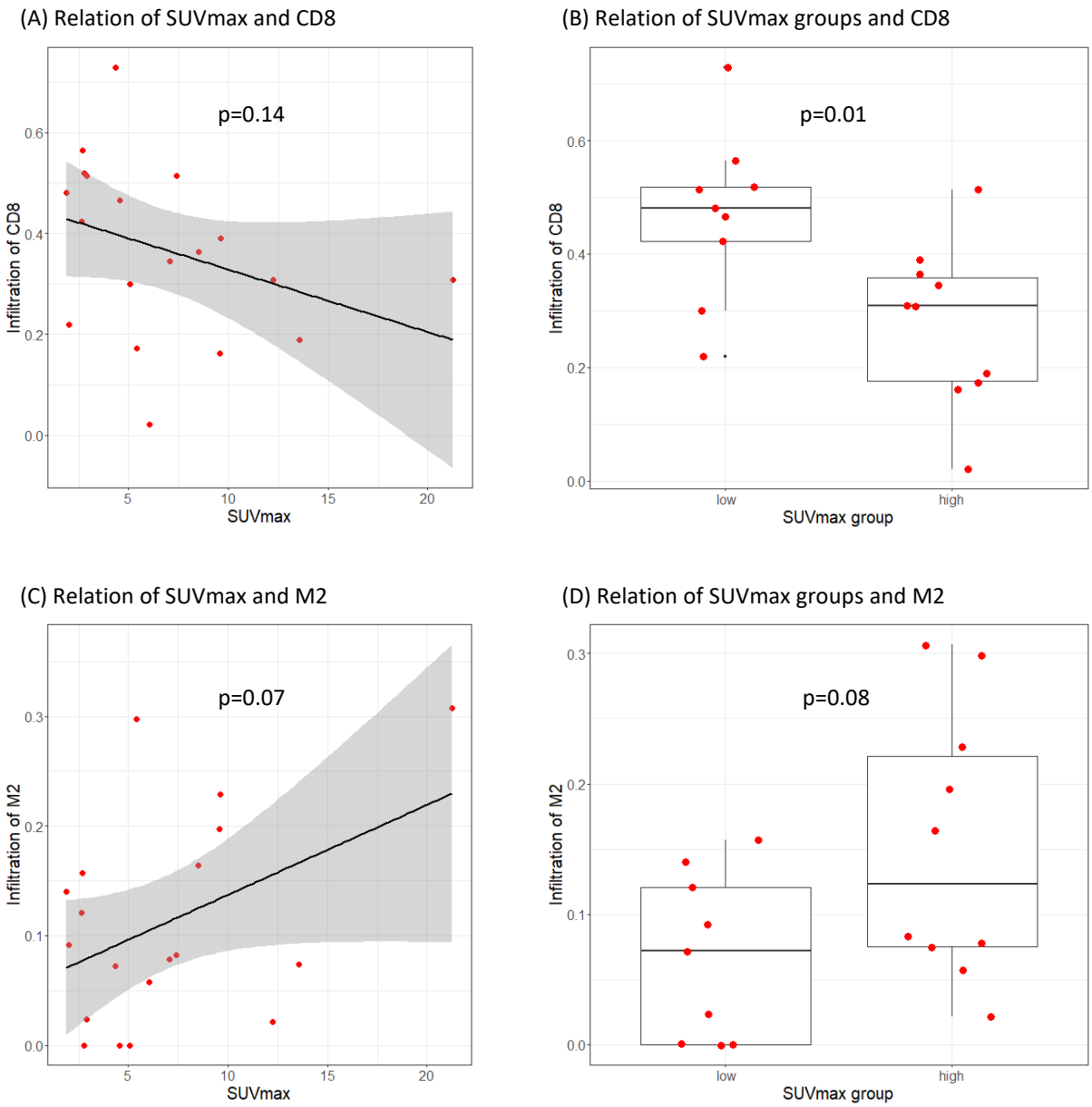


Figure 12: Correlation of SUVmax and absolute infiltration of CD8 T cells (A + B) and M2 macrophages (C + D).*(A) Absolute CD8 T cell abundance correlates negatively with SUVmax in linear regression modeling (not significant, p-value=0.14). The plot shows the regression line and confidence interval. (B) Comparison of SUV groups indicates significantly less CD8 infiltration in tumors with high SUV (p-value=0.01 from Welch two-sample t-test). (C) Absolute M2 macrophage abundance correlates positively with SUVmax in linear regression modeling (not significant, p-value=0.07). The plot shows the regression line and confidence interval. (D) Comparison of SUV groups indicates more M2 infiltration in tumors with high SUV (not significant, p-value=0.08 from the Welch two-sample t-test).*

3.5.3 Linear Regression of SUVmax and Gene Expression

The main step in our analysis workflow was the relation of transcriptomic data with glucose uptake (see Figure 1, the panel at the bottom right, bottom row). We wanted to know whether gene expression was associated with SUVmax? To identify associations in 19 EwS samples, we applied linear regression modeling using the R package limma [112, 113] and examined all 1376 genes that remained after the preprocessing and filtering steps of expression analysis. Recapped, these genes met the quality criteria and were predictive for survival in 3 external EwS datasets. For each gene, a regression model was fit, a p-value was calculated and adjusted for multiple testing using the Benjamini-Hochberg procedure. The volcano plot (Figure 13A) illustrates the effect size (i.e., the slope of the regression line) and r^2 (i.e., the squared Pearson correlation coefficient). r^2 indicates how well the data fit the regression model and implies which proportion of the variance in gene expression can be explained by SUVmax. Additionally, the slope and the adjusted p-value are depicted in Figure 13B. To increase the interpretability of the effect size, the slopes were standardized using Z-score normalization. The distribution of slopes and Z-scaled slopes is depicted in Suppl. Figure 7. Altogether, 645/1376 genes were positively correlated with SUVmax, 731/1376 genes negatively.

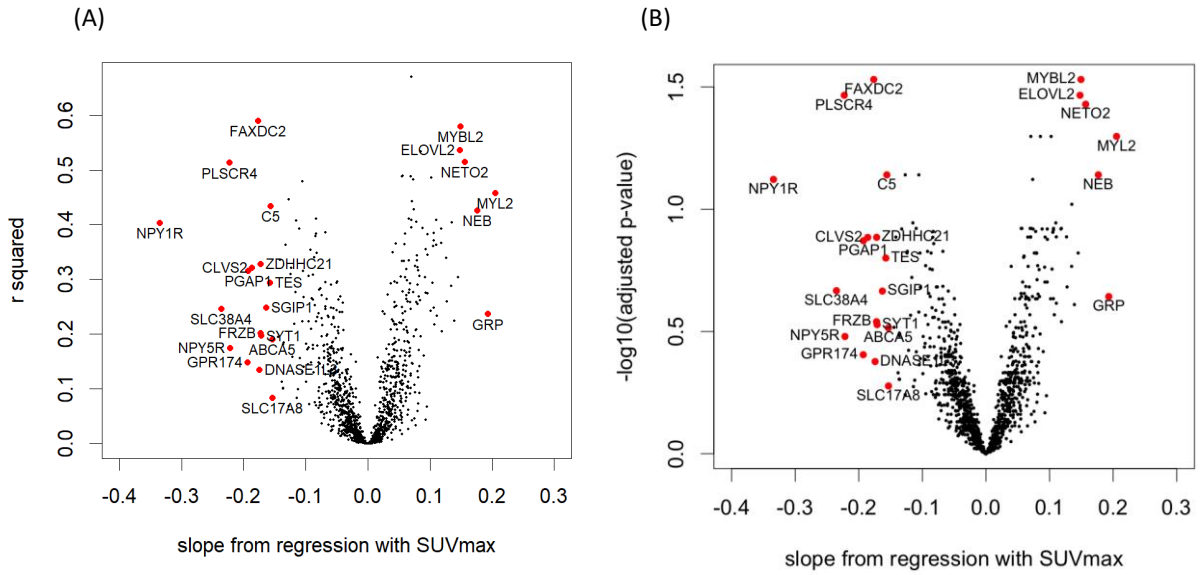


Figure 13: Volcano plot showing results from linear regression modeling of SUVmax and gene expression. *5 genes are significantly correlated with SUVmax with adjusted p-value<0.05 (FAXDC2, MYBL2, PLSCR4, ELOVL2, and NETO2). 23 genes show a high effect size of abs(slope)>0.146 (labelled genes). (A) For each gene tested, the regression line slope and r^2 (Pearson correlation) are plotted. (B) The regression line slope and -log10(adjusted p-value) are plotted.*

3.5.3.1 *Five Genes Significantly Correlated With SUVmax*

The linear regression analysis of gene expression and SUVmax showed 5 significant correlations with adjusted p-value<0.05 (see Suppl. Table 5).
3/5 genes were positively associated with SUVmax (Figure 14, top row):
MYBL2 showed a slope of the regression line of 0.15 (95% CI [0.09; 0.21]) indicating that the gene expression doubles over 6.69 SUV units (r^2= 0.58). ELOVL2 showed a slope of 0.15 (CI [0.08; 0.21]) indicating that the expression doubles over 6.76 SUV units (r^2=0.54). NETO2 showed a slope of 0.16 (CI [0.09; 0.23]) indicating that the expression doubles over 6.38 SUV units (r^2=0.51).

In contrast, 2/5 genes were negatively associated with SUVmax (Figure 14, bottom row):
FAXDC2 showed a slope of -0.18 (CI [-0.25; -0.11]) indicating that the expression halves over 5.67 SUV units (r^2=0.59). PLSCR4 showed a slope of -0.22 (CI [-0.32; -0.12]) indicating that the expression halves over 4.49 SUV units (r^2=0.51).

The Z-scores of the slopes of these 5 genes were ≤-3.37 and ≥2.99, respectively, which corresponds to a probability of <0.003 to obtain more extreme slopes. In summary, the effect sizes were high for these genes, as expression doubles over maximal 6.76 SUV units, which is a difference in SUV values that is commonly seen amongst patients [55, 57-60, 139].

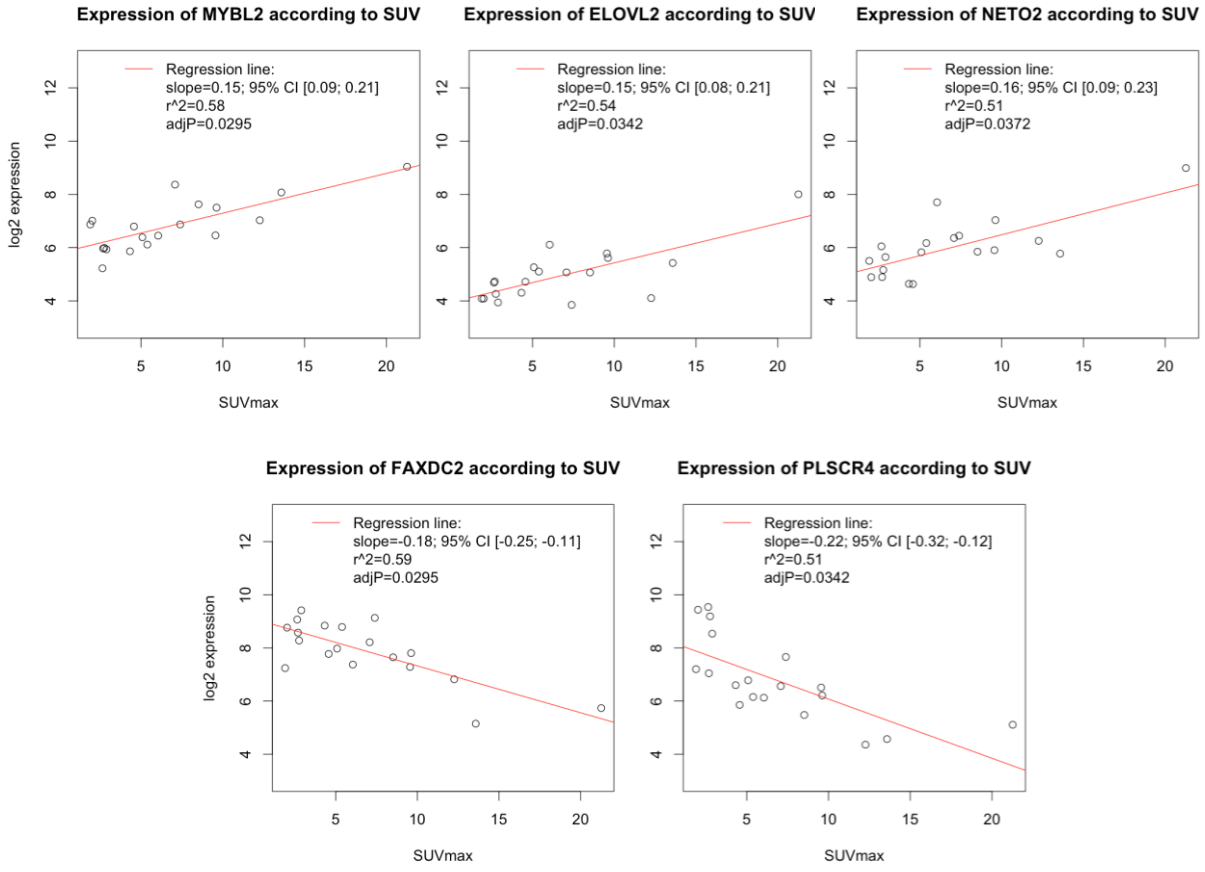


Figure 14: Scatterplots of significant correlations of SUVmax and gene expression (adjusted p-value<0.05).*Expression of MYBL2, ELOVL2, and NETO2 is positively associated with SUVmax (top row), whereas FAXDC2 and PLSCR4 are negatively associated (bottom row).*

3.5.3.2 23 Genes Correlating With SUVmax with High Effect

When interpreting the results of the linear regression of gene expression and SUVmax, especially with our limited sample size, one must not focus on significance alone. Another important aspect is the effect size, which indicates how much gene expression changes in relation to SUVmax. Genes with a high difference in expression levels in relation to SUVmax are more likely to have an impact of biological relevance.

To define genes with high effect size, we chose a cutoff for the regression line slope, namely abs(slope)>0.146. This cutoff corresponds to the standard log fold change of 1, which is normally applied when comparing gene expression levels between two groups, and which we transferred to regression analysis. A slope of 0.146 corresponds to a doubling (or halving) of expression levels over 6.8 SUV units, which seems appropriate regarding the SUV range in our cohort.

Applying this cutoff to the results from linear regression, we obtained 23 genes with high effect size (Figure 13, labelled genes plotted in red). The majority of genes was negatively correlated with SUVmax, namely 17/23, whereas 6/23 were positively associated. The 23 genes included the 5 significant correlations, so the previously described 5 genes had a high effect size by our definition.

Normalizing the slopes to yield Z-scores, we found that the 23 genes had absolute Z-scores>2.91 (Figure 13B, labelled genes plotted in red). This corresponds to a probability of 0.0036 to obtain a more extreme effect, which would be only 5 genes in our dataset by expectation. In summary, the slopes of the regression lines for these 23 genes are very extreme compared to the mean of all slopes in our analysis.

The expression of the 23 genes in the 19 samples is illustrated in a heatmap (Figure 15), together with clinical variables of the patients and their classification into 2 groups of low SUV or high SUV according to median SUVmax.

The hierarchical clustering of the samples based on expression displayed outliers, namely the 2 tumors with the highest glucose uptake in our cohort (SUVmax 13.6 and 21.3), and 2 slight outliers with low SUVmax (2.7 and 4.3). Without these outliers, the samples split into 2 groups: a cluster with lower SUVmax (5/6 with low SUV), and a cluster with higher SUVmax (7/9 with high SUV). As this sample clustering was based on genes whose expression changes strongly in relation to SUVmax, it nicely reflected the samples' spectrum of SUVmax values.

The hierarchical clustering of the genes displayed 2 clusters without outliers. In the smaller cluster, 5/6 genes correlated positively with SUVmax. In the larger cluster 16/17 genes correlated negatively with SUVmax. Thus, the clustering reflected the two directions of association with SUVmax.

A list of the 23 genes with high effect size is provided in Suppl. Table 5 giving the results from linear regression (slope, Z-score of slope, adjusted p-value, r^2 from Pearson correlation) and a functional annotation of the genes. In summary, these 23 genes showed a substantial change in expression level with regard to increasing SUVmax, and may have effects of biological relevance. The clustering based on their expression suggests that there is an expression signature for metabolic activity.

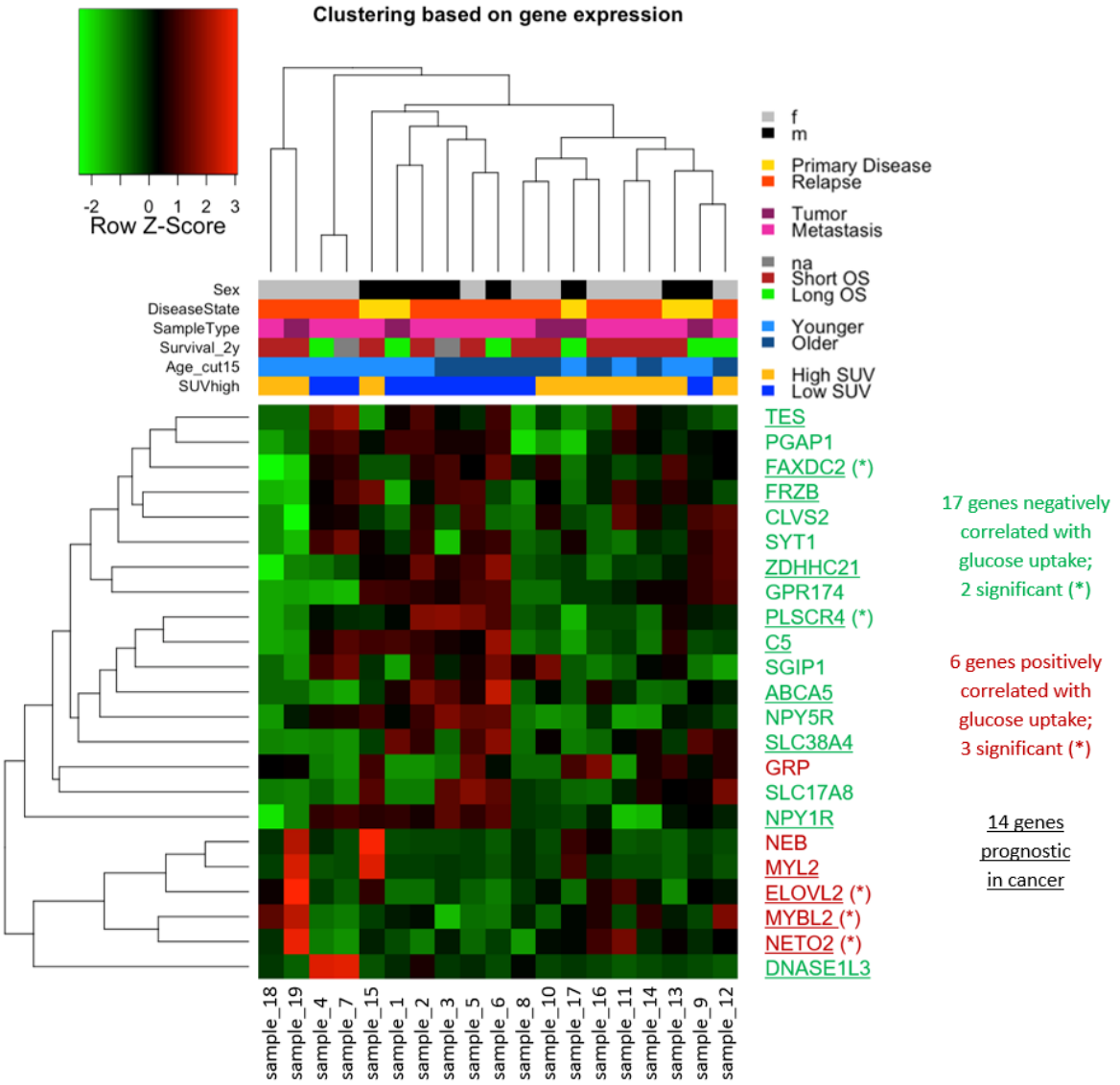


Figure 15: Heatmap depicting expression of 23 genes with high effect (abs(slope)>0.146) in linear regression with SUVmax. *On top, clinical data is provided about sex (male or female), disease state (primary disease or relapse), sample type (tumor or metastasis), 2-year overall survival, age (≤15 years or >15 years), and categorical partitioning of the samples into high SUV or low SUV (split by median SUVmax). 17/23 genes were negatively associated with SUVmax, 6/23 positively. 14/23 genes were prognostic in cancer according to the Human Protein Atlas.*

3.5.3.3 Linear Regression Without Sample_19 of Extreme SUVmax

Examining the distribution of SUVmax values in our cohort, 18/19 samples had SUVmax values between 1.9 and 13.6. One sample with a much higher SUVmax of 21.3 stood out (sample_19). This skewed distribution of SUVmax values may cause that the sample with the extreme SUVmax affects the linear regression more than a single sample in the group of SUVmax values. This could be problematic if sample_19 is an outlier in the dataset. Arguing against sample_19 being an outlier is the clustering of samples in the heatmap of genes with high effect size (Figure 15). In this heatmap, sample_19 is no outlier but is placed next to the sample with the second highest SUVmax of 13.6 (sample_18). This indicates that both samples are similar in their expression pattern.
Nevertheless, we checked how sensitive our results were with respect to sample_19. To this end, we repeated the linear regression of gene expression values for 18 samples – removing sample_19 – and compared how much the correlations had changed.

The effect sizes of all 1376 genes for both regression analyses are depicted in Figure 16. The horizontal and vertical blue line corresponds to an effect size of 0, respectively. The red diagonal line indicates identical effect sizes, meaning that the slope is independent of sample_19.
For the genes in the red highlighted areas, the presence of sample_19 heightened the effect in comparison to the linear regression without sample_19 (294 genes). These were 165/294 genes with positive effect size and 129/294 with negative.
For the genes in the yellow highlighted areas, sample_19 diminished the effect size. Thus, the effect increased when sample_19 was absent (791 genes). These were 301/791 genes with positive effect size and 490/791 with negative.
In short, the presence of sample_19 attenuates the effect size for most genes (791 vs. 294 genes).

Of particular interest is the location of the genes relative to the green lines marking the threshold for high effect size (abs(slope) 0.146). The results falling into the green rectangles were independent of the presence of the sample_19 (Suppl. Table 6). 15 of the 23 genes with high effect size still showed it after removing sample_19: 2/6 with positive correlation (NEB and GRP) and 13/17 with negative correlation (NPY5R, NPY1R, PLSCR4, PGAP1, TES, DNASE1L3, ABCA5, ZDHHC21, SLC38A4, C5, FAXDC2, SGIP1, and GPR174).
The genes next to these green rectangles showed a high effect size in either the analysis without sample_19 (47 genes) or in the analysis with sample_19 (8 genes). Thus, sample_19 contributed strongly to the high effect size we originally found for 8/23 genes (CLVS2, MYL2, ELOVL2, NETO2, SYT1, FRZB, SLC17A8, and MYBL2). These findings only showed up due to the presence of sample_19.
In the original regression analysis, 5 genes were significantly correlated with SUVmax. Of these, only PLSCR4 was significant without sample_19. Instead, 4 other genes showed significant correlations: SNX13, ESPL1, TET1, and NAPEPLD. Thus, the significant results relied on the presence of sample_19.

Overall, the presence of sample_19 mostly diminished the effect size. Only a smaller portion of the results was dependent on sample_19.
The results for the 23 genes with high effect size were mostly stable. This supports that the results based on effect size are more robust findings than the results based on significance and that it is important to focus on robust methods such as enrichment analysis.

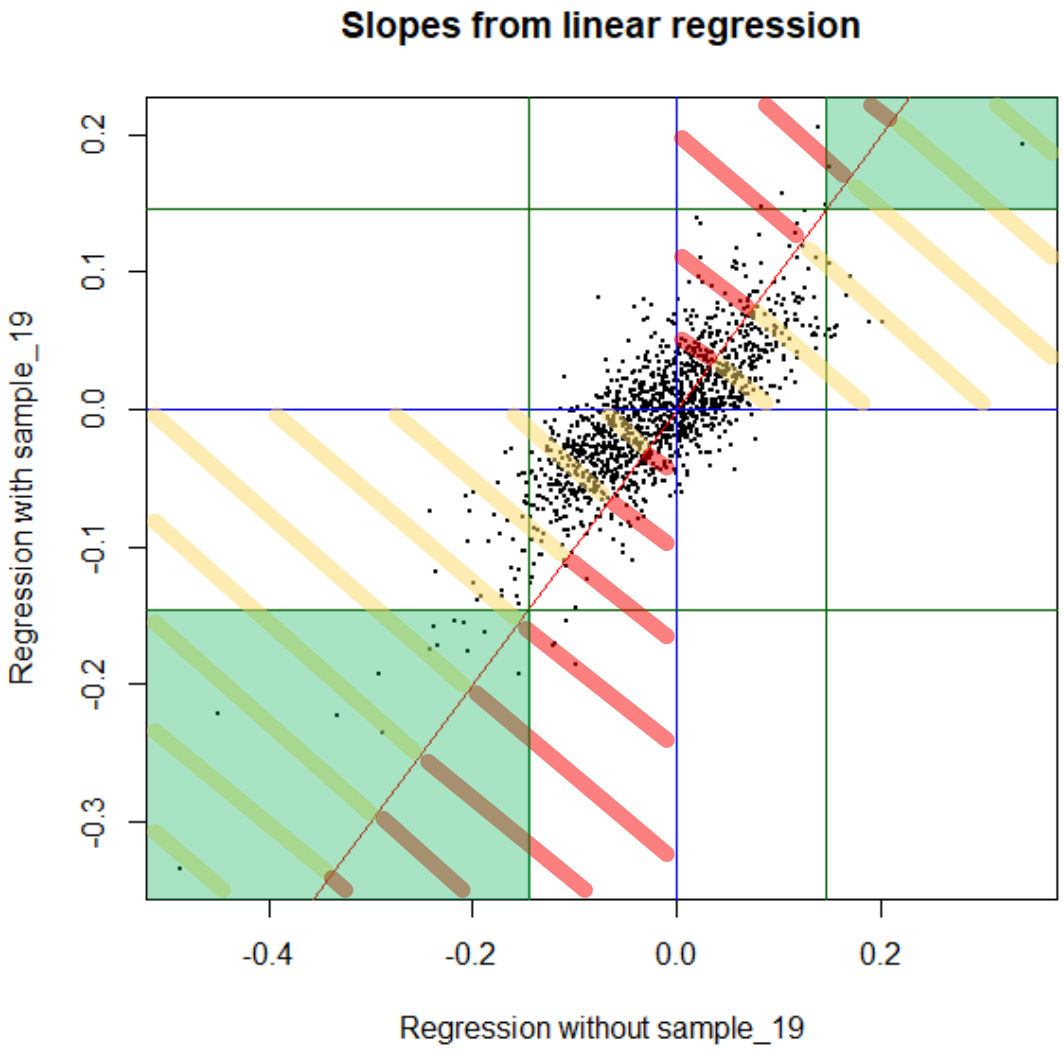


Figure 16: Effect sizes from linear regression analysis with and without sample_19. *For each gene (1376 genes) the slope from linear regression is depicted for both regression analyses. The horizontal and vertical blue line corresponds to an effect size of 0, respectively. The red diagonal line indicates identical effect sizes, meaning that the slope is independent of sample_19. The green lines mark the threshold for high effect size (abs(slope) 0.146). The red highlighted areas indicate that the presence of sample_19 increases the effect size (294 genes), while the yellow highlighted areas indicate that the effect size is diminished in presence of sample_19 (791 genes). Genes falling into the green rectangles have a high effect size independent of sample_19 (15 genes).*

3.6 Enrichment Analysis of Regression Results

After analyzing single correlations of gene expression and SUVmax, we asked for shared pathways and processes that correlated with glucose uptake. The aim was to summarize and generalize the results of the correlation analysis on a functional level. We investigated whether there were pathways and functionally annotated gene sets characterizing tumors with varying glucose uptake by applying different approaches of enrichment analysis. Enrichment analyses are robust to false positive findings because many genes are considered at once and a single false positive will be of less consequence. This is an advantage especially with our small sample size.

3.6.1 Annotation of 23 Genes With High Effect Size

For the 23 genes that correlated with SUVmax with high effect size, we looked for shared pathways and functions. To this end, we scanned their annotation for their prognostic value in cancer entities (Suppl. Table 5), and then performed enrichment analysis.

First, we wanted to know whether the 23 genes were prognostic in cancer. In total, 14/23 genes predict survival in several cancer types [140]. These are underlined in the heatmap of gene expression (Figure 15). Expression of ABCA5, C5, DNASE1L3, ELOVL2, FAXDC2, NPY1R, SLC38A4, TES, and ZDHHC21 predicts a favorable outcome in breast, liver, pancreatic, renal, and urothelial cancer. In contrast, expression of FRZB, MYBL2, MYL2, NETO2, PLSCR4, and TES predicts unfavorable outcome in endometrial, head and neck, liver, pancreatic, and renal cancer.

3.6.1.1 Genes With High Effect Size Enriched for NPY Signaling Pathway

Next, we tested for enrichment of pathways and functions systematically using the tool Enrichr [117-119]. We focused on pathways in “Reactome 2016” and the Gene Ontology (GO) knowledgebase including “GO biological process 2021”, “GO molecular function 2021” and “GO cellular component 2021”. As we decided to be more stringent, we considered enrichments with adjusted p-value (q-value)<0.01 as significant.

The set of 23 genes with high effect size showed significant enrichment for the Reactome pathway “Peptide ligand-binding receptors Homo sapiens R-HAS-375276” with q=0.004 (Figure 17A), which included a subset of the rhodopsin-like G protein-coupled receptor (GPCR) family. This enrichment was due to 4 genes: C5, NPY1R, NPY5R, and GRP.

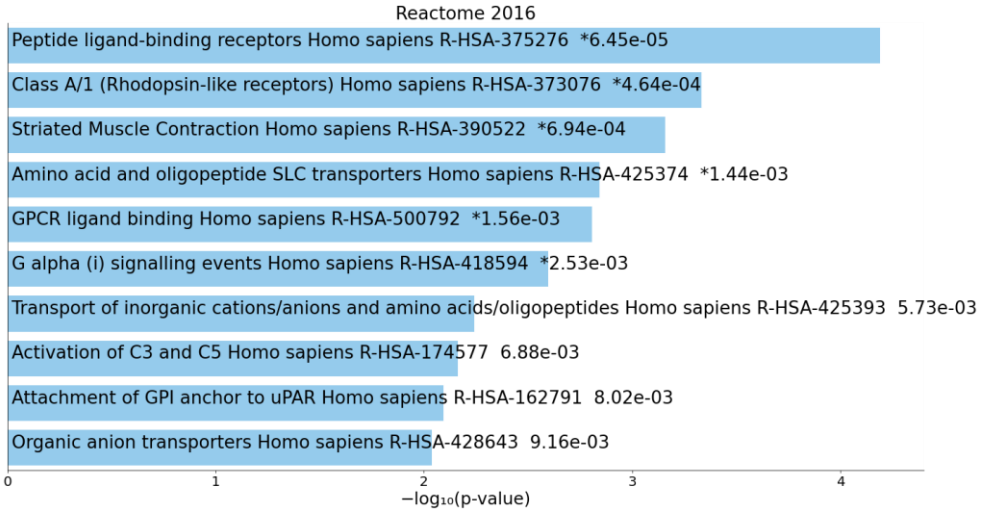
There were no significant enrichments in GO biological process and GO cellular component.

However, the 23 genes were significantly enriched for the GO molecular function "neuropeptide Y receptor activity (GO:0004983)" with q-value=0.0007 (Figure 17B). NPY receptors are rhodopsin-like receptors, too. This enrichment was based on the 2 genes NPY1R and NPY5R.

Figure 17: Enrichments among the 23 genes with high effect size (abs(slope)>0.146) found by Enrichr.*Details for all terms with adjusted p-value<0.05 are provided below, respectively. Here, “overlap” indicates the proportion of genes with a high effect size among the genes in the gene set. (A) Enrichments of “Reactome Database”. Considering terms with p-value<0.01 as significant, one term obtains a significant adjusted p-value (q=0.0043): “Peptide ligand-binding receptors Homo sapiens R-HAS-375276”. (B) Enrichments of “GO molecular function”. Considering terms with p-value<0.01 as significant, one term obtains a significant adjusted p-value (q=0.0007): “neuropeptide Y receptor activity (GO:0004983)”.*

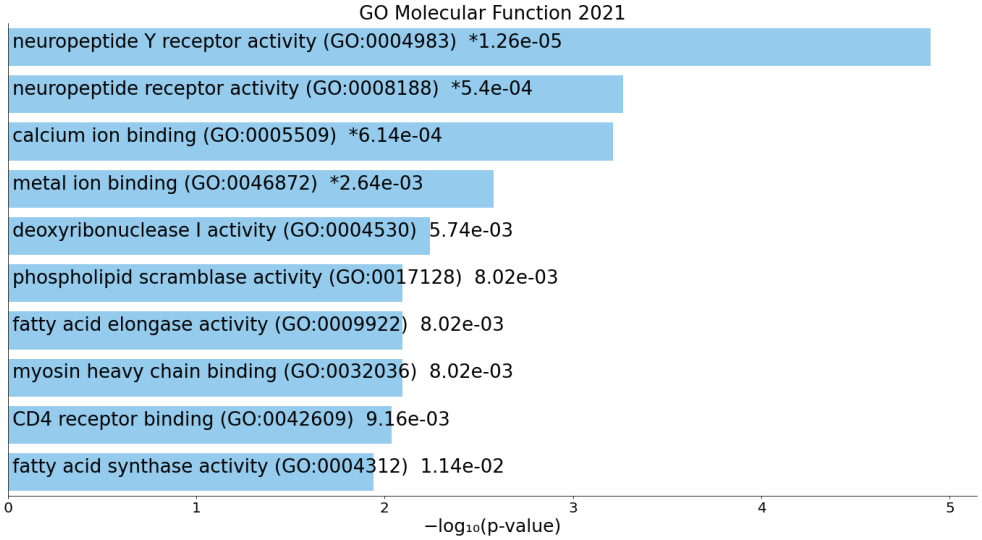
(continued →)

(A)



Term	Overlap	Genes	P-value	Adjusted P-value (q)
Peptide ligand-binding receptors Homo sapiens R-HSA-375276	4/193	C5; NPY5R; GRP; NPY1R	0.00006	0.004
Class A/1 (Rhodopsin-like receptors) Homo sapiens R-HSA-373076	4/323	C5; NPY5R; GRP; NPY1R	0.0005	0.015
Striated Muscle Contraction Homo sapiens R-HSA-390522	2/34	MYL2; NEB	0.0007	0.015
Amino acid and oligopeptide SLC transporters Homo sapiens R-HSA-425374	2/49	SLC17A8; SLC38A4	0.001	0.021
GPCR ligand binding Homo sapiens R-HSA-500792	4/447	C5; NPY5R; GRP; NPY1R	0.002	0.021
G alpha (i) signalling events Homo sapiens R-HSA-418594	3/240	C5; NPY5R; NPY1R	0.003	0.028

(B)



Term	Overlap	Genes	P-value	Adjusted P-value
neuropeptide Y receptor activity (GO:0004983)	2/5	NPY5R; NPY1R	0.00001	0.0007
neuropeptide receptor activity (GO:0008188)	2/30	NPY5R; NPY1R	0.0005	0.011
calcium ion binding (GO:0005509)	4/348	SYT1; MYL2; PLSCR4; DNASE1L3	0.0006	0.011
metal ion binding (GO:0046872)	4/517	SYT1; MYL2; PLSCR4; DNASE1L3	0.003	0.037

3.6.1.2 NPY Pathway Negatively Associated With Glucose Uptake

As the NPY receptors contributed to both enrichments, we further investigated the role of the NPY pathway in our dataset. The expression of the genes in the NPY signaling axis was decreased as SUVmax increased (Figure 18).
The signaling molecule NPY showed a slope of the regression line of -0.14 (95% CI [-0.26; -0.0087], Z-score=-2.57), implying that expression halved per 7.37 SUV units (r^2=0.18).
The NPY receptor NPY1R showed a slope of -0.33 (95% CI [-0.51; -0.15], Z-score=-6.46), indicating that expression halved per 2.99 SUV units (r^2=0.40).
NPY5R, another NPY receptor, showed a slope of -0.22 (95% CI [-0.43; -0.01], Z-score=-4.26), indicating that expression halveed per 4.51 SUV units (r^2=0.17).

In contrast, 2 other NPY receptors (NPY2R and pseudogene NPY6R) were expressed constantly regardless of SUVmax. NPY2R showed a slope of 0.00007 (95% CI [-0.03; 0.05]), and NPY6R showed a slope of 0.009 (95% CI [-0.04; 0.04]).
In addition, there are 2 paralogs of the signaling molecule NPY, namely PYY (peptide YY) and PPY (pancreatic polypeptide). Their expression was independent of SUVmax as well: PYY showed a slope of -0.005 (95% CI [-0.06;0.05]), and PPY showed a slope of 0.015 (95% CI [-0.02; 0.05]).

3.6.1.3 Summary of Annotation of Genes With High Effect

All in all, of the 23 genes that were strongly associated with SUVmax, most are associated with survival in several cancer entities and therefore play a role in distinguishing subgroups in entities other than EwS. Our findings, namely that these genes have different expression levels in EwS tumors with high or low glucose uptake indicating a different prognosis, suggest that they may also distinguish subgroups in EwS.
Looking for functional similarities of the 23 genes, we found that the NPY signaling axis correlated negatively with glucose uptake.

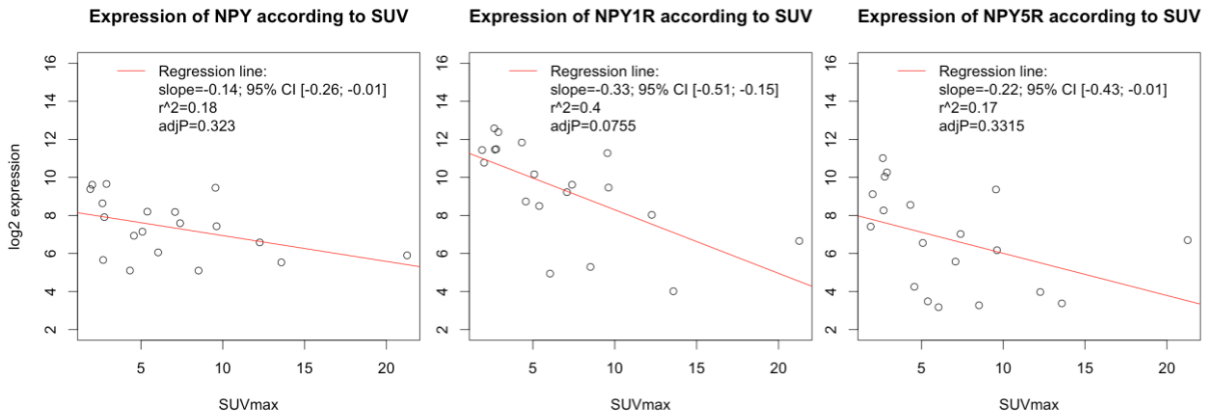


Figure 18: Correlation of SUVmax and the NPY axis.*Expression of NPY axis members (NPY, NPY1R, NPY5R) decreases as SUVmax increases.*

3.6.2 GSEA of the Total Regression Results

In addition to enrichments in the set of 23 genes with high effect size, we investigated enrichments across the entire results obtained from linear regression. To this end, we used gene set enrichment analysis (GSEA), which has several advantages compared to enrichment methods working on a set of genes of interest. First, no cutoff has to be chosen to determine the gene set for enrichment testing, which makes the whole analysis less arbitrary. Second, GSEA utilizes much more information because the entire linear regression results serve as input, namely a list of genes ranked by their correlation with SUVmax. Using this ranking of genes, the direction of the association is considered: whether enrichment for a gene set occurs at the top of the list (among genes that correlate positively with SUVmax) or at the bottom (among genes that correlate negatively with SUVmax). Thus, GSEA provides a broader view of which functional gene sets are related to glucose uptake compared to the analysis limited to the 23 genes with high effect size.

With GSEA, we tested different categories of gene sets. First, we used "hallmark gene sets" (H), which provided initial insight and a general overview of all categories. We then focused on more specific aspects. To investigate pathways, we used "curated gene sets: canonical pathway" (C2cp), which summarized pathways from 5 databases (BIOCARTA, KEGG, PID, REACTOME, and WikiPathways). We also scanned for transcription factors (TFs) whose targets were positively or negatively correlated with SUVmax. Thus, we used "regulatory target gene sets: transcription factor targets" (C3tft), which contained gene sets that share TF binding sites or motifs. Finally, we tested "ontology gene sets" (C5) containing terms from GO and the Human Phenotype Ontology (HPO).

For each gene set tested for enrichment, GSEA calculates a normalized enrichment score (NES). The NES indicates the extent to which this gene set is enriched at the top or bottom of the given ranked gene list. Additionally, a p-value is calculated and corrected for multiple testing (FDR q-value). As the authors suggest less stringent q-value cutoffs for formulating hypothesis, we considered $q<0.1$ as significant.

3.6.2.1 Basic Cellular Functions Positively Associated With Glucose Uptake

The results of GSEA – the number of enriched terms found for each category – is listed in Table 7. Across all categories, we observed more significant enrichments associated with high SUVmax than enrichments associated with low SUVmax. In H, there were 3 terms enriched at high SUVmax, in C2cp 16 terms, in C5 71 terms – all representing basic functions such as cell cycle, DNA replication and repair, transcription, cytoskeleton, actin-myosin interaction, muscle and muscle development. These processes appeared to be upregulated in EwS cells in the context of increased glucose uptake.

3.6.2.2 Rhodopsin-Like GPCRs Negatively Associated With Glucose Uptake

However, we also found enrichments that covered more specific aspects.
Looking at category C2cp for canonical pathways, 2 pathways were significantly enriched at low SUVmax (Figure 19A): "REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS" (NES -1.90; q=0.042) and "WP_GPCRS_CLASS_A_RHODOPSINLIKE" (NES -1.86; q=0.036), which both involved rhodopsin-like GPCRs. The genes that contributed most to the enrichment (so-called core enrichment) were NPY1R, NPY5R, C5, NPY for "REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS", and NPY1R, NPY5R, GPR174, LPAR6, LPAR4 for "WP_GPCRS_CLASS_A

_RHODOPSINLIKE" (Suppl. Table 7A-B). For both terms, the enrichments were predominantly due to genes of the NPY signaling axis. As NPY receptors belong to the class of rhodopsin-like receptors, these findings mirror the results of the previous enrichment analysis of Enrichr on the 23 genes with high effect size.

3.6.2.3 Transcription Factor Activity of RNF2, the E2F family, and TCF3 Positively Associated With Glucose Uptake

Furthermore, in category C3tft of TF targets, 3 terms showed significant enrichment. The targets of RNF2 (NES 2.26; q=0.003), of the E2F family of TFs (NES 2.00; q=0.065), and of TCF3 (NES 1.99; q=0.048) all showed enrichment among genes positively associated with SUVmax (Figure 19B, genes in core enrichment see Suppl. Table 7C-E). This may indicate that the activity of RNF2, the E2F family and TCF3 was related to the glucose uptake by EwS cells.

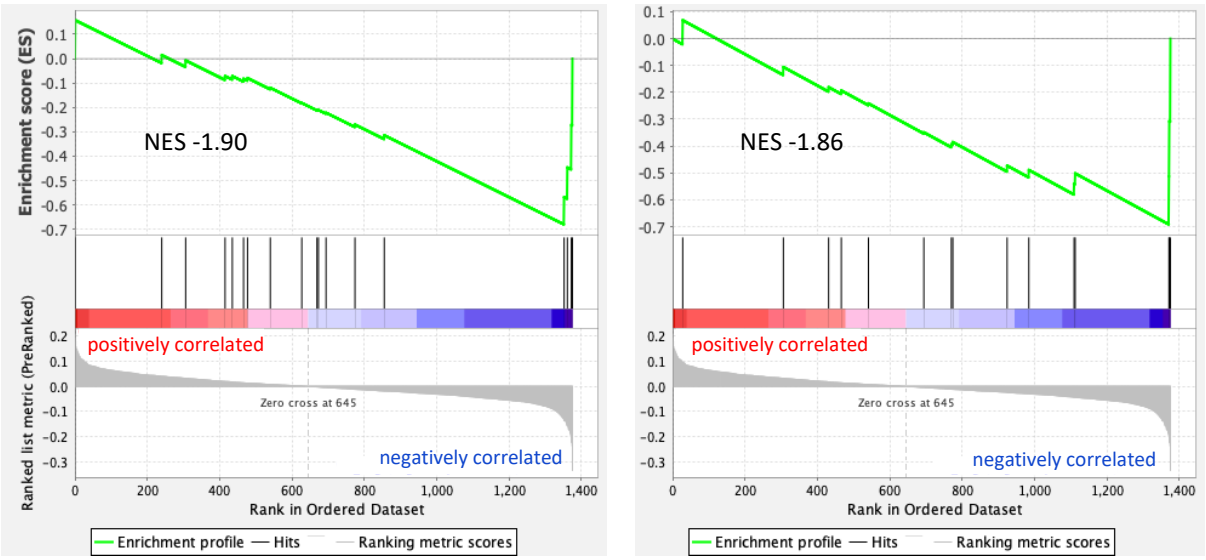
3.6.2.4 Summary of GSEA Enrichments

In summary, we identified pathways and TFs that characterize tumors in terms of glucose uptake. Tumors with high glucose uptake had more enriched terms than tumors with low glucose uptake, especially terms referring to increased turnover like cell cycle, replication and transcription. Furthermore, the activity of the 3 TFs RNF2, the E2F family and TCF3 might be positively associated with glucose uptake, whereas rhodopsin-like receptor pathways might be negatively associated with glucose uptake in EwS tumors.

			Enriched among genes with positive correlation with SUVmax		Enriched among genes with negative correlation with SUVmax	
Category	**Category name**	**Number of gene sets used for analysis / number of gene sets in category**	**Number of gene sets**	**Sets with FDR<0.1**	**Number of gene sets**	**Sets with FDR<0.1**
H	Hallmark gene sets	32/50	13	3	19	0
C2cp	Canonical pathway	129/2922	74	16	55	2
C3tft	Transcription factor targets	730/1133	285	3	445	0
C5	Ontology gene sets	1748/14998	852	71	896	0

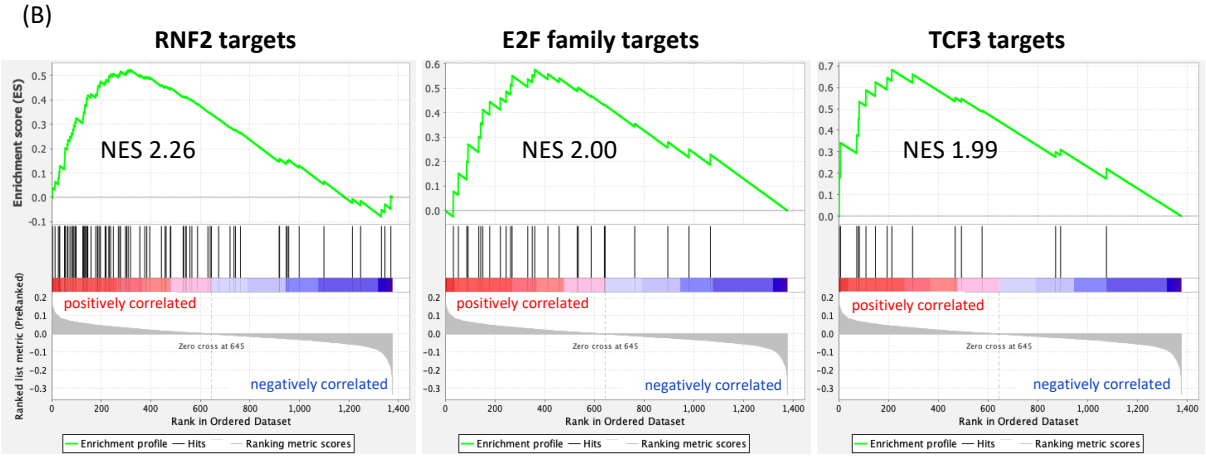
Table 7: Results from GSEA analysis on 1376 genes ranked by their correlation with SUVmax. *For each category (H "hallmark gene sets", C2cp "canonical pathway", C3tft "transcription factor targets", C5 "ontology gene sets"), the number of annotated gene sets is given. Some gene sets are filtered out before analysis based on size (default criteria for analysis: gene set size min=15 and max=500). The remaining gene sets are either enriched among genes with positive (red columns) or with negative correlation (green columns) with SUVmax. Finally, the number of gene sets with significant enrichments (FDR<0.1) is given.*

(A)



Gene set	Size	ES	NES	p-value	FDR q-value
REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS	18	-0.68	-1.90	<0.0005	0.042
WP_GPCRS_CLASS_A_RHODOPSINLIKE	15	-0.69	-1.86	0.002	0.036

(B)



Gene set (name in GSEA)	Size	ES	NES	p-value	FDR q-value
Target genes of RNF2 (HSD17B8_TARGET_GENES*)	81	0.52	2.26	<0.0005	0.003
Target genes of E2F family (E2F_Q6)	27	0.58	2.00	<0.0005	0.065
Target genes of TCF3 (E47_01)	16	0.68	1.96	<0.0005	0.048

* HSD17B8 was wrong name in GSEA at the time of analysis (presumably corrected in the meantime)

Figure 19: Enrichment plots of GSEA analysis on 1376 genes ordered by positive to negative correlation with SUVmax. *(A) For canonical pathways (C2cp), 2 gene sets of rhodopsin-like receptors are enriched among genes that are negatively associated with SUVmax: "REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS" and "WP_GPCRS_CLASS_A_ RHODOPSINLIKE". (B) 3 transcription factors (TFs) (in C3tft) show target gene enrichment among genes that are positively associated with SUVmax: RNF2, E2F family, TCF3.*

4 Discussion

Investigating a systemic disease requires whole-body examinations like functional FDG-PET imaging. EwS has long been known for its predilection for glucose utilization and FDG-PET sensitivity [10]. However, the biological basis for elevated but variable glycolytic activity is not well understood. Furthermore, it is not known how radiomic data from PET imaging relates to transcriptomic data in EwS.
Long-term treatment effects and poor survival rates in EwS require the discovery of new treatment methods and individualization of therapy. This necessitates the identification of genes that can serve as biomarkers for risk stratification, i.e., that correlate with survival and malignancy features. Such genes can eventually provide opportunities for novel treatment options.

As the first large-scale, functional radiogenomic study in EwS, the present study comprehensively examined the correlations between gene expression and radiomic data from PET imaging.
We found genes, signaling pathways, immune infiltration and transcription factors (TFs) characterizing EwS tumors in terms of SUVmax, i.e., maximal glucose uptake and consecutive glycolysis.
What most studies investigating the prognostic value of PET have in common is that high SUV values indicate a worse survival and more aggressive stage of disease, although the exact cutoff values vary. Nevertheless, it is established that higher glucose uptake, in general associated with the Warburg effect [141-144], indicates a higher level of malignancy. As PET signals like SUVmax are prognostic in EwS, we anticipate that our findings will provide novel biomarker candidates for risk stratification, which may also open up new therapeutic options in the future.

4.1 Discussion of Assumptions and Methodology

4.1.1 SUVmax as Prognostic Phenotype

First of all, it needs to be discussed whether IFs and especially SUVmax are suitable as prognostic phenotype. In this context, some limitations of PET imaging and SUVmax should be considered.

4.1.1.1 Relationship of SUVmax and Tumor Volume

First, it can be a drawback that SUVmax correlates with lesion volume, which is stated in different studies [124-128] and also found in a subgroup of our patient cohort. However, this indicates that SUVmax is associated with another prognostic factor, namely tumor size. Thus, this correlation is consistent with expectations. It is not reasonable to correct for this correlation as tumor size may contribute to the prognostic value of SUVmax.

4.1.1.2 Variability of PET Signals and SUVmax

A drawback of PET imaging is that signals are variable due to technical influences, biological and physical sources of error, and different image reconstruction methods [33]. SUVmax signals varied by 10-16% in a test-retest study of 62 patients from 8 imaging centers [145].

Even with the same scanner, 15% variance was measured at different time points [33]. Kinahan et al. [33] discuss that SUVmax may vary by more than 15-20% in reality.
In addition, SUV is not measured correctly for lesions smaller than 3 cm due to the partial volume effect [33, 37]. However, Kinahan et al. [33] state that SUVmax is more reproducible and stable than SUVmean, and a better representation of the ground truth [5]. Although SUVmax – unlike SUVmean – describes only one pixel in the tumor region, the variance and bias is lower than expected [33].
In our cohort, we had 5/19 small lesions with less than 3 cm^3. If we would select samples by size, we would introduce another bias. Thus, we used all samples in our cohort, which also reflects reality better as tumors do vary in size. At least, using SUVmax seems to be a valid choice to estimate ground truth [33].

4.1.1.3 Inconsistency of the Literature Regarding the Prognostic Value of SUVmax

One limitation regarding the prognostic value of SUVmax is that study results are not completely consistent. Some studies found pretreatment SUVmax to be prognostic for survival [55, 56, 59] but others did not – yet, post-treatment SUVmax was prognostic here [57, 58]. The SUVmax cutoff value to define prognostic subgroups was also study-dependent. For example, Hwang et al. identified a cutoff of 5.8 [55], Salem et al. identified 11.6 [59], while Jamet et al. [56] found a significant effect of SUVmax on overall survival but could not determine a cutoff value at all.

These deviations in the studies indicate that one cannot rely on cutoff values for SUVmax. The aforementioned general variability of PET signals certainly plays a role in these studies as well. It is not straightforward to find a universal cutoff, and it is questionable whether cutoff definitions are useful at all if they have limited validity [33].
Our approach is to circumvent the problem by working without cutoff definitions but correlating SUVmax values directly with gene expression using linear regression. Overall, we focus on cutoff independent results, which is why we specifically rely on the GSEA results.

4.1.1.4 Bias Due to Joint Analysis of PET-CT and PET-MR

Another source of bias in our study may be the joint analysis of PET-CT and PET-MR. Studies analyzing simultaneous PET-CT and PET-MR runs on tumors of different entities show that signals are not identical [146-148].
Reasons for this may be technical effects like different PET scanners, and different methods of attenuation correction [146, 147]. In addition, different time intervals between injection and imaging for PET-CT and PET-MR could have an effect as uptake may vary [147].
SUVmean and SUVmax measured in PET-CT are slightly higher than in PET-MR. For example, average SUVmax was 5.8 in PET-MR and 7.9 in PET-CT [147], and average SUVmean was 5.4±3.1 in PET-MR and 6.1±3.1 in PET-CT [148]. At least, both modalities showed a strong correlation in the studies (r=0.85 [146], r=0.855 [147], and r=0.93 [148]). Thus, there seems to be a linear relationship between SUVs in PET-CT and PET-MR. Sachpekidis et al. [147] and Drzezga et al. [148] described these relationship using formulas from linear regression lines. However, the two studies differ as the regression line was SUVmax_MR=0.735*SUVmax_CT-0.003 in Sachpekidis et al. [147], and SUVmean_MR=0.79*SUVmean_CT+0.57 in Drzezga et al. [148]. Again, the variability of PET signals probably affects these analyses.

In our dataset, we also analyzed the correlation of SUVs between PET-CT and PET-MR using 21 simultaneous runs in 15 sarcoma patients. The intraclass correlation coefficient (ICC) of SUVmean and SUVmax was 0.72, respectively, which is interpreted as "moderate" agreement according to Koo and Li [99] or "good" agreement according to Cicchetti [100]. This corresponds to the findings in the literature.
In our dataset for correlation analysis with gene expression values, the majority of 15/19 samples were PET-CT. 4/19 samples were PET-MR, distributed over the entire SUVmax spectrum: sample_1 (SUVmax 1.9), sample_9 (SUVmax 5.1), sample_13 (SUVmax 7.4), and sample_15 (SUVmax 9.6) (see Suppl. Table 1). It is likely that these SUVmax values are little underestimated compared with the SUVmax values from PET-CT. One idea was to correct for this and estimate SUVmax_CT from SUVmax_MR based on the literature. However, since the formulas in [147] and [148] differ and do not provide a universal description of the relationship, this was not applicable.
As the minority of 4/19 PET-MR in our cohort are spread across the SUVmax spectrum not introducing bias selectively, and SUVmax values have intrinsic variance, we assume that there is no major impact of imaging modality in our analysis. Furthermore, studies agree about comparability of modalities to some extend [146, 148]. To eliminate any bias in this regard, we will validate our results on an external cohort of the same imaging modality in a follow-up project (n=7 samples of PET-CT series from Münster).

4.1.1.5 SUVmax as Established Prognostic Marker Despite Limitations

One must consider the discussed limitations of imaging and SUVmax. However, SUVmax is an established prognostic marker used in clinical practice. Although cutoffs and signal strength vary between studies, it is widely accepted that high values indicate poor prognosis and malignancy. Although the ground truth can only be measured with variance, SUVmax seems to be the most appropriate value reflecting this ground truth [33]. Therefore, we posit that SUVmax is suitable as prognostic quantitative phenotype.

4.1.2 Approach of Linear Regression Versus Comparison of Groups

In our study, we correlated gene expression data with SUVmax values of tumors. For doing so, there are different approaches: comparing groups of patients or linear regression analysis.

4.1.2.1 Standard Approach of Comparing Phenotypic Groups

An established approach is to stratify tumors into two groups based on PET signal and compare the distribution of expression levels for each gene. To determine differentially expressed genes, a fold change of 2 is usually chosen. Even though this cutoff is used by default, it is arbitrary.
One advantage of the group comparison is that you can consider variances for each gene to interpret the results. You normalize the results by comparing the between-group variance to the within-group variance.

However, the group comparison approach also has drawbacks. One issue is to find an appropriate criterion for defining the groups. As mentioned earlier, there is no universal cutoff for SUVmax with prognostic relevance, which could be used to stratify patients into groups. Thus, it is common practice to use the median SUVmax in the cohort to define tumors with SUVmax below average (group with low SUV) and tumors with SUVmax above average (group with high SUV). This cutoff ensures that the groups are of equal size but is otherwise arbitrary.

A drawback of using median SUVmax is that tumors with intermediate SUVmax are assigned to either high or low. This may weaken signals in the dataset. If both groups contain some tumors with intermediate SUVmax that can be expected to have a similar phenotype, it is more difficult to detect differences between the groups.
Another drawback is that the median depends on the composition of the cohort, since it is defined by the SUVmax distribution. Cohorts with different SUV spectra have different groupings. This results in different effect sizes for the same gene, as the fold change only refers to high versus low neglecting the absolute SUV values.

4.1.2.2 Linear Regression Approach to Overcome The Drawbacks Of Group Comparison

Given these drawbacks, we used a method other than group comparison. Since SUVmax and expression values are both continuous, we applied linear regression. To do this, we first set up the standards for interpreting the results.
When you are not comparing groups, you cannot calculate a fold change and need a different measure of effect size. The slope of the regression line is appropriate as it indicates how "steep" the correlation between gene expression and SUVmax is. To facilitate interpretation of the effect size, the slope can be represented as change in gene expression levels by a factor of 2. This describes how many SUV units are needed to obtain a doubling or halving of gene expression. The expression change by a factor of 2 is chosen arbitrarily and is based on the standard fold change cutoff of 2. Still, this conveys a good impression of how much gene expression varies with respect to SUVmax. It provides an estimate of the expected change in expression level when comparing two tumors.
To define correlations with high effect size, we transferred the standard fold change cutoff of 2 to a slope cutoff of 0.146. Thus, our results are more comparable with studies that used group comparison and calculated fold changes. However, this involves the aforementioned disadvantage of an arbitrary cutoff. To avoid this, we focused on the results of analyses that do not rely on a cutoff definition, such as GSEA.

4.1.2.2.1 Disadvantages of Linear Regression

A disadvantage of the approach without grouping is that it is less straightforward to account for variance of gene expression. Thus, to better interpret our results, we considered the distribution of outcomes and used the slope distribution to calculate Z-scores.
Another drawback of linear regression is that the exact SUVmax values are used, which means that this approach is more sensitive to variation of the values. In comparison, the accuracy of the values is less relevant in a grouping approach as the values are assigned to only two categories of "high" or "low". This implies that in linear regression the intrinsic variance of SUVmax has a greater impact. In addition, the bias due to the joint analysis of PET-CT and PET-MR and the bias due to small lesion size have a greater influence than in the approach based on group comparison.
As a last limitation of the linear regression approach, it could be discussed that it only tests whether the correlations are linear. This may not correspond to reality. If a linear model does not fit well, this results in a low r^2 and a high p-value. This indicates that there is no relation between gene expression and SUVmax – or that there may be relation other than linear. A nonlinear, more complex model could fit better. However, the use of more complex models increases the risk of overfitting. Due to the limited sample size, this was not an appropriate approach for our cohort. In addition, the clinical relevance of models with higher complexity is questionable. The goal of this study was not to accurately derive gene expression values

based on SUVmax, but to identify patterns in the dataset. Based on this, we could hypothesize which genes might explain elevated glucose uptake and may be associated with tumor malignancy.

4.1.2.2.2 Advantages of Linear Regression

The linear regression approach has advantages that basically overcome all the disadvantages of the group comparison approach.
In linear regression, as mentioned above, the exact SUVmax values are included in the analysis. Thus, more detailed information about the tumors is used than in binary grouping.
Furthermore, tumors with intermediate phenotype do not attenuate the signal but contribute to pattern recognition.
Another advantage over group comparison is that no SUVmax cutoff needs to be defined. Therefore, this analysis is less sensitive to the phenotypic composition of the cohort and the SUVmax spectrum. Moreover, the slope is independent of cohort composition and therefore comparable between studies. Taking into account the distribution of the results and calculating the Z-scores of the slopes, these give a good impression of how extreme an association is.

Given all this, we conclude that the advantages outweigh the disadvantages, and therefore we used linear regression.

4.2 Limitations of the Study

4.2.1 Limited Sample Size and Heterogeneous Dataset

Due to the low prevalence, studies of EwS have a small sample size [2]. Thus, validation on an external dataset is important. As radiogenomic analyses are not established for pediatric sarcoma yet, public datasets of imaging and expression data are not available. So, we searched for collaboration partners who could supply EwS patients for validation. To this end, we have evaluated several large treatment centers for EwS patients across Germany. We collaborate with Universitätsklinikum Münster that provides data from 7 EwS patients. Using this cohort, the results of this study will be validated in a follow-up project.

Another limitation is the heterogeneity in our cohort regarding clinical variables. To obtain a more homogeneous cohort we aimed to analyze subsets of our dataset stratified by risk factors. However, such a stratification was not feasible for our dataset. There are multiple prognostic factors in EwS such as age, ethnic background, localized or metastatic disease, primary site, tumor volume, response to therapy, primary disease or recurrence, in which the time to recurrence has also an influence [130, 131]. With this complex interplay of risk factors, it is futile to stratify patients just by one risk factor, and limited sample sizes do not allow to create subgroups with respect to multiple risk factors. To compensate for this, we checked for distribution of SUVmax values with respect to risk factors that were available for our dataset. Among the clinical variables, no SUVmax-related confounding factors were identified suggesting that different risk factors do not introduce an obvious bias into our data.

4.2.2 Skewed SUVmax Distribution

A third limitation is the skewed SUVmax distribution in our cohort. The sample with a very high SUVmax of 21.3 (sample_19) may influence the linear regression more than one of the other samples with SUVmax between 1.9 to 13.6. To evaluate the sensitivity of our findings towards sample_19, we compared the results from regression analysis with and without sample_19. We conclude that the sample with high SUVmax has an impact on the correlation results in some cases. The significant results rely on the presence of sample_19, while the findings with high effects size are mostly independent of it.
However, there is no reason to consider that the results based on the 18 samples are more correct than the results based on the 19 samples. For completeness, we have pointed out the problem that might be caused by the skewed SUVmax distribution in our cohort. To dispel doubts and to validate the results, we propose to repeat this analysis on a larger dataset with a better covered range of SUVmax values.

4.2.3 Importance of Robust Methods to Overcome Limitations

A consequence of our small, heterogeneous cohort are broad confidence intervals and that results like single significant relationships may be little generalizable. To compensate for these limitations, we focused on the results of robust methods like functional enrichment analyses.

4.3 Glycolytic Pathway Gene Expression Did Not Correspond with PET Signal Intensities

When correlating gene expression with glucose uptake, one would expect glycolytic gene expression to be upregulated with increasing SUVmax. Macpherson et al. [61] describe a frequent upregulation of glucose transporters in sarcoma, which results in raised SUV levels.
However, this was not observed in our study. Glycolytic gene expression pattern did not turn up in enrichment analysis, and glucose transporters showed only minor association with SUVmax. The member of the SLC2 family of glucose transporters with highest correlation with SUVmax was SLC2A1 with a slope of 0.061.
This observation is not unique to our study. PET-transcriptomic studies in lung and breast cancer got similar results [64-66]: only modest enrichments of glycolytic expression patterns were found, and glycolytic genes were not under the most significant correlations.
This indicates that PET imaging captures much more than glucose uptake alone.

4.4 Five Genes Significantly Associated with Glucose Uptake

We found 5 genes whose expression was significantly associated with SUVmax: PLSCR4, FAXDC2, NETO2, ELOVL2, and MYBL2, which are set into context regarding functionality in the following.

4.4.1 PLSCR4

PLSCR4 (phospholipid scramblase 4) is an organizer of the plasma membrane [149]. High expression is related to worse prognosis in endometrial and renal cancer [140]. However, in metastatic melanoma drugs are proposed that upregulate PLSCR4 as it is significantly downregulated there [150]. This implies an opposite role of PLSCR4 depending on cancer entity.
Our data showed downregulation of PLSCR4 in EwS tumors with increased SUVmax. Due to the prognostic role of SUVmax, this indicates that high PLSCR4 expression might be beneficial in EwS. This warrants future research on the possibility to apply drugs to upregulate PLSCR4 especially for patients with high SUVmax and poor prognosis.

4.4.2 FAXDC2

FAXDC2 (fatty acid hydroxylase domain containing 2) has oxidoreductase activity and binds iron ions [149]. It is associated with vascular diseases [149]. In addition, it plays a role in megakaryocyte differentiation [149, 151]. It has low expression levels in acute myeloid leukemia, where drugs are used to foster differentiation [151]. Additionally, FAXDC2 expression is prognostic for better outcome in renal and liver cancer [140].
In our cohort, FAXDC2 was negatively associated with SUVmax, which may indicate a prognostic role for better outcome in EwS. Furthermore, it suggests higher differentiation in EwS tumors with low SUVmax.

4.4.3 NETO2

NETO2 (neuropilin and tolloid like 2) is a transmembrane protein, a subunit of neuronal glutamate receptors [149]. NETO2 is upregulated in several cancer types like osteosarcoma, infantile hemangioma, hepatocellular carcinoma, nasopharyngeal carcinoma, colorectal, gastric, lung, pancreatic, and renal cancer [152, 153]. Its expression is associated with worse survival and/or metastasis [140, 152, 153]. Furthermore, NETO2 activates tumorigenic, stemness-related signaling pathways [152, 153].
In our analysis, we found a positive association of NETO2 expression and SUVmax in EwS. This may contribute to increased stemness in EwS tumors with high glucose uptake.

4.4.4 ELOVL2

ELOVL2 (ELOVL fatty acid elongase 2) is involved in fatty acid biosynthesis of membrane lipids and lipids for signaling [149]. Its role in cancer is controversial and depends on cancer type as it has a favorable prognostic function in breast cancer [140, 154], but promotes progression and worse outcome in renal cell carcinoma [155]. In general, Gimple et al. [156] summarize that fatty-acid synthesis is upregulated in different cancer types (glioma, breast, prostate, colon, and ovarian cancer) and often relates with worse prognosis. In particular, ELOVL2 is upregulated in glioma stem cells in glioblastoma, which is mediated by stem cell enhancers like SOX2 (SRY-box transcription factor 2) [156]. SOX2 expression in EwS was shown to increase malignancy [157].

In our EwS cohort, ELOVL2 was significantly positively correlated with SUVmax, and SOX2 was positively correlated with SUVmax but below our cutoff for high effect size (slope=0.12; 95% CI=[0.01; 0.22]; Z-score=2.42; r^2=0.21). We hypothesize that SOX2 upregulates ELOVL2, so that the correlation with SUVmax is stronger on downstream-levels, and that both genes may be associated with undifferentiated state in tumors with high SUVmax.

4.4.5 MYBL2

MYBL2 (MYB proto-oncogene like 2) is a TF – activator and repressor – regulating cell cycle progression, cell survival, and proliferation [149], and maintaining an undifferentiated state of cells [158]. High expression of MYBL2 is prognostic for poor outcome in renal, liver, and endometrial cancer [140]. Overexpression or amplification also correlates with poor outcome and metastatic ability in lung [159], prostate [160], and breast cancer [161]. It is an oncogene in ovarian cancer fostering tumor growth and metastases [162]. Due to the oncogenic role of MYBL2 across different cancer types, its inhibition is considered for treatment [158]. As no MYBL2 inhibitor is available, downstream inhibition of MYBL2 targets is considered, for example by using CDK2 (cyclin dependent kinase 2) inhibitors [158].
The role of MYBL2 in EwS was analyzed by Musa et al. [163], who found MYBL2 upregulation by EWS-FLI1. However, the intensity of MYBL2 expression did not depend on EWS-FLI1 levels but on the genetic composition of the EWS-FLI1 binding locus [163]. High expression was related to poor survival of patients, and MYBL2 silencing in EwS cell lines reduced proliferation rate [163]. However, high levels of MYBL2 sensitized EwS towards treatment with CDK2 inhibitors *in vitro* and *in vivo* [163].
In our cohort, we found upregulation of MYBL2 in EwS tumors with high SUVmax indicating poor prognosis. This replicates the finding by Musa et al. [163] that MYBL2 contributes to malignancy. As MYBL2 levels can be a biomarker for efficacy of CDK2 inhibitors [163], we speculate that this therapy may be most effective for EwS tumors with high SUVmax.

4.4.6 Significant Correlations Warrant Experimental Validation

In summary, these genes are candidates for experimental validation to uncover their mechanistic role in EwS. Furthermore, our results warrant future research on drugs upregulating PLSCR4 or inhibiting CDK2 downstream of MYBL2, especially for patients with strong PET signal and poor prognosis.

However, due to our limited sample size, we will rather focus on more robust methods such as the enrichment analyses.

4.5 Function of NPY Signaling Axis

In our enrichment analyses, the expression of genes in the NPY signaling axis and rhodopsin-like GPCRs were found to be decreased as glycolysis increased.

Generally, NPY signaling promotes inflammation [164-167] and differentiation of cells as rhodopsin-like receptors [168].
The role of NPY and its receptors in cancer is not completely understood. NPY receptors are overexpressed in different cancer entities [169]. Yet studies are sometimes contradictory and indicate cancer type-specific mechanism. For example, high NPY expression predicted poor prognosis in endometrial cancer [140]. In gastric cancer high expression of NPY indicated better survival [170] but high NPY1R poor survival [171]. In hepatocellular carcinoma an inverse pattern was found as NPY signaling via NPY5R promoted progression, and NPY5R expression correlated with shorter survival time [172], whereas NPY1R expression was a marker for longer survival and showed tumor inhibiting functionality [173]. Similarly, in breast cancer NPY signaling via NPY5R promoted migration and proliferation *in vitro* [174]. Though, NPY1R expression was a marker for longer survival [140]. Another study found high NPY5R expression to indicate better prognosis, activated apoptosis and cell cycle arrest [175]. NPY5R expression also sensitized the breast cancer cells to doxorubicin treatment [175].
In conclusion, these studies suggest a very context-specific role of NPY signaling in cancer types other than EwS.

NPY pathway expression and function in EwS has been studied as well [21, 176-184]. NPY and its receptors NPY1R and NPY5R are targets of EWS-FLI1, and therefore upregulated in EwS [180, 183, 185]. One study showed NPY signaling to foster bone metastasis *in vivo* [184]. However, this study did not examine different receptor types and their impact on NPY downstream effects. The pro-metastatic and proliferation signaling of NPY was conveyed by the receptors NPY2R and NPY5R [182, 185]. However, Tilan et al. [180] showed *in vitro* that NPY signaling via the receptors NPY1R and NPY5R promoted cell death.
A survival analysis of a publicly available EwS dataset (Savola dataset, n=44) on the R2 platform [186] showed significantly longer overall and event-free survival for tumors with high NPY1R or NPY5R expression.

In our cohort, the expression of NPY, NPY1R and NPY5R was negatively associated with SUVmax. We infer from that that there is not a uniform upregulation of these genes in all EwS tumors. Instead, the expression seems to be related to the glucose uptake of the tumor. We hypothesize that there is more NPY signaling promoting cell death in EwS tumors with low glucose uptake indicating less malignancy, possibly associated with neuroectodermal differentiation. A beneficial effect on survival of NPY1R and NPY5R expression could be replicated in the Savola dataset.

In conclusion, more studies are needed to elucidate the context specific role of NPY signaling in cancer in general and EwS in particular. Our analysis in the context of the current literature proposes a prognostic and favorable role of NPY signaling through NPY1R and NPY5R in EwS tumors with low glucose uptake.

4.6 Discussion of GSEA Results

In addition to the enrichment analysis for individual genes with high effect size, the enrichment analysis for all genes using GSEA provides a broader overview.

4.6.1 General Cellular Processes Reflecting the Hallmarks Of Cancer

Several enrichments from GSEA with high SUVmax represent basic functions such as cell cycle, DNA replication and repair, transcription, cytoskeleton, actin-myosin interaction, muscle and muscle development. These processes seem to be upregulated in cells relative to increased glucose uptake and reflect raised metabolic activity and the hallmarks of cancer. Such findings are consistent with other radiogenomic studies [63, 64, 66-68].

4.6.2 Expression Levels Regulated by the Activity of Transcription Factors Correlating With Glucose Uptake

GSEA also indicated 3 TFs with targets enriched at high SUVmax: RNF2 (ring finger protein 2), the E2F family of TFs, and TCF3 (transcription factor 3).

4.6.2.1 Transcription Factor RNF2

RNF2 is a core component, namely the catalytic subunit, of the Polycomb repressive complex 1 (PRC1) belonging to the Polycomb group (PcG) [149]. PcG genes are crucial developmental genes and epigenetically maintain the transcriptionally repressive stemness state [19, 22, 149, 187, 188]. As such, RNF2 plays a role in cancer development and proliferation, and its expression correlates with progression, invasiveness, recurrence and sensitivity to therapies [189]. Its expression is a predictive marker for poor survival in renal and liver cancer [140]. However, RNF2 can also have contradictory roles as high expression in basal breast cancer indicates worse survival, but in ER+ breast cancer it indicates longer survival [189]. Furthermore, RNF2 affects the tumor microenvironment and suppresses antitumor immune responses [190]. In short, RNF2 is discussed as promising target for therapies in tumor entities other than EwS [189].

Our analysis suggests a role of RNF2 in EwS as well. As its targets are mostly upregulated in tumors with high glucose uptake, we hypothesize that the activity of RNF2 correlates with metabolic activity, and is prognostic for worse outcome. Based on this, we suggest further studies to elucidate the role of RNF2 in EwS and therapeutic options.

4.6.2.2 E2F Family of Transcription Factors

The E2F family of TFs is studied for more than 30 years but the mechanisms are not completely understood yet [191]. The family members have various functions in cell cycle control – like in DNA replication, proliferation, apoptosis, differentiation and stemness – and are linked to energy metabolism enhancing glycolysis [191-194]. E2F is upregulated in response to DNA damage as it maintains genomic stability in DNA damage repair [194].

Most of the E2F TFs are active across all cancer entities [194]. High expression of certain family members is prognostic for poor survival in liver, endometrial, renal, pancreatic, bladder, and breast cancer [140, 191] but prognostic for longer survival in cervical, thyroid, and ovarian cancer [140]. However, in general E2Fs foster invasiveness, angiogenesis, and drug resistance [191-194]. This can be due to the fact that E2F target genes maintain stemness features, and thereby E2F contributes to the characteristics of cancer stem cells, which are associated with treatment failure [191]. Thus, the E2F TFs are discussed as targets for therapies [194],

although no specific drugs are known that directly inhibit E2F [191]. An alternative is to inhibit upstream regulators of E2F, like inhibiting CDK4 (cyclin dependent kinase 4) and CDK6 (cyclin dependent kinase 6) [194]. Many studies are ongoing on CDK inhibition, and some inhibitors are FDA approved drugs for treatment of advanced breast cancer: palbociclib, ribociclib and abemaciclib [194]. These drugs have an effect on E2F and its target genes, as in breast cancer resistant to palbociclib no changes in E2F target gene expression were observed [194]. Hence, E2F and its target genes indicate sensitivity towards CDK inhibitors [194].

Just as in other cancer entities, E2Fs are active in EwS. EWS-FLI1 activates the expression of the E2F family, except of E2F4 [195]. Furthermore, Schwentner et al. [195] describe a clear overlap of E2F and EWS-FLI1 binding sites in the promotors of target genes. Normally, E2F4 is bound to the promotors and represses transcription. However, in presence of EWS-FLI1, E2F4 is replaced by a E2F3 having activating function, which leads to increased expression of E2F target genes [195].
Therefore, studies were conducted for EwS treatment by inhibiting E2F upstream regulators using CDK inhibitors. In general, CDK inhibition is no standard treatment in EwS. However, Bailey et al. [196] describe CDK4/6 inhibitors to be of high priority for testing in clinical trials. *In vivo* data showed longer survival and reduced tumor growth under CDK4/6 inhibition. So far, CDK4/6 inhibitors have been tested in completed or ongoing phase I trials in children showing beneficial effects [196-198]. Yet, an appropriate combination treatment has to be identified [196].
In our analysis, we saw that E2F targets were enriched among genes that were upregulated in tumors with high metabolic activity. As E2F enhances glycolysis and was shown to promote aerobic glycolysis [199], this may be the reason for the observed PET signal. We hypothesize that E2Fs are more active in EwS tumors with higher level of malignancy. Since indirect targeting of E2F by CDK inhibitors showed effects in EwS treatment trials, we speculate that CDK inhibition may be beneficial especially for patients who suffer from EwS with high SUVmax and poor prognosis.

4.6.2.3 Transcription Factor TCF3

TCF3 is a helix-loop-helix TF with context and cell lineage specific functionality [200]. It recruits different proteins, and dependent on this protein partner TCF3 has activating or repressive effects on target gene transcription [201]. As such, TCF3 can act as tumor suppressor fostering differentiation, or as tumor promotor fostering proliferation [201]. Its function is cancer type and context specific: high expression of TCF3 is prognostic for better survival in colorectal cancer [200, 201]; however, another study found TCF3 as tumor promotor and/or prognostic marker in colorectal cancer [202], as well as in renal, liver, breast, gastric, and prostate cancer, lymphoma, cervical squamous cell carcinoma and hepatocellular carcinoma [140, 200, 201]. Furthermore, TCF3 is associated with mesenchymal to epithelial transition [149, 201, 202] and stemness as it is highly expressed in stem cells and poorly differentiated cancers [203].

Despite its role in various cancer types, TCF3 has not been studied in EwS so far. Our radiogenomic analysis showed that the targets of TCF3 are mostly upregulated in EwS tumors with high glucose uptake.
Based on our findings, we hypothesize that the activity of TCF3 correlates with metabolic activity, and is prognostic for worse outcome in EwS. Considering its role in multiple cancer entities, our results suggest further research to uncover the mechanism of TCF3 action in EwS.

4.7 Spectrum of Differentiation to Stemness

Strikingly, many of our findings have a link to stemness or differentiation. 4/5 genes that were significantly associated with SUVmax in our EwS cohort indicate a link between stemness or differentiation and SUVmax. Furthermore, the 3 TFs (RNF2, E2F family, TCF3), whose targets were enriched among the genes that were positively correlated with SUVmax, are associated with maintaining stemness. Additionally, based on the findings of the NPY axis, we speculate about potential neuroectodermal and endothelial differentiation of cells at low SUVmax. Similar findings were reported in studies on esophageal cancer [49] and lung carcinoma [124], where the authors found a correlation of high SUVmax and poorly differentiated tumors.

A stem-cell like phenotype is a basic characteristic of EwS, and maintaining stemness plays an important role [15-17, 19, 22]. Our group [19, 22] and Sheffield et al. [26] found that EwS tumors exhibit a spectrum of stemness varying from stem-like towards mesenchymal, neuroectodermal or endothelial differentiation. However, correlation analyses with outcome were not performed in these studies. Another group, Sheffield et al. [26], showed that tumors with STAG2 mutations have a tendency towards stem-like characteristics, and STAG2 mutations are associated with worse prognosis [2, 3]. An analysis of stemness and prognosis was conducted by Stahl et al. [18], who investigated immune infiltrates in EwS, and found an association of stemness genes (HIF1A expression) and unfavorable immune cell infiltration (positive correlation with M2 macrophages and neutrophils, negative correlation with T cell infiltration). Furthermore, they described a trend of high HIF1A expression – reflecting stemness – and shortened overall and event-free survival in EwS [18].
This is in concordance with the general observation that stemness in cancer is predominantly associated with poor prognosis [204-206].

In our analyses, we used SUVmax as indicator for malignancy and outcome. Given our findings described above, we hypothesize that the spectrum of stemness to differentiation, which was described for EwS tumors by Sheffield et al. [26], is reflected in the SUVmax of the tumor. While stemness may be predominant in EwS tumors with high glucose uptake and seems to indicate increased malignancy, differentiation may be associated with low glucose uptake indicating better prognosis.

4.8 Spectrum between Immune Activation and Immune Suppression

Several analyses we have performed suggest a relationship between glucose uptake and immune signatures. To our knowledge, there are no studies on the relationship between immune response and PET in EwS. So far, there is only one study in soft tissue sarcoma that observed a positive association of macrophage infiltration and SUVmax [135]. We were able to replicate this correlation in our EwS cohort, as we found a positive correlation between M2 macrophage signature and SUVmax. In addition, we found a negative correlation between CD8 T cell infiltration and SUVmax. One has to consider that some therapies, e.g. allogeneic stem cell transplantation, have an effect on T cell infiltration. Since not all clinical data were available, we cannot exclude whether this might be the case in some of our patients.
However, a limitation in this analysis is that the absolute immune cell infiltration is very low in our cohort as in EwS in general.

The detected correlations of DC8 and M2 infiltration and glucose uptake match with our results from additional analyses indicating a link to immune signatures. NPY signaling, which correlated with low SUVmax, acts proinflammatory. We speculate that a NPY induced proinflammatory microenvironment may favor CD8 infiltration. In contrast, RNF2, whose target genes were predominantly upregulated at high SUVmax, suppresses an antitumor immune response [190]. Taken together, we hypothesize that immune response may be associated with low SUV and immunosuppression with high SUV.

4.8.1 Activated T Cells and Macrophages are PET Active

However, this hypothesis contradicts the observation that inflammation is PET active. Activated T cells have an increased metabolism and thus contribute to measured FDG uptake [207, 208]. The same holds for macrophages, which could strengthen the measured positive correlation with SUVmax [135]. In short, both the infiltrates we observed at low SUVmax and high SUVmax could amplify the PET signal. However, we think this has only minor impact in our data. The sum of all infiltrated immune cells is maximal 2.5% of cells in the bulk sample, much lower than the number of tumor cells. We assume that the signal measured in PET is predominantly due to the tumor cells.

4.8.2 Hypotheses for Characteristic Absence of T Cell Infiltration in EwS

In general, EwS is considered as immunogenic "cold” [18]. Stahl et al. [18] discuss two hypotheses regarding the scarcity of T cell infiltration in EwS.
The first hypothesis is based on the genetic homogeneity of EwS. The low mutational burden results in few neoantigens that could be recognized by T cells [18, 138]. Therefore, the number of T cells in the tumor is low and immune response is poor [18]. This should apply in general to all EwS tumors. However, we observe variation of CD8 T cell infiltration with respect to SUVmax. Further analyses should examine whether there is a correlation between mutational load and SUVmax. Otherwise, this hypothesis (low mutational burden makes cold tumors) may contribute to the low T cell infiltration in general but cannot explain the pattern observed in our cohort.
The second hypothesis is that hypoxia has an effect on immune cells [134]. Hypoxia fosters the attraction of macrophages to tumor tissue and the development of M2 macrophages [134, 209]. It also promotes features of aggressive tumors and glycolytic reprogramming [134]. We can detect this altered metabolism – called Warburg effect – in PET and observe that increased

SUV correlates with poor survival. Two factors affect T cell infiltration [18]. On the one hand, an environment develops that excludes effector T cells and suppresses their function. Thus, there is no anti-tumor response and the "cold" state is maintained [18, 134, 209]. On the other hand, T cells and tumor cells compete for energy supply [18, 134]. This is consistent with our observation that tumors with high energy consumption, i.e., high glucose uptake, present with lower CD8 T cell infiltration. Thus, our results – the relation of CD8 infiltration and SUVmax – support the second hypothesis. This implies that one could expect hypoxia in tumors with high SUVmax. This is consistent with our finding of a positive correlation between SUVmax and M2 infiltration, which is related to hypoxia as described above. However, we do not find a correlation of hypoxia with glucose uptake. The effects of hypoxia are mainly mediated by HIF1A [18, 210]. In our data, the expression of HIF1A is independent of SUVmax (slope 0.003 in linear regression, i.e., doubling of gene expression over 291 SUV units). Furthermore, the term hypoxia does not pop up in any of our enrichment analyses. This suggests that the mechanism can be independent of hypoxia and proceed, for example, directly via glycolytic reprogramming. Further analysis in larger cohorts is needed to confirm the underlying mechanism.

4.8.3 Link Between Immune Infiltration, Hypoxia and Stemness

There is also a link between hypoxia and stemness [209], for example, HIF1A mediates hypoxia but also represents stemness [18].
All of this is reflected in PET. We found stemness in tumors with high glucose uptake, both of which indicate a poor prognosis. Immune infiltration, which is prognostic as well, was also associated with glucose uptake. However, against expectations, there was no association between hypoxia indicating worse survival and glucose uptake. This suggests that hypoxia may be not a necessary mechanism. Other mechanisms such as glycolytic reprogramming might be sufficient for the observed relations.

In short, we summarize from the literature and our observations that there seems to be an association between stemness and immunosuppression, both of which are reflected in PET signal. We hypothesize that these mechanisms might contribute to the prognostic role of PET imaging.

5 Conclusion

This study is the first radiogenomic study in EwS relating large-scale transcriptomic data and functional radiomic data from PET imaging. Since EwS tumors are uniform at the genomic level, we assessed gene expression. To investigate the information content of transcriptomic and radiomic data and whether they complement each other, we performed prediction studies on tumor entity, genetic fusion type, and outcome. For outcome prediction in our cohort, the integration of transcriptomic plus radiomic was equivalent compared with prediction based on either transcriptomic or radiomic data alone.
Furthermore, we characterized EwS tumors in terms of their glucose uptake measured as SUVmax in PET, which correlates with the level of malignancy. With this, we aimed to elucidate the transcriptomic basis for the variant glucose uptake and to identify novel prognostic biomarker candidates.
Due to the low incidence of EwS, we focused on the results of enrichment analyses, as these are more robust to single false positive findings. Thus, we identified correlations between SUVmax and infiltration of immune cells, neuroectodermal signaling pathways, and stemness-related transcription factors. We hypothesize that stemness is associated with increased glucose uptake and poor prognosis. Furthermore, increased differentiation may indicate better prognosis correlating with low glucose uptake.
These findings warrant prospective validation in an external cohort, as well as in experimental and clinical settings. They may eventually provide a better risk stratification of patients and open up new therapeutic opportunities.

The present study tested the potential of a radiogenomic approach in EwS. Large-scale analyses are not yet established in clinical practice. Therefore, it is necessary that they will be part of the clinical routine in the future. However, clinical procedures are often not suited for such analyses. Out of 75 patients referred to our institution during the duration of the study, only 19 met all of our quality standards. Thus, if we are to reap the benefits of large-scale analyses in the future, clinical practice must be adapted to the requirements of large-scale analyses.

6 Appendix

6.1 Supplementary Figures and Tables

Patient	ID	Sex	Genetics	Disease State	Sample Type	Therapy	Age	Event death (0=no, 1=yes)	Overall survival (months)	Overall survival group	PET modality	SUVmax	SUV group
patient_1	sample_1	male	EWS-FLI1	Primary Disease	Tumor	Untreated	10	0	27	long survival	PET-MR	1,898	low
patient_2	sample_2	male	EWS-FLI1	Relapse	Metastasis	Treated	12	1	18	short survival	PET-CT	2,019	low
patient_3	sample_3	male	EWS-FLI1	Relapse	Metastasis	Untreated	28	0	1	na	PET-CT	2,641	low
patient_4	sample_4	female	EWS-ETV1	Relapse	Metastasis	Untreated	6	0	24	long survival	PET-CT	2,685	low
patient_5	sample_5	female	EWSR1	Relapse	Metastasis	Treated	18	1	21	short survival	PET-CT	2,756	low
patient_6	sample_6	male	EWS-FLI1	Relapse	Metastasis	Untreated	17	1	32	long survival	PET-CT	2,889	low
patient_4	sample_7	female	EWS-ETV1	Relapse	Metastasis	Treated	7	0	19	na	PET-CT	4,323	low
patient_7	sample_8	female	EWS-FLI1	Relapse	Metastasis	Untreated	31	1	21	short survival	PET-CT	4,567	low
patient_8	sample_9	male	EWS-FLI1	Primary Disease	Tumor	Untreated	10	0	34	long survival	PET-MR	5,084	low
patient_9	sample_10	female	EWS-FLI1	Relapse	Tumor	Untreated	24	1	22	short survival	PET-CT	5,387	high
patient_10	sample_11	female	EWS-FLI1	Relapse	Metastasis	Untreated	3	1	3	short survival	PET-CT	6,046	high
patient_11	sample_12	female	EWS-FLI1	Relapse	Metastasis	Untreated	17	0	37	long survival	PET-CT	7,08	high
patient_12	sample_13	male	EWS-ERG	Primary Disease	Metastasis	Treated	14	1	21	short survival	PET-MR	7,391	high
patient_13	sample_14	female	EWS-FLI1	Relapse	Metastasis	Untreated	27	1	10	short survival	PET-CT	8,517	high
patient_12	sample_15	male	EWS-ERG	Primary Disease	Metastasis	Treated	15	1	21	short survival	PET-MR	9,553	high
patient_14	sample_16	female	na	Relapse	Metastasis	Untreated	18	1	7	short survival	PET-CT	9,617	high
patient_15	sample_17	male	EWS-FLI1	Primary Disease	Tumor	Untreated	5	0	76	long survival	PET-CT	12,26	high
patient_16	sample_18	female	EWS-FLI1	Relapse	Metastasis	Treated	8	1	2	short survival	PET-CT	13,578	high
patient_17	sample_19	female	EWS-FLI1	Relapse	Tumor	Treated	11	1	6	short survival	PET-CT	21,269	high

Suppl. Table 1: Patients' metadata including clinical data, genetic data, tissue data, survival data and image data. *na=not available/ not applicable.*

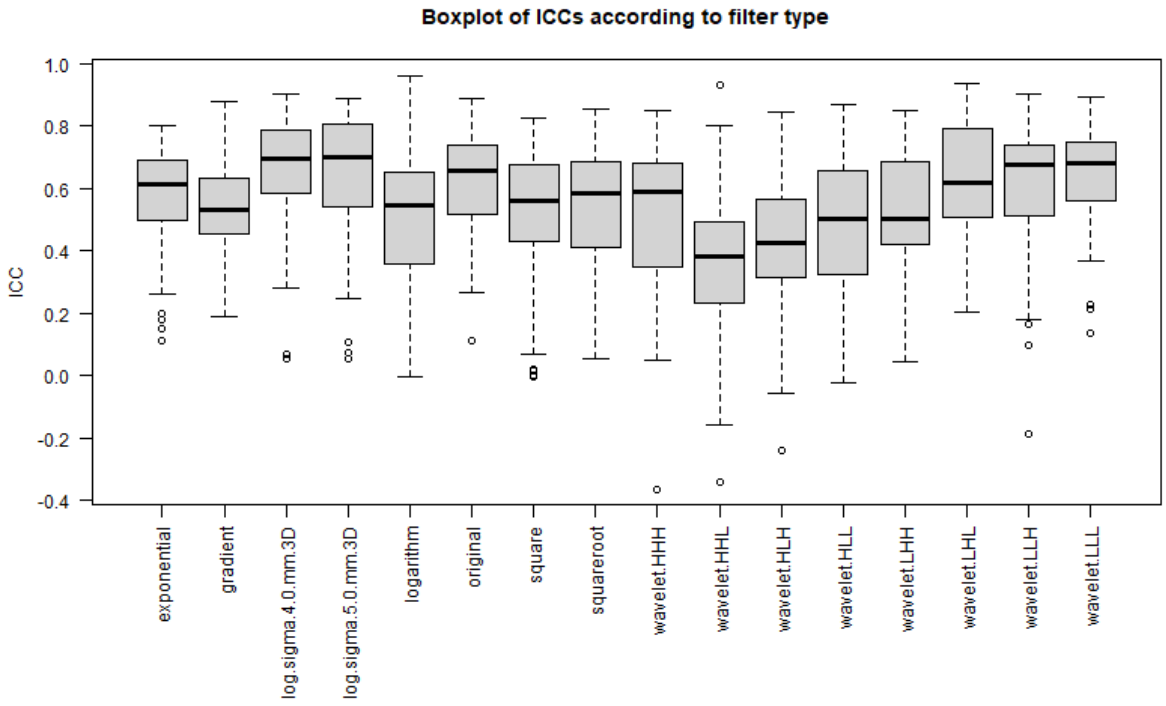
(A)

Feature class	Description	Number of IFs
shape	3D size and shape description (based on mask neglecting voxel values; only for original image)	14
first-order	Histogram based intensity statistics	18
glcm	Gray level co-occurrence matrix features	24
glrlm	Gray level run length matrix features	16
glszm	Gray level size zone matrix features	16
gldm	Gray level dependence matrix features	14
ngtdm	Neighboring gray tone difference matrix features	5
		107

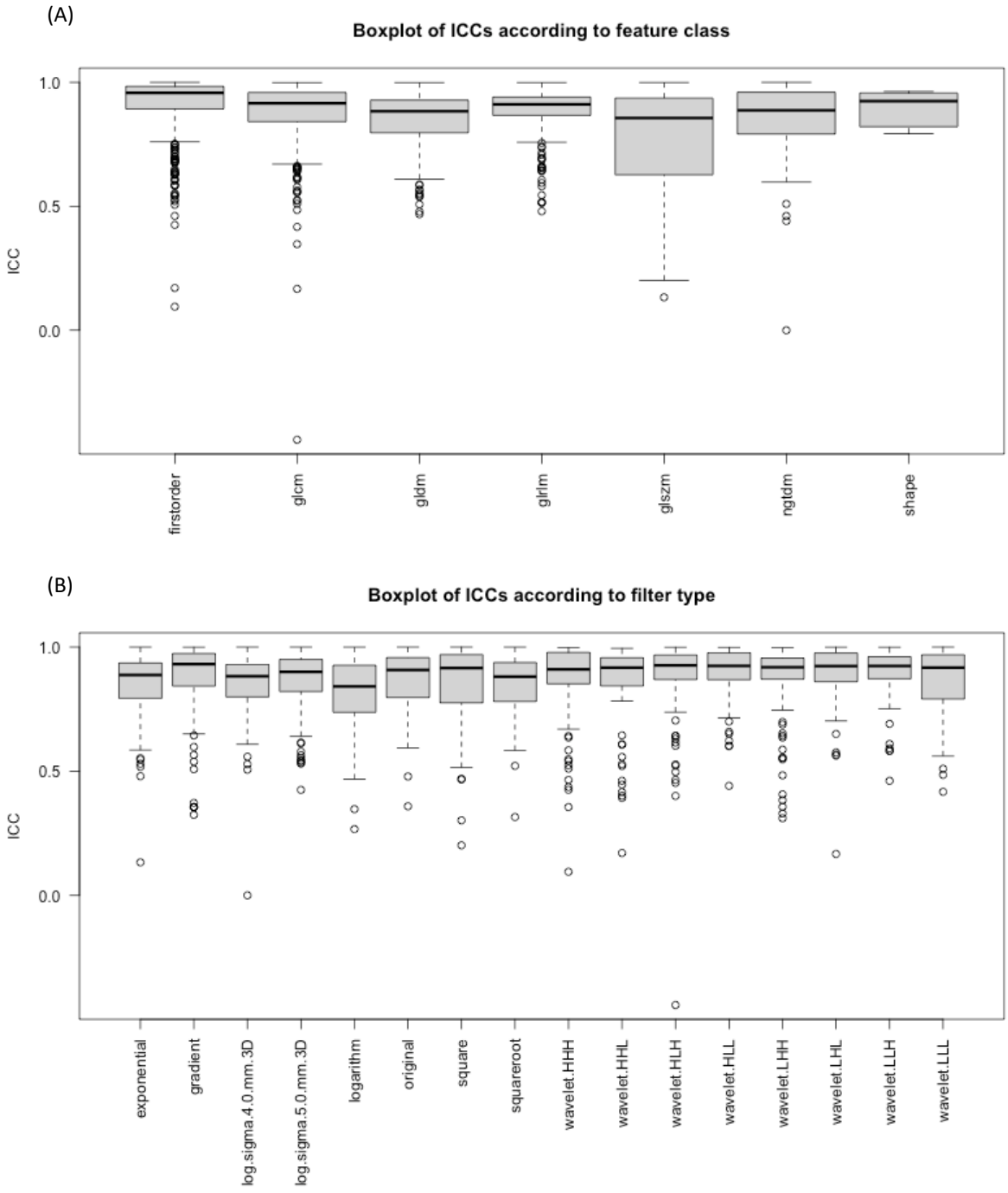
(B)

Filter type	Description	Number of IFs
original	Original image without filter	107
log-sigma-4-0-mm-3D	Laplacian of Gaussian filter, sigma 4 (emphasizing finer textures)	93
log-sigma-5-0-mm-3D	Laplacian of Gaussian filter, sigma 5 (emphasizing coarser textures)	93
wavelet-LLH	Wavelet filter, decomposition low-low-high	93
wavelet-LHL	Wavelet filter, decomposition low-high-low	93
wavelet-LHH	Wavelet filter, decomposition low-high-high	93
wavelet-HLL	Wavelet filter, decomposition high-low-low	93
wavelet-HLH	Wavelet filter, decomposition high-low-high	93
wavelet-HHL	Wavelet filter, decomposition high-high-low	93
wavelet-HHH	Wavelet filter, decomposition high-high-high	93
wavelet-LLL	Wavelet filter, decomposition low-low-low	93
square	Squaring and scaling of voxel intensities	93
squareroot	Applying square root and scaling of voxel intensities	93
logarithm	Logarithm of absolute voxel intensities +1 and scaling	93
exponential	Applying exponential (absolute intensities) and scaling	93
gradient	Returning magnitude of local gradient	93
		1502

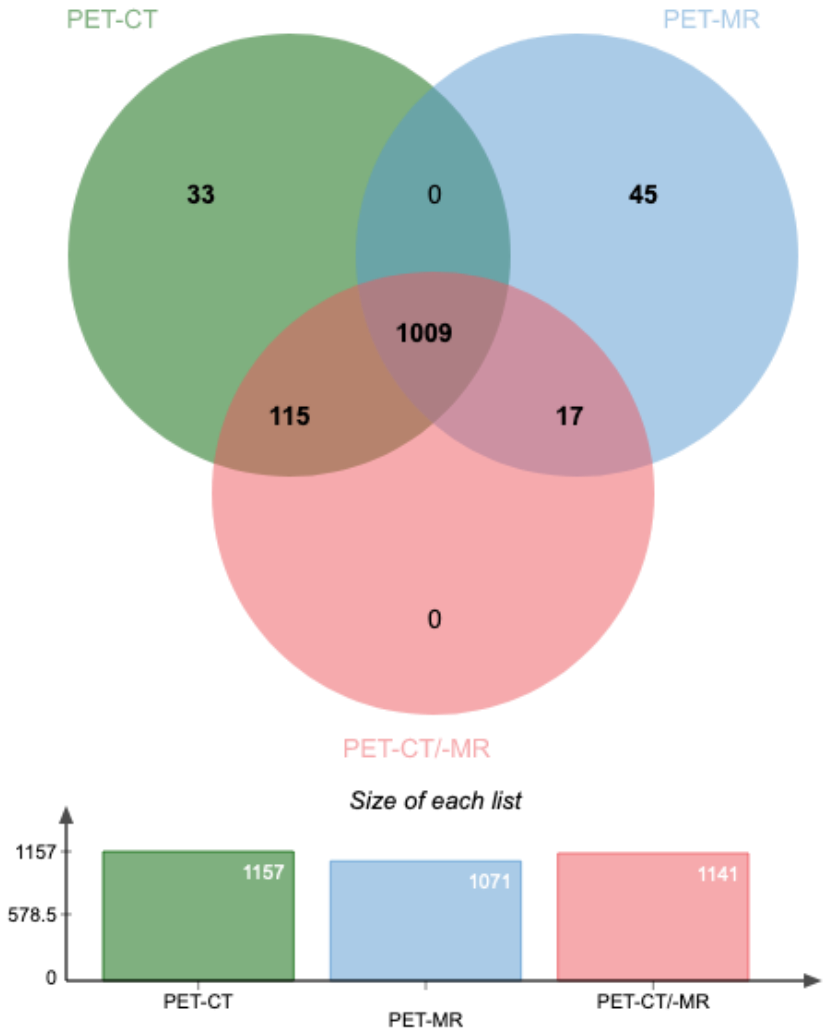
Suppl. Table 2: Overview and short description of image features (IFs).*(A) Overview of 7 feature classes. For each class, the number of contained IFs is given. Altogether, there are 107 distinct IFs: 14 shape features and 93 first- and second-order (glcm, glrlm, glszm, gldm, and ngtdm) IFs. (B) Overview of image filter types: original image and 15 filter types to obtain derived images. For each filter type, the number of IFs is given. Shape features are only calculated on the original image together with first-order and second-order IFs, which yield 107 IFs. On derived images, first-order and second-order IFs are calculated yielding 93 higher-order IFs, respectively. Altogether, there are 1502 IFs.*



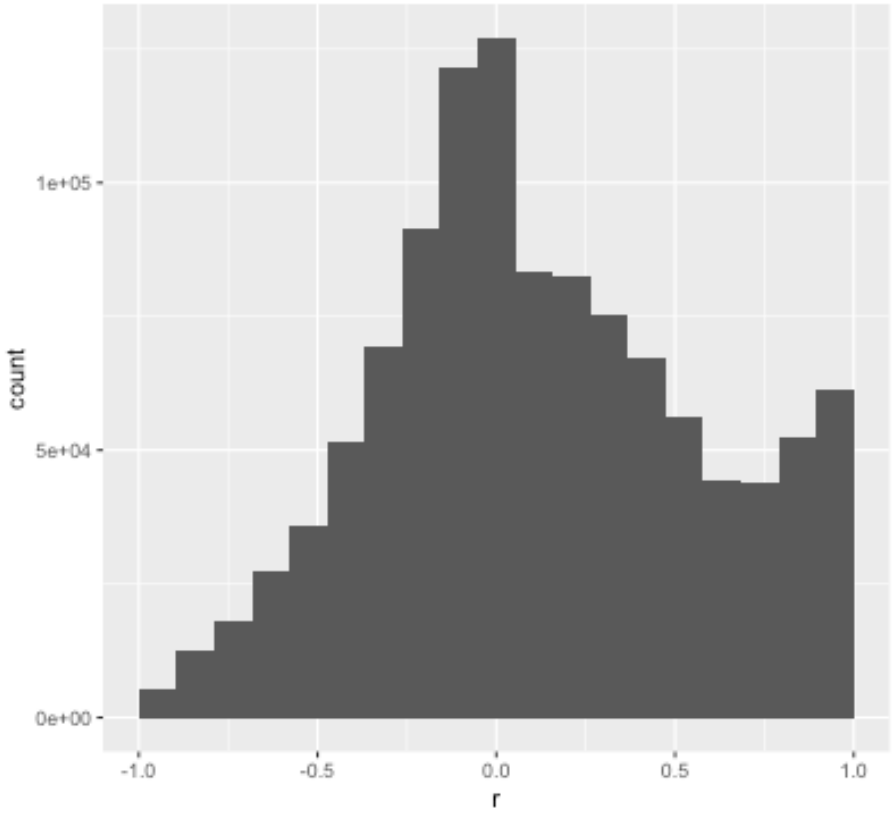
Suppl. Figure 1: ICC values distribution to analyze robustness to imaging modality (PET-CT versus PET-MR).*The ICC distributions are split by filter type. Median ICC varies between 0.4 (wavelet.HHL filter) and 0.7 (LoG filters).*



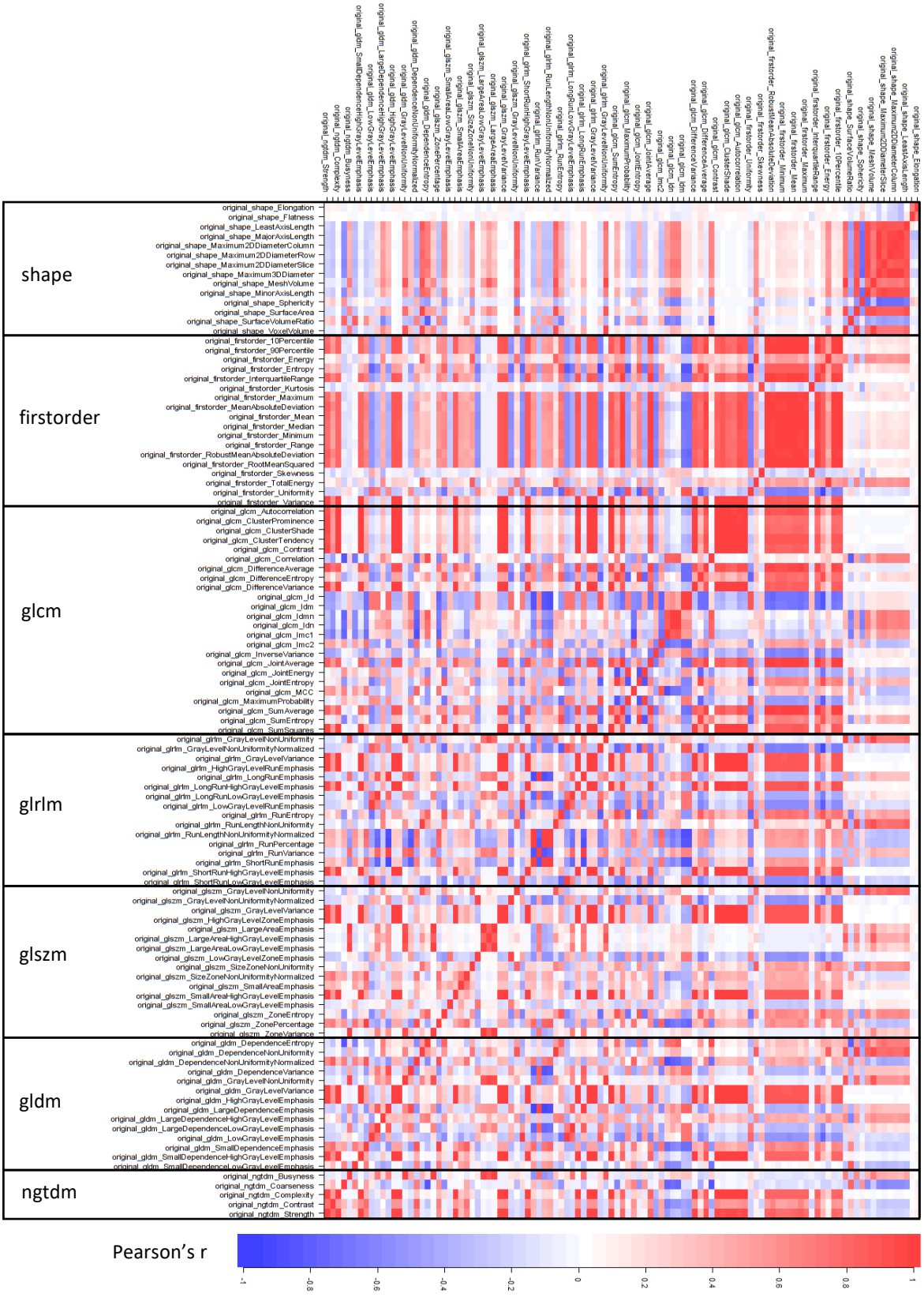
Suppl. Figure 2: ICC values distribution to analyze robustness to delineation variability.*Lesion segmentation was based on 20%, 30%, and 40% of SUVmax, and examined in PET-CT series (n=106). (A) The ICC distributions are split by filter type. For all filter types, ICC distributions are in a similar range with median ICC around 0.9, except for the logarithm filter. This shows that the filter type has minor impact on the analysis of delineation sensitivity. (B) The ICC distributions are split by feature class. All feature classes, except glszm features, achieved consistently high ICC values (median ICC around 0.9).*



Suppl. Figure 3: Venn diagram depicting the overlaps of delineation-robust IFs identified in PET-CT series (green), in PET-MR series (blue), or in a mixed dataset (red).*1009 IFs are stable in all 3 datasets. Below, the total number of delineation-robust IFs per dataset is given: 1157 for our PET-CT series, 1071 for our PET-MR series, and 1141 for our mixed dataset.*



Suppl. Figure 4: Histogram of pairwise correlation coefficients (Pearson's r) for all 1502 IFs.*IFs originate from 160 PET series (PET-CT and PET-MR) based on 40% SUVmax segmentation. Though most IFs are uncorrelated (r around 0), a portion of IFs shows high correlation between 0.8 to 1, which indicates redundant information content.*



Suppl. Figure 5: Correlogram depicting pairwise correlation of all 107 IFs on original images (160 PET series). *Shape features show high intraclass correlation but no correlation with features of other classes. First-order features show high intraclass correlation but also some correlations with features of second order. Except for shape features, there is no obvious partition by feature class.*

(A)

Feature class	Number of non-redundant IFs/ total number of IFs in this class	% non-redundant of total IFs in this feature class
shape	5/14	35.7%
first-order	46/288	16%
glcm	74/384	19.3%
glrlm	18/256	7%
glszm	94/256	36.7%
gldm	65/224	29%
ngtdm	19/80	23.8%
	321/1502	

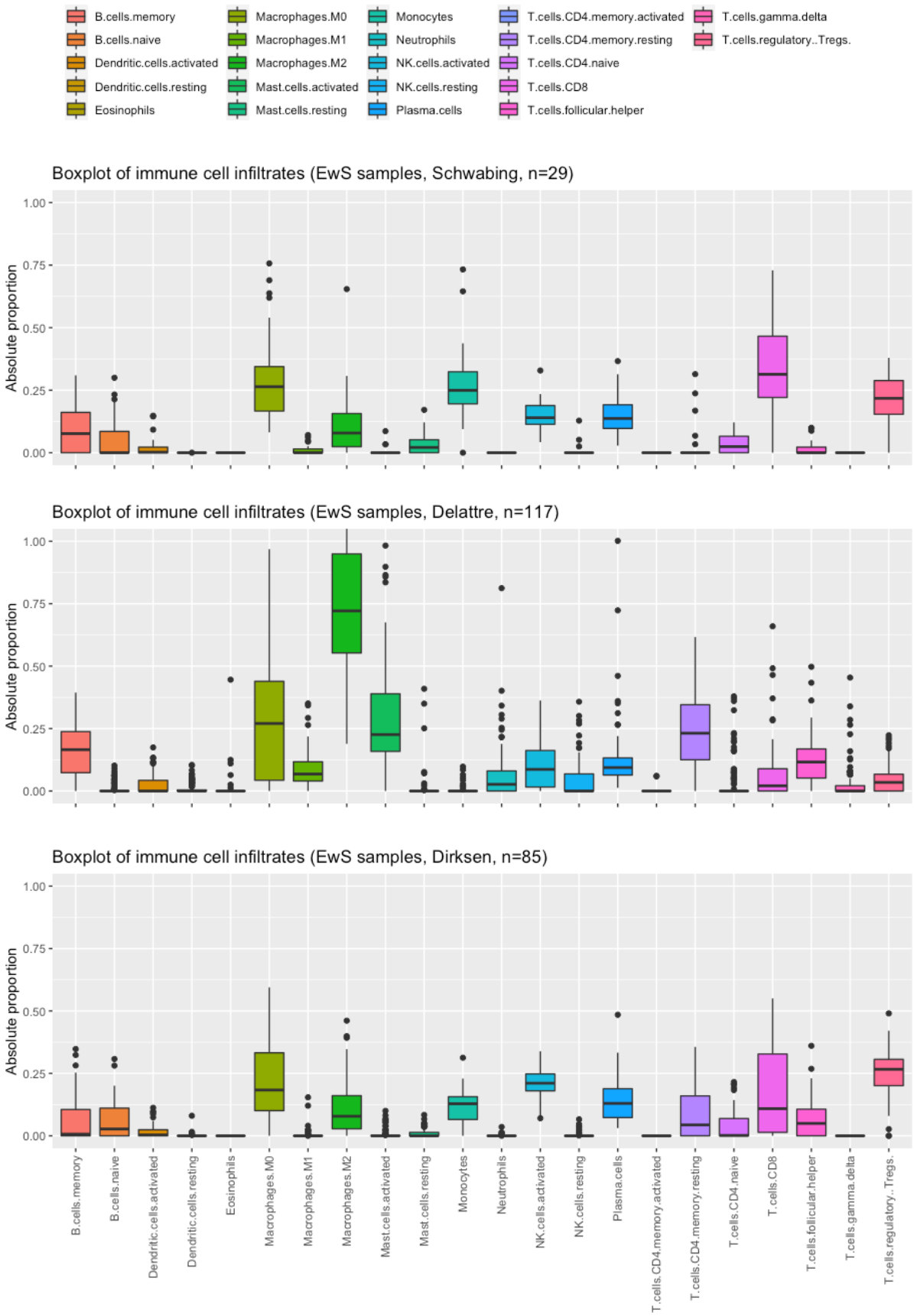
(B)

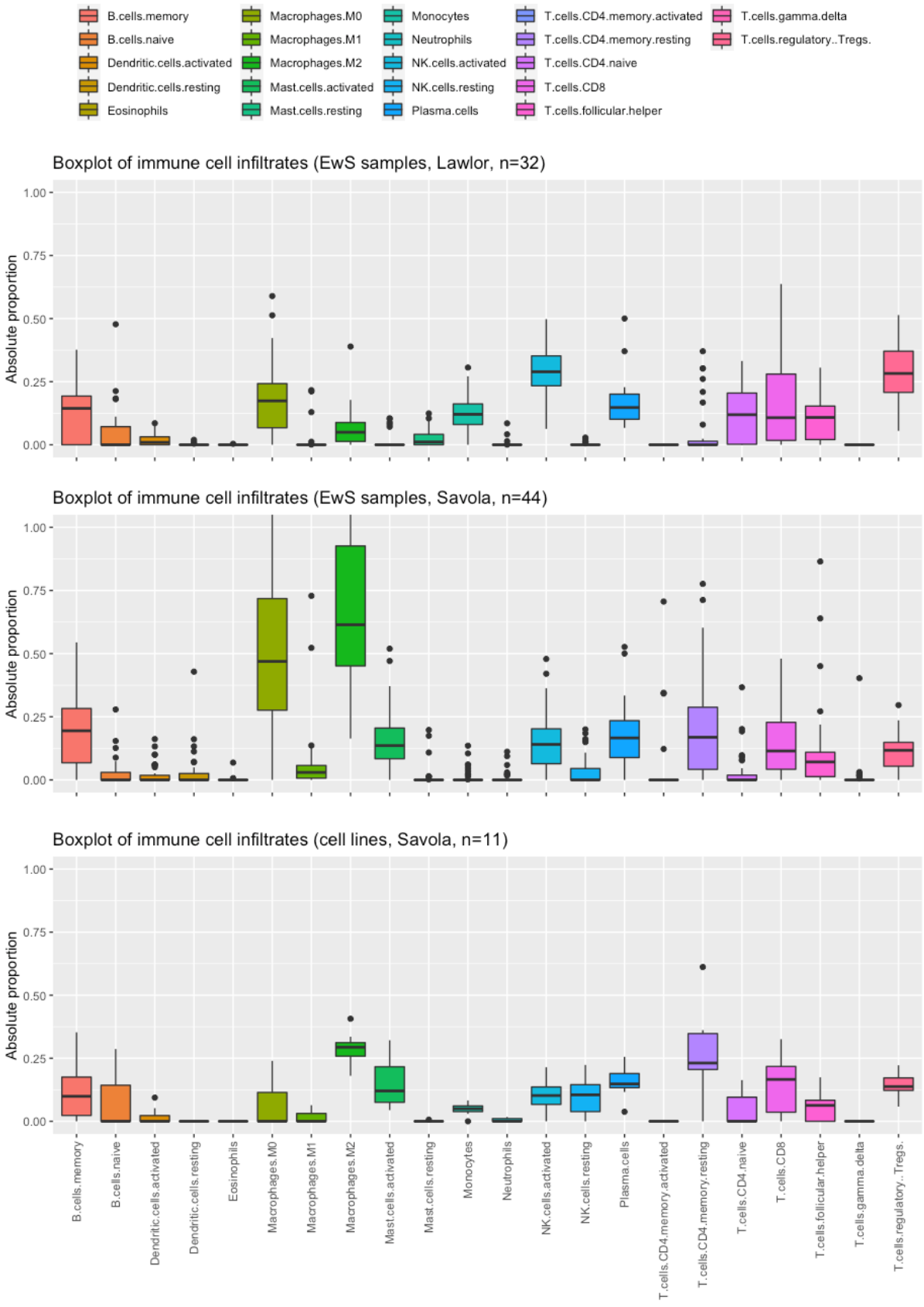
Filter type	Number of non-redundant IFs/ total number of IFs for this filter	% non-redundant of total IFs for this filter type
original	16/107	15%
log-sigma-4-0-mm-3D	18/93	19.4%
log-sigma-5-0-mm-3D	17/93	18.3%
wavelet-LLH	25/93	26.9%
wavelet-LHL	20/93	21.5%
wavelet-LHH	21/93	22.6%
wavelet-HLL	21/93	22.6%
wavelet-HLH	20/93	21.5%
wavelet-HHL	22/93	23.7%
wavelet-HHH	27/93	29%
wavelet-LLL	12/93	12.9%
square	19/93	20.4%
squareroot	13/93	14%
logarithm	25/93	26.9%
exponential	18/93	19.4%
gradient	27/93	29%
	321/1502	

Suppl. Table 3: Overview of non-redundant IFs.*321/1502 IFs are non-redundant, which means they have pairwise absolute correlations ≤0.9. (A) Number of non-redundant IFs split by feature class. (B) Number of non-redundant IFs split by filter type.*

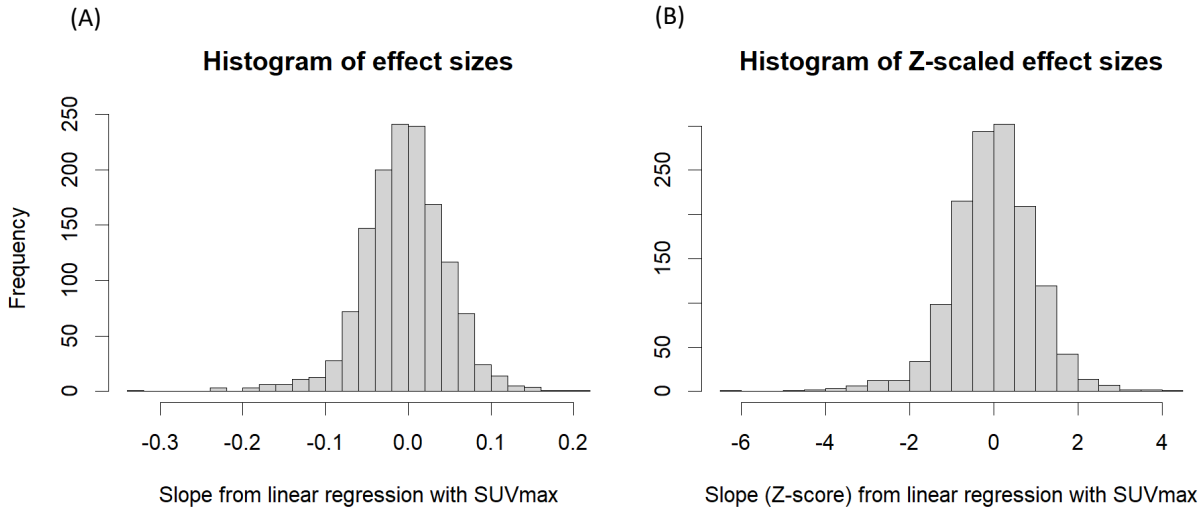
Immune cell type	Relative infiltration	Standard deviation
T cells CD8	18.0%	0.082
Macrophages M0	16.2%	0.088
Monocytes	14.6%	0.068
T cells regulatory (Tregs)	11.9%	0.048
Plasma cells	8.1%	0.041
NK cells activated	8.1%	0.033
Macrophages M2	6.1%	0.066
B cells memory	5.4%	0.057
B cells naive	2.9%	0.043
T cells CD4 naive	2.1%	0.024
Mast cells resting	1.8%	0.023
T cells CD4 memory resting	1.5%	0.040
Dendritic cells activated	1.1%	0.021
T cells follicular helper	0.9%	0.017
Macrophages M1	0.6%	0.011
NK cells resting	0.3%	0.013
Mast cells activated	0.3%	0.010
Dendritic cells resting	0.0%	0.000
T cells CD4 memory activated	0.0%	0.000
T cells gamma delta	0.0%	0.000
Eosinophils	0.0%	0.000
Neutrophils	0.0%	0.000

Suppl. Table 4: Immune cell infiltration in our cohort (Schwabing, n=29) deconvoluted by Cibersortx.*Relative numbers with regard to total infiltrating immune cells are given. The most abundant cell types are CD8 T cells, M0 macrophages, monocytes, and regulatory T cells.*


B.cells.memory
B.cells.naive
Dendritic.cells.activated
Dendritic.cells.resting
Eosinophils
Macrophages.M0
Macrophages.M1
Macrophages.M2
Mast.cells.activated
Mast.cells.resting
Monocytes
Neutrophils
NK.cells.activated
NK.cells.resting
Plasma.cells
T.cells.CD4.memory.activated
T.cells.CD4.memory.resting
T.cells.CD4.naive
T.cells.CD8
T.cells.follicular.helper
T.cells.gamma.delta
T.cells.regulatory..Tregs.
Boxplot of immune cell infiltrates (EwS samples, Schwabing, n=29)
Absolute proportion
1.00
0.75
0.50
0.25
0.00
Boxplot of immune cell infiltrates (EwS samples, Delattre, n=117)
Absolute proportion
1.00
0.75
0.50
0.25
0.00
Boxplot of immune cell infiltrates (EwS samples, Dirksen, n=85)
Absolute proportion
1.00
0.75
0.50
0.25
0.00
B.cells.memory
B.cells.naive
Dendritic.cells.activated
Dendritic.cells.resting
Eosinophils
Macrophages.M0
Macrophages.M1
Macrophages.M2
Mast.cells.activated
Mast.cells.resting
Monocytes
Neutrophils
NK.cells.activated
NK.cells.resting
Plasma.cells
T.cells.CD4.memory.activated
T.cells.CD4.memory.resting
T.cells.CD4.naive
T.cells.CD8
T.cells.follicular.helper
T.cells.gamma.delta
T.cells.regulatory..Tregs.



Suppl. Figure 6: Infiltration of 22 immune cell types in 6 EwS datasets (Schwabing, Delattre, Dirksen, Lawlor, Savola, Savola cell lines) deconvoluted by Cibersortx.*For each cell type, the estimated absolute proportion in the sample is given in percent of total cells. For better comparability, the plots y axes are limited to the same range (0.00% to 1.00%).*



Suppl. Figure 7: Distribution of effect sizes from linear regression modeling of SUVmax and gene expression. *(A) Distribution of slopes. (B) Distribution of Z-scaled slopes.*

Gene symbol (EntrezID)	adjusted p-value	r squared (Pearson correlation)	Slope from linear regression	Z-score of slope	p-value of Z-score	Doubling (pos.) or halving (neg.) of expression over x SUV units	Gene name	Favourable prognosis	Unfavourable prognsis	Short annotation
NPY1R (4886)	0,08	0,40	-0,334	-6,46	1,1E-10	-2,99	Neuropeptide Y Receptor Y1	breast cancer	/	NPY axis; GPCR
SLC38A4 (55089)	0,22	0,25	-0,235	-4,52	6,2E-06	-4,25	Solute Carrier Family 38 Member 4	liver cancer	/	amino acid transport
PLSCR4 (57088)	0,03	0,51	-0,223	-4,28	1,9E-05	-4,49	Phospholipid Scramblase 4	/	endometrial cancer, renal cancer	
NPY5R (4889)	0,33	0,17	-0,222	-4,26	2,1E-05	-4,51	Neuropeptide Y Receptor Y5	/	/	NPY axis; GPCR
MYL2 (4633)	0,05	0,46	0,205	4,12	3,8E-05	4,87	Myosin Light Chain 2	/	head and neck cancer	motor proteins / muscle
GRP (2922)	0,23	0,24	0,193	3,88	1,1E-04	5,18	Gastrin Releasing Peptide	/	/	muscle; proliferation; cancer
GPR174 (84636)	0,39	0,15	-0,193	-3,69	2,2E-04	-5,18	G Protein-Coupled Receptor 174	/	/	GPCR
PGAP1 (80055)	0,13	0,32	-0,193	-3,69	2,3E-04	-5,19	Post-GPI Attachment To Proteins Inositol Deacylase 1	/	/	
CLVS2 (134829)	0,13	0,32	-0,186	-3,55	3,8E-04	-5,38	Clavesin 2	/	/	
NEB (4703)	0,07	0,43	0,177	3,56	3,8E-04	5,66	Nebulin	/	/	muscle; membrane
FAXDC2 (10826)	0,03	0,59	-0,176	-3,37	7,6E-04	-5,67	Fatty Acid Hydroxylase Domain Containing 2	renal cancer, liver cancer	/	differentiation
DNASE1L3 (1776)	0,42	0,13	-0,174	-3,33	8,8E-04	-5,73	Deoxyribonuclease 1 Like 3	renal cancer, liver cancer	/	apoptosis; differentiation
FRZB (2487)	0,29	0,20	-0,172	-3,28	1,0E-03	-5,81	Frizzled Related Protein	/	renal cancer	GPCR; Wnt signaling
ZDHHC21 (340481)	0,13	0,33	-0,172	-3,28	1,0E-03	-5,82	Zinc Finger DHHC-Type Palmitoyltransferase 21	urothelial cancer, renal cancer	/	metabolism
SYT1 (6857)	0,30	0,20	-0,171	-3,26	1,1E-03	-5,84	Synaptotagmin 1	/	/	signaling
SGIP1 (84251)	0,22	0,25	-0,163	-3,10	1,9E-03	-6,14	SH3GL Interacting Endocytic Adaptor 1	/	/	signaling; energy homeostasis
TES (26136)	0,16	0,29	-0,158	-3,00	2,7E-03	-6,35	Testin LIM Domain Protein	renal cancer	pancreatic cancer	tumor suppressor
NETO2 (81831)	0,04	0,51	0,157	3,16	1,6E-03	6,38	Neuropilin And Tolloid Like 2	/	renal cancer	signaling
C5 (727)	0,07	0,43	-0,156	-2,97	3,0E-03	-6,41	Complement C5	liver cancer	/	inflammation
ABCA5 (23461)	0,31	0,19	-0,154	-2,92	3,5E-03	-6,51	ATP Binding Cassette Subfamily A Member 5	pancreatic cancer	/	molecule transport
SLC17A8 (246213)	0,20	0,08	-0,153	-2,91	3,6E-03	-6,52	Solute Carrier Family 17 Member 8	/	/	neurotransmitter transport; signaling
MYBL2 (4605	0,03	0,58	0,149	3,02	2,5E-03	6,69	MYB Proto-Oncogene Like 2	/	renal cancer, liver cancer, endometrial cancer	cell survival, proliferation, and differentiation
ELOVL2 (54898)	0,03	0,54	0,148	2,99	2,8E-03	6,76	ELOVL Fatty Acid Elongase 2	breast cancer	/	lipids and membrane

Suppl. Table 5: Results from linear regression analysis correlating gene expression and SUVmax.*23 genes with high effect size (abs(slope)>0.146) are annotated, sorted by decreasing effect size.*

GeneSymbol_EntrezID	Slope in regression with sample_19 (original analysis)	Slope in regression without sample_19	Change in effect size with sample_19 present compared to absent
NPY5R_4889	-0.222	-0.450	0.229
NPY1R_4886	-0.334	-0.486	0.152
GRP_2922	0.193	0.339	0.146
PLSCR4_57088	* -0.223	* -0.333	0.111
PGAP1_80055	-0.193	-0.292	0.099
TES_26136	-0.158	-0.239	0.081
DNASE1L3_1776	-0.174	-0.242	0.068
ABCA5_23461	-0.154	-0.218	0.064
ZDHHC21_340481	-0.172	-0.235	0.063
SLC38A4_55089	-0.235	-0.289	0.054
C5_727	-0.156	-0.209	0.053
FAXDC2_10826	* -0.176	-0.205	0.028
SGIP1_84251	-0.163	-0.187	0.025
MYBL2_4605	* 0.149	0.145	-0.004
NEB_4703	0.177	0.149	-0.027
GPR174_84636	-0.193	-0.154	-0.039
SLC17A8_246213	-0.153	-0.108	-0.045
FRZB_2487	-0.172	-0.121	-0.051
SYT1_6857	-0.171	-0.119	-0.052
NETO2_81831	* 0.157	0.104	-0.053
ELOVL2_54898	* 0.148	0.083	-0.065
MYL2_4633	0.205	0.139	-0.066
CLVS2_134829	-0.186	-0.099	-0.087

Suppl. Table 6: Linear regression results for 23 genes with high effect size.*For each gene, the slope from regression with and without sample_19 is given. Significant results are marked by * (adjusted p-value<0.05). High effects (abs(slope)>0.146) are highlighted in green. 15 of the 23 genes with high effect size still showed it after removing sample_19. The last column specifies increased (yellow) or decreased (red) effect sizes in the regression with sample_19 compared to the regression without.*

(A) Term "REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS"

Gene symbol	Rank in gene list	Rank metric score (= slope)	Running ES	Core enrichment
GRP	1	0.193	0.16	No
PENK	240	0.039	0.01	No
APLNR	306	0.032	-0.01	No
CXCL13	415	0.019	-0.07	No
CCL4	435	0.017	-0.07	No
F2R	466	0.013	-0.08	No
EDN1	477	0.013	-0.08	No
PPBP	540	0.007	-0.12	No
CXCL16	627	0.001	-0.18	No
CCL19	669	-0.002	-0.21	No
HEBP1	673	-0.002	-0.21	No
ACKR3	694	-0.004	-0.22	No
GRPR	774	-0.012	-0.27	No
C3	855	-0.018	-0.31	No
NPY	1352	-0.136	-0.57	Yes
C5	1361	-0.156	-0.45	Yes
NPY5R	1372	-0.222	-0.27	Yes
NPY1R	1375	-0.334	0.00	Yes

(B) Term "WP_GPCRS_CLASS_A_RHODOPSINLIKE"

Gene symbol	Rank in gene list	Rank metric score (= slope)	Running ES	Core enrichment
ADORA3	28	0.096	0.07	No
APLNR	306	0.032	-0.11	No
CHRM3	431	0.017	-0.18	No
F2R	466	0.013	-0.19	No
PTGER2	541	0.007	-0.24	No
ACKR3	694	-0.004	-0.35	No
OR10H1	770	-0.011	-0.39	No
GRPR	774	-0.012	-0.39	No
HTR1F	924	-0.024	-0.47	No
GPR20	984	-0.029	-0.49	No
LPAR4	1109	-0.043	-0.54	Yes
LPAR6	1112	-0.044	-0.50	Yes
GPR174	1371	-0.193	-0.51	Yes
NPY5R	1372	-0.222	-0.31	Yes
NPY1R	1375	-0.334	0.00	Yes

(C) Targets of transcription factor RNF2

Gene symbol	Rank in gene list	Rank metric score (= slope)	Running ES	Core enrichment
MYBL2	4	0.149	0.04	Yes
GINS1	14	0.118	0.06	Yes
ESPL1	27	0.096	0.08	Yes
MND1	31	0.093	0.10	Yes
RRM2	33	0.092	0.13	Yes
TROAP	52	0.080	0.14	Yes

E2F8	53	0.079	0.16	Yes
UBE2C	54	0.078	0.18	Yes
DLGAP5	55	0.078	0.20	Yes
NCAPH	63	0.074	0.22	Yes
FOXM1	66	0.073	0.24	Yes
CENPA	74	0.069	0.25	Yes
SPC25	81	0.068	0.26	Yes
SMC4	86	0.067	0.28	Yes
UBE3D	90	0.066	0.29	Yes
PLCXD1	95	0.064	0.31	Yes
RAD51AP1	99	0.064	0.32	Yes
MCM10	126	0.058	0.32	Yes
PTTG1	128	0.057	0.34	Yes
CDCA8	130	0.057	0.35	Yes
CIT	134	0.056	0.36	Yes
CDCA7	135	0.056	0.38	Yes
ZWINT	141	0.055	0.39	Yes
KIF20A	142	0.055	0.41	Yes
KIF14	146	0.055	0.42	Yes
KIF15	159	0.052	0.42	Yes
MCM5	178	0.047	0.42	Yes
TACC3	185	0.047	0.43	Yes
POLQ	186	0.046	0.44	Yes
CCNB2	187	0.046	0.46	Yes
CENPK	194	0.045	0.46	Yes
CENPO	197	0.045	0.47	Yes
GATA2	216	0.043	0.47	Yes
CHD7	220	0.043	0.48	Yes
NCAPG2	231	0.041	0.49	Yes
BLM	232	0.041	0.50	Yes
STIL	238	0.040	0.50	Yes
HEY2	249	0.039	0.51	Yes
FANCD2	269	0.036	0.50	Yes
NDC80	275	0.035	0.51	Yes
SHCBP1	278	0.035	0.52	Yes
ASPM	298	0.033	0.51	Yes
CEP55	303	0.032	0.52	Yes
KPNA2	309	0.031	0.52	Yes
POLE	318	0.030	0.52	Yes
C1orf112	356	0.026	0.50	No
FANCI	376	0.023	0.49	No
LSM5	384	0.022	0.49	No
PARPBP	396	0.021	0.49	No
RFC3	443	0.016	0.46	No
SMC6	458	0.014	0.45	No
RPL39L	462	0.014	0.45	No
WNK1	479	0.012	0.45	No

CDC25C	481	0.012	0.45	No
INTS7	531	0.008	0.41	No
MRPL13	533	0.008	0.41	No
ZW10	542	0.007	0.41	No
ARHGAP19	543	0.007	0.41	No
DLX1	558	0.006	0.40	No
NUDCD2	566	0.005	0.40	No
CDKN2C	589	0.004	0.38	No
SNRPD2	631	0.001	0.35	No
TOPBP1	640	0.000	0.35	No
DNMT3B	645	0.000	0.34	No
EPOR	674	-0.002	0.32	No
RGS5	720	-0.006	0.29	No
MASTL	736	-0.008	0.28	No
HOXD11	742	-0.008	0.28	No
TRMT13	762	-0.010	0.27	No
DZIP3	918	-0.023	0.15	No
HPGD	920	-0.023	0.16	No
RFWD3	947	-0.026	0.14	No
CEP192	951	-0.027	0.15	No
FBXO15	957	-0.027	0.15	No
DIAPH3	999	-0.031	0.13	No
ZNF43	1099	-0.041	0.07	No
SLFN11	1214	-0.057	-0.01	No
TMEFF2	1247	-0.063	-0.01	No
EPHA5	1331	-0.099	-0.05	No
TNKS	1345	-0.119	-0.03	No
CLVS2	1369	-0.186	0.00	No

(D) Targets of E2F family of transcription factors

Gene symbol	**Rank in gene list**	**Rank metric score (= slope)**	**Running ES**	**Core enrichment**
RRM2	33	0.092	0.08	Yes
E2F8	53	0.079	0.15	Yes
GMNN	87	0.066	0.20	Yes
NELL2	92	0.065	0.27	Yes
CDCA7	135	0.056	0.30	Yes
E2F7	143	0.055	0.36	Yes
ATF5	151	0.054	0.41	Yes
DNMT1	179	0.047	0.44	Yes
NRP2	222	0.042	0.46	Yes
POLE2	245	0.039	0.49	Yes
RASAL2	263	0.037	0.52	Yes
FANCD2	269	0.036	0.55	Yes
ID3	332	0.029	0.54	Yes
MCM2	350	0.027	0.55	Yes
DMD	360	0.025	0.58	Yes
STAG1	413	0.019	0.56	No

SMC6	458	0.014	0.54	No
INTS7	531	0.008	0.50	No
PCSK1	535	0.007	0.50	No
PRRC2C	587	0.004	0.47	No
TOPBP1	640	0.000	0.43	No
PODN	641	0.000	0.43	No
KPNB1	643	0.000	0.43	No
TRMT13	762	-0.010	0.36	No
SMC3	895	-0.021	0.28	No
FMO4	980	-0.029	0.25	No
ALDH6A1	1066	-0.037	0.23	No

(E) Targets of transcription factor TCF3

Gene symbol	Rank in gene list	Rank metric score (= slope)	Running ES	Core enrichment
NETO2	3	0.157	0.18	Yes
DPYSL5	8	0.135	0.34	Yes
FLNC	73	0.070	0.38	Yes
BCL11B	80	0.068	0.45	Yes
STC2	83	0.068	0.53	Yes
CD47	110	0.062	0.59	Yes
CDH2	148	0.054	0.62	Yes
MACF1	195	0.045	0.64	Yes
MTSS1	214	0.043	0.68	Yes
PODXL	297	0.033	0.66	No
HOXB7	468	0.013	0.55	No
PARP8	493	0.011	0.55	No
DGKD	576	0.004	0.49	No
KLF10	871	-0.019	0.30	No
DSCAM	891	-0.021	0.31	No
DAAM1	1075	-0.039	0.22	No

Suppl. Table 7: List of genes in GSEA core enrichments, which contribute most to the found enrichments.*(A) Term "REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS". (B) Term "WP_GPCRS_CLASS_A_RHODOPSINLIKE". (C) Targets of transcription factor RNF2. (D) Targets of E2F family of transcription factors. (E) Targets of transcription factor TCF3.*

6.2 List of Figures

6.3 List of Tables

6.4 List of Suppl. Figures

6.5 List of Suppl. Tables

6.6 List of Abbreviations

18F-FDG	18-F-fluorodeoxyglucose
ABCA5	ATP binding cassette subfamily A member 5
abs	Absolute value
avg	Average
BH	Benjamini- Hochberg
C2cp	Curated gene sets: canonical pathway (GSEA)
C3tft	Regulatory target gene sets: transcription factor targets (GSEA)
C5	Ontology gene sets (GSEA)
CDK	Cyclin dependent kinase
CDK2	Cyclin dependent kinase 2
CDK4	Cyclin dependent kinease 4
CDK6	Cyclin dependent kinease 6
CDKN2A	Cyclin dependent kinase inhibitor 2A
CI	Confidence interval
CLVS2	Clavesin 2
CT	Computed tomography
DNA	Deoxyribonucleic acid
DNASE1L3	Deoxyribonuclease 1 like 3
E2F3	E2F transcription factor 3
E2F4	E2F transcription factor 4
EANM	European Association of Nuclear Medicine
ELOVL2	ELOVL fatty acid elongase 2
ERG	ETS transcription factor ERG
ES	Enrichment score
ETS	ETS proto-oncogene 1, transcription factor
ETV1	ETS variant transcription factor 1
ETV4	ETS variant transcription factor 4
EwS	Ewing sarcoma
EWS	EWS RNA binding protein 1 (EWSR1)
f	Female
FAXDC2	Fatty acid hydroxylase domain containing 2
FDA	Food and Drug Administration
FDR	False discovery rate
FEV	FEV transcription factor, ETS family member
FLI1	Fli-1 proto-oncogene, ETS transcription factor
FRZB	Frizzled related protein
glcm	Gray level co-occurrence matrix
gldm	Gray level dependence matrix
glrlm	Gray level run length matrix
glszm	Gray level size zone matrix
GO	Gene Ontology

GPCR	G protein-coupled receptor
GPR174	G protein-coupled receptor 174
GRP	Gastrin releasing peptide
GSEA	Gene set enrichment analysis
H	Hallmark gene sets (GSEA)
HIF1A	Hypoxia inducible factor 1 subunit alpha
HLA	Human leukocyte antigen
HPO	Human Phenotype Ontology
HR	Hazard ratio
IBSI	Image biomarker standardization initiative
ICC	Intraclass correlation coefficient
IF	Image feature
limma	Linear models for microarray data
LoG	Laplacian of Gaussian filter
logFC	Log2 fold change
LPAR4	Lysophosphatidic acid receptor 4
LPAR6	Lysophosphatidic acid receptor 6
m	Male
MCC	Matthew's correlation coefficient
MCM2	Minichromosome maintenance complex component 2
MKI67	Marker of proliferation Ki-67
MR	Magnetic resonance
MYBL2	MYB proto-oncogene like 2
MYL2	Myosin light chain 2
n	Number
na	Not available / not applicable
NEB	Nebulin
NES	Normalized enrichment score
NETO2	Neuropilin and tolloid like 2
ngtdm	Neighboring gray tone difference matrix features
NPY	Neuropeptide Y
NPY1R	Neuropeptide Y receptor Y1
NPY2R	Neuropeptide Y receptor Y2
NPY5R	Neuropeptide Y receptor Y5
NPY6R	Neuropeptide Y receptor Y6 (pseudogene)
OS	Overall survival
p	p-value
PcG	Polycomb group
PCNA	Proliferating cell nuclear antigen
PET	Positron emission tomography
PGAP1	Post-GPI attachment to proteins inositol deacylase 1
PLSCR4	Phospholipid scramblase 4
PPY	Pancreatic polypeptide
PR AUC	Area under the precision-recall curve

PRC1	Polycomb repressive complex 1
PYY	Peptide YY
q	q-value
r^2	Squared Pearson correlation coefficient
RMA	Robust Multichip Average
RNA	Ribonucleic acid
RNF2	Ring finger protein 2
ROC AUC	Area under the receiver operating characteristic curve
SGIP1	SH3GL interacting endocytic adaptor 1
SLC17A8	Solute carrier family 17 member 8
SLC38A4	Solute carrier family 38 member 4
SMGs	Significantly mutated genes
SOX2	SRY-box transcription factor 2
STAG2	Stromal antigen 2
SUV	Standardized uptake value
SUVmax	Maximal standardized uptake value
SUVmean	Mean standardized uptake value
SYT1	Synaptotagmin 1
TCF3	Transcription factor 3
TES	Testin LIM domain protein
TF	Transcription factor
TOP2A	DNA topoisomerase II alpha
TP53	Tumor protein p53
ZDHHC21	Zinc finger DHHC-type palmitoyltransferase 21

6.7 Acknowledgements

First of all, I would like to thank my supervisor Prof. Dr. med. Stefan Burdach for providing me the opportunity to work on this highly interesting topic about radiogenomics. He introduced me to the world of pediatric oncology and clinical research in general and connected me to many colleagues in this field. I appreciate his support, motivating discussions and valuable advice, and his deep expertise in pediatric oncology.

I'm also grateful to my supervisor Prof. Dr. Hans-Werner Mewes and my mentor Prof. Dr. Günther Richter for their critical view on the project from a different perspective, and for their always helpful input and constructive feedback. Thank you for pointing out gaps that needed to be addressed and for helping me not to lose focus. Günther Richter always provided kind advice and had an open ear for problems of any kind.

Many thanks to Dr. Janina Erlewein-Schweizer for the lively professional exchange, her critical eye and the joint efforts to solve problems.

I am also indebted to all physicians for their support from the clinical side, especially Dr. med. Marie Sophie Knape for organizing and establishing the procedures of the project with me. Furthermore, I would like to thank all colleagues involved in the Kinderklinik Schwabing, in the Klinikum rechts der Isar and throughout Germany in Essen and Münster for their contribution and helpful support. Special thanks go to the team of the tissue bank for the reliable cooperation and to the team of the nuclear medicine department – especially Prof. Dr. med. Wolfgang Weber, Dr. med. Wiebke Seeman, and Dr. med. Sebastian Kleiner – for their fruitful professional input and help with image processing. They took a lot of time to work with me on the PET images.

Sincere thanks are given to all former and current colleagues in the lab for the pleasant, inspiring and motivating working atmosphere and the personal as well as professional exchange. I appreciate their helpful discussions, comments and advice.

Furthermore, I am indebted to the patients for their approval to provide their data for research. These data form the basis for such studies.

Last not least, I am deeply grateful to my family and friends for their constant support, understanding and patience; for building me up and motivating me to keep going even in difficult moments.

This project would not have been possible without those mentioned here. Once again, thank you very much.

7 References

1. Grünewald, T.G.P., et al., *Ewing sarcoma.* Nature Reviews Disease Primers, 2018. **4**(1): p. 5.
2. Tirode, F., et al., *Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations.* Cancer Discov, 2014. **4**(11): p. 1342-53.
3. Brohl, A.S., et al., *The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation.* PLoS Genet, 2014. **10**(7): p. e1004475.
4. Guerra, R.B., et al., *Comparative analysis between osteosarcoma and Ewing's sarcoma: evaluation of the time from onset of signs and symptoms until diagnosis.* Clinics (Sao Paulo), 2006. **61**(2): p. 99-106.
5. Schmidkonz, C., et al., *Assessment of treatment responses in children and adolescents with Ewing sarcoma with metabolic tumor parameters derived from (18)F-FDG-PET/CT and circulating tumor DNA.* Eur J Nucl Med Mol Imaging, 2020. **47**(6): p. 1564-1575.
6. Salguero-Aranda, C., et al., *Breakthrough Technologies Reshape the Ewing Sarcoma Molecular Landscape.* Cells, 2020. **9**(4).
7. Burdach, S. and H. Jürgens, *High-dose chemoradiotherapy (HDC) in the Ewing family of tumors (EFT).* Crit Rev Oncol Hematol, 2002. **41**(2): p. 169-89.
8. Burdach, S., et al., *Myeloablative radiochemotherapy and hematopoietic stem-cell rescue in poor-prognosis Ewing's sarcoma.* J Clin Oncol, 1993. **11**(8): p. 1482-8.
9. Burdach, S., et al., *Total body MRI-governed involved compartment irradiation combined with high-dose chemotherapy and stem cell rescue improves long-term survival in Ewing tumor patients with multiple primary bone metastases.* Bone Marrow Transplant, 2010. **45**(3): p. 483-9.
10. Burdach, S.E.G., et al., *Klinik und multidisziplinäre Therapie der Ewing-Sarkome, MANUAL Knochentumoren und Weichteilsarkome.* Zuckschwerdt Verlag 6. Auflage 2017.
11. Grobner, S.N., et al., *The landscape of genomic alterations across childhood cancers.* Nature, 2018. **555**(7696): p. 321-327.
12. Burdach, S.E.G., et al., *Precision medicine in pediatric oncology.* Mol Cell Pediatr, 2018. **5**(1): p. 6.
13. Delattre, O., et al., *Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours.* Nature, 1992. **359**(6391): p. 162-5.
14. Theisen, E.R., et al., *Transcriptomic analysis functionally maps the intrinsically disordered domain of EWS/FLI and reveals novel transcriptional dependencies for oncogenesis.* Genes Cancer, 2019. **10**(1-2): p. 21-38.
15. Agelopoulos, K., et al., *Deep Sequencing in Conjunction with Expression and Functional Analyses Reveals Activation of FGFR1 in Ewing Sarcoma.* Clin Cancer Res, 2015. **21**(21): p. 4935-46.
16. Miller, H.E., et al., *Reconstruction of Ewing Sarcoma Developmental Context from Mass-Scale Transcriptomics Reveals Characteristics of EWSR1-FLI1 Permissibility.* Cancers (Basel), 2020. **12**(4).
17. Sand, L.G., K. Szuhai, and P.C. Hogendoorn, *Sequencing Overview of Ewing Sarcoma: A Journey across Genomic, Epigenomic and Transcriptomic Landscapes.* Int J Mol Sci, 2015. **16**(7): p. 16176-215.
18. Stahl, D., et al., *Prognostic profiling of the immune cell microenvironment in Ewing´s Sarcoma Family of Tumors.* Oncoimmunology, 2019. **8**(12): p. e1674113.
19. Richter, G.H., et al., *EZH2 is a mediator of EWS/FLI1 driven tumor growth and metastasis blocking endothelial and neuro-ectodermal differentiation.* Proc Natl Acad Sci U S A, 2009. **106**(13): p. 5324-9.
20. Ewing, J., *Classics in oncology. Diffuse endothelioma of bone. James Ewing. Proceedings of the New York Pathological Society, 1921.* CA Cancer J Clin, 1972. **22**(2): p. 95-8.

21. Staege, M.S., et al., *DNA microarrays reveal relationship of Ewing family tumors to both endothelial and fetal neural crest-derived cells and define novel targets.* Cancer Res, 2004. **64**(22): p. 8213-21.
22. Burdach, S., et al., *Epigenetic maintenance of stemness and malignancy in peripheral neuroectodermal tumors by EZH2.* Cell Cycle, 2009. **8**(13): p. 1991-6.
23. Hong, B., et al., *Single-cell transcriptional profiling reveals heterogeneity and developmental trajectories of Ewing sarcoma.* J Cancer Res Clin Oncol, 2022.
24. Schmidt, D., D. Harms, and S. Burdach, *Malignant peripheral neuroectodermal tumours of childhood and adolescence.* Virchows Arch A Pathol Anat Histopathol, 1985. **406**(3): p. 351-65.
25. Crompton, B.D., et al., *The genomic landscape of pediatric Ewing sarcoma.* Cancer Discov, 2014. **4**(11): p. 1326-41.
26. Sheffield, N.C., et al., *DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma.* Nat Med, 2017. **23**(3): p. 386-395.
27. Reuter, D., et al., *Immunostimulation by OX40 Ligand Transgenic Ewing Sarcoma Cells.* Front Oncol, 2015. **5**: p. 242.
28. Bandopadhayay, P. and M. Meyerson, *Landscapes of childhood tumours.* Nature, 2018. **555**(7696): p. 316-317.
29. Ma, X., et al., *Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours.* Nature, 2018. **555**(7696): p. 371-376.
30. Burdach, S., *Molecular precision chemotherapy: overcoming resistance to targeted therapies?* Clin Cancer Res, 2014. **20**(5): p. 1064-6.
31. Sala, E., et al., *Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging.* Clin Radiol, 2017. **72**(1): p. 3-10.
32. Ganeshan, B., et al., *Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival.* Eur Radiol, 2012. **22**(4): p. 796-802.
33. Kinahan, P.E. and J.W. Fletcher, *Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy.* Seminars in ultrasound, CT, and MR, 2010. **31**(6): p. 496-505.
34. Kumar, V., et al., *Radiomics: the process and the challenges.* Magn Reson Imaging, 2012. **30**(9): p. 1234-48.
35. Gillies, R.J., P.E. Kinahan, and H. Hricak, *Radiomics: Images Are More than Pictures, They Are Data.* Radiology, 2016. **278**(2): p. 563-77.
36. Guiot, J., et al., *A review in radiomics: Making personalized medicine a reality via routine imaging.* Med Res Rev, 2021.
37. Incoronato, M., et al., *Radiogenomic Analysis of Oncological Data: A Technical Survey.* Int J Mol Sci, 2017. **18**(4).
38. Shankar, L.K., et al., *Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials.* J Nucl Med, 2006. **47**(6): p. 1059-66.
39. Leijenaar, R.T., et al., *Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability.* Acta Oncol, 2013. **52**(7): p. 1391-7.
40. Singh, I., et al., *(18)F-FDG-PET SUV as a Prognostic Marker of Increasing Size in Thyroid Cancer Tumors.* Endocr Pract, 2017. **23**(2): p. 182-189.
41. Floberg, J.M., et al., *Pre-Treatment [F-18]FDG-PET SUVmax as a Prognostic and Radiogenomic Marker in Cervical Cancer.* International Journal of Radiation Oncology • Biology • Physics, 2018. **102**(3): p. S82-S83.
42. Diao, W., F. Tian, and Z. Jia, *The prognostic value of SUVmax measuring on primary lesion and ALN by (18)F-FDG PET or PET/CT in patients with breast cancer.* Eur J Radiol, 2018. **105**: p. 1-7.
43. Zhang, Q., et al., *Prognostic Value of MTV, SUVmax and the T/N Ratio of PET/CT in Patients with Glioma: A Systematic Review and Meta-Analysis.* J Cancer, 2019. **10**(7): p. 1707-1716.
44. Perrone, A.M., et al., *Potential Prognostic Role of (18)F-FDG PET/CT in Invasive Epithelial Ovarian Cancer Relapse. A Preliminary Study.* Cancers (Basel), 2019. **11**(5).

45. Kim, C.Y., et al., *Quantitative metabolic parameters measured on F-18 FDG PET/CT predict survival after relapse in patients with relapsed epithelial ovarian cancer.* Gynecol Oncol, 2015. **136**(3): p. 498-504.
46. Torizuka, T., et al., *Prognostic value of 18F-FDG PET in patients with head and neck squamous cell cancer.* AJR Am J Roentgenol, 2009. **192**(4): p. W156-60.
47. Pankowska, V., et al., *FDG PET/CT as a survival prognostic factor in patients with advanced renal cell carcinoma.* Clin Exp Med, 2019. **19**(1): p. 143-148.
48. Berghmans, T., et al., *Primary tumor standardized uptake value (SUVmax) measured on fluorodeoxyglucose positron emission tomography (FDG-PET) is of prognostic value for survival in non-small cell lung cancer (NSCLC): a systematic review and meta-analysis (MA) by the European Lung Cancer Working Party for the IASLC Lung Cancer Staging Project.* J Thorac Oncol, 2008. **3**(1): p. 6-12.
49. Cerfolio, R.J. and A.S. Bryant, *Maximum standardized uptake values on positron emission tomography of esophageal cancer predicts stage, tumor biology, and survival.* Ann Thorac Surg, 2006. **82**(2): p. 391-4; discussion 394-5.
50. Pan, L., et al., *Prognostic significance of SUV on PET/CT in patients with esophageal cancer: a systematic review and meta-analysis.* Eur J Gastroenterol Hepatol, 2009. **21**(9): p. 1008-15.
51. Lee, W., et al., *Metabolic activity by FDG-PET/CT after neoadjuvant chemotherapy in borderline resectable and locally advanced pancreatic cancer and association with survival.* Br J Surg, 2021. **109**(1): p. 61-70.
52. Hellwig, D., et al., *Diagnostic performance and prognostic impact of FDG-PET in suspected recurrence of surgically treated non-small cell lung cancer.* Eur J Nucl Med Mol Imaging, 2006. **33**(1): p. 13-21.
53. Jiang, Y., et al., *The maximum standardized uptake value and extent of peritoneal involvement may predict the prognosis of patients with recurrent ovarian cancer after primary treatment: A retrospective clinical study.* Medicine (Baltimore), 2020. **99**(8): p. e19228.
54. Sala, E., et al., *Recurrent ovarian cancer: use of contrast-enhanced CT and PET/CT to accurately localize tumor recurrence and to predict patients' survival.* Radiology, 2010. **257**(1): p. 125-34.
55. Hwang, J.P., et al., *Prognostic Value of SUVmax Measured by Pretreatment Fluorine-18 Fluorodeoxyglucose Positron Emission Tomography/Computed Tomography in Patients with Ewing Sarcoma.* PLoS One, 2016. **11**(4): p. e0153281.
56. Jamet, B., et al., *Initial FDG-PET/CT predicts survival in adults Ewing sarcoma family of tumors.* Oncotarget, 2017. **8**(44): p. 77050-77060.
57. Raciborska, A., et al., *Response to chemotherapy estimates by FDG PET is an important prognostic factor in patients with Ewing sarcoma.* Clin Transl Oncol, 2016. **18**(2): p. 189-95.
58. Hawkins, D.S., et al., *[18F]Fluorodeoxyglucose positron emission tomography predicts outcome for Ewing sarcoma family of tumors.* J Clin Oncol, 2005. **23**(34): p. 8828-34.
59. Salem, U., et al., *(18)F-FDG PET/CT as an Indicator of Survival in Ewing Sarcoma of Bone.* J Cancer, 2017. **8**(15): p. 2892-2898.
60. Annovazzi, A., et al., *[(18)F]FDG PET/CT quantitative parameters for the prediction of histological response to induction chemotherapy and clinical outcome in patients with localised bone and soft-tissue Ewing sarcoma.* Eur Radiol, 2021. **31**(9): p. 7012-7021.
61. Macpherson, R.E., et al., *Retrospective audit of 957 consecutive (18)F-FDG PET-CT scans compared to CT and MRI in 493 patients with different histological subtypes of bone and soft tissue sarcoma.* Clin Sarcoma Res, 2018. **8**: p. 9.
62. Gevaert, O., et al., *Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data--methods and preliminary results.* Radiology, 2012. **264**(2): p. 387-96.
63. Hatt, M., et al., *Characterization of PET/CT images using texture analysis: the past, the present... any future?* Eur J Nucl Med Mol Imaging, 2017. **44**(1): p. 151-165.

64. Nair, V.S., et al., *Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer.* Cancer Res, 2012. **72**(15): p. 3725-34.
65. Osborne, J.R., et al., *18F-FDG PET of locally invasive breast cancer and association of estrogen receptor status with standardized uptake value: microarray and immunohistochemical analysis.* J Nucl Med, 2010. **51**(4): p. 543-50.
66. Heiden, B.T., et al., *18F-FDG PET intensity correlates with a hypoxic gene signature and other oncogenic abnormalities in operable non-small cell lung cancer.* PLoS One, 2018. **13**(7): p. e0199970.
67. Lee, J.D., et al., *Analysis of gene expression profiles of hepatocellular carcinomas with regard to 18F-fluorodeoxyglucose uptake pattern on positron emission tomography.* Eur J Nucl Med Mol Imaging, 2004. **31**(12): p. 1621-30.
68. Tixier, F., et al., *Transcriptomics in cancer revealed by Positron Emission Tomography radiomics.* Sci Rep, 2020. **10**(1): p. 5660.
69. Tixier, F., et al., *Signaling pathways alteration involved in head and neck cancer can be identified through textural features analysis in 18F-FDG PET images: a prospective study.* Journal of Nuclear Medicine, 2015. **56**(supplement 3): p. 449-449.
70. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
71. Hanahan, D., *Hallmarks of Cancer: New Dimensions.* Cancer Discovery, 2022. **12**(1): p. 31-46.
72. R Core Team, *R: A Language and Environment for Statistical Computing*. 2020, R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
73. Volchenboum, S.L., et al., *Gene Expression Profiling of Ewing Sarcoma Tumors Reveals the Prognostic Importance of Tumor-Stromal Interactions: A Report from the Children's Oncology Group.* J Pathol Clin Res, 2015. **1**(2): p. 83-94.
74. Mitchell, M.W., *Bias of the Random Forest Out-of-Bag (OOB) Error for Certain Input Parameters.* Open Journal of Statistics, 2011. **2011**: p. 205-211.
75. Dittman, D.J., T.M. Khoshgoftaar, and A. Napolitano. *The Effect of Data Sampling When Using Random Forest on Imbalanced Bioinformatics Data*. in *2015 IEEE International Conference on Information Reuse and Integration*. 2015.
76. Deist, T.M., et al., *Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers.* Med Phys, 2018. **45**(7): p. 3449-3459.
77. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.* Bioinformatics, 2003. **19**(2): p. 185-93.
78. Irizarry, R.A., et al., *Summaries of Affymetrix GeneChip probe level data.* Nucleic Acids Res, 2003. **31**(4): p. e15.
79. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics, 2003. **4**(2): p. 249-64.
80. Carvalho, B.S. and R.A. Irizarry, *A framework for oligonucleotide microarray preprocessing.* Bioinformatics, 2010. **26**(19): p. 2363-7.
81. Dai, M., et al., *Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data.* Nucleic Acids Res, 2005. **33**(20): p. e175.
82. Lu, X. and X. Zhang, *The effect of GeneChip gene definitions on the microarray study of cancers.* Bioessays, 2006. **28**(7): p. 739-46.
83. Sandberg, R. and O. Larsson, *Improved precision and accuracy for microarrays using updated probe set definitions.* BMC Bioinformatics, 2007. **8**: p. 48.
84. Galili, T., *dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering.* Bioinformatics, 2015. **31**(22): p. 3718-20.
85. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic Acids Res, 2002. **30**(1): p. 207-10.

86. Savola, S., et al., *High Expression of Complement Component 5 (C5) at Tumor Site Associates with Superior Survival in Ewing's Sarcoma Family of Tumour Patients.* ISRN Oncol, 2011. **2011**: p. 168712.
87. Hackstadt, A.J. and A.M. Hess, *Filtering for increased power for microarray data analysis.* BMC Bioinformatics, 2009. **10**: p. 11.
88. Kuhn, M., *caret: Classification and Regression Training*. 2021, R package version 6.0-90. https://CRAN.R-project.org/package=caret.
89. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
90. Sherman, B.T., et al., *DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update).* Nucleic Acids Res, 2022.
91. Rosset, A., L. Spadola, and O. Ratib, *OsiriX: an open-source software for navigating in multidimensional DICOM images.* J Digit Imaging, 2004. **17**(3): p. 205-16.
92. El Naqa, I., et al., *Exploring feature-based approaches in PET images for predicting cancer treatment outcomes.* Pattern Recognit, 2009. **42**(6): p. 1162-1171.
93. Orlhac, F., et al., *Tumor Texture Analysis in 18F-FDG PET: Relationships Between Texture Parameters, Histogram Indices, Standardized Uptake Values, Metabolic Volumes, and Total Lesion Glycolysis.* Journal of Nuclear Medicine, 2014. **55**(3): p. 414-422.
94. Galavis, P.E., et al., *Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters.* Acta Oncol, 2010. **49**(7): p. 1012-6.
95. Zwanenburg, A., et al., *The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping.* Radiology, 2020. **295**(2): p. 328-338.
96. van Griethuysen, J.J.M., et al., *Computational Radiomics System to Decode the Radiographic Phenotype.* Cancer Res, 2017. **77**(21): p. e104-e107.
97. Leijenaar, R.T., et al., *The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis.* Sci Rep, 2015. **5**: p. 11075.
98. Gamer, M., J. Lemon, and I.F.P. Singh, *irr: Various Coefficients of Interrater Reliability and Agreement*. 2019, R package version 0.84.1. https://CRAN.R-project.org/package=irr.
99. Koo, T.K. and M.Y. Li, *A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.* J Chiropr Med, 2016. **15**(2): p. 155-63.
100. Cicchetti, D.V., *Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology*. 1994, American Psychological Association: US. p. 284-290.
101. Bardou, P., et al., *jvenn: an interactive Venn diagram viewer.* BMC Bioinformatics, 2014. **15**(1): p. 293.
102. Revelle, W., *psych: Procedures for Psychological, Psychometric, and Personality Research*. 2021, R package version 2.1.9. https://CRAN.R-project.org/package=psych.
103. Kassambara, A., *ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2'*. 2019, R package version 0.1.3. https://CRAN.R-project.org/package=ggcorrplot.
104. Segal, E., et al., *Decoding global gene expression programs in liver cancer by noninvasive imaging.* Nat Biotechnol, 2007. **25**(6): p. 675-80.
105. Therneau, T.M., *A Package for Survival Analysis in R*. 2021, R package version 3.2-13, URL https://CRAN.R-project.org/package=survival.
106. Therneau, T.M. and P.M. Grambsch, *Modeling Survival Data: Extending the Cox Model*. 2000, Springer, New York. ISBN 0-387-98784-3.
107. Kassambara, A., M. Kosinski, and P. Biecek, *survminer: Drawing Survival Curves using 'ggplot2'*. 2021, R package version 0.4.9. https://CRAN.R-project.org/package=survminer.
108. John, C.R., *MLeval: Machine Learning Model Evaluation*. 2020, R package version 0.3. https://CRAN.R-project.org/package=MLeval.
109. Newman, A.M., et al., *Determining cell type abundance and expression from bulk tissues with digital cytometry.* Nat Biotechnol, 2019. **37**(7): p. 773-782.

110. Postel-Vinay, S., et al., *Common variants near TARDBP and EGR2 are associated with susceptibility to Ewing sarcoma.* Nat Genet, 2012. **44**(3): p. 323-7.
111. Svoboda, L.K., et al., *Overexpression of HOX genes is prevalent in Ewing sarcoma and is associated with altered epigenetic regulation of developmental transcription programs.* Epigenetics, 2014. **9**(12): p. 1613-25.
112. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
113. Phipson, B., et al., *Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression.* Ann Appl Stat, 2016. **10**(2): p. 946-963.
114. Jeanmougin, M., et al., *Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies.* PLoS One, 2010. **5**(9): p. e12336.
115. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the Royal Statistical Society Series B, 1995. **57**: p. 289-300.
116. *heatmap.3*. obtained from https://raw.githubusercontent.com/obigriffith/biostar-tutorials/master/Heatmaps/heatmap.3.R.
117. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.* BMC Bioinformatics, 2013. **14**: p. 128.
118. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W90-7.
119. Xie, Z., et al., *Gene Set Knowledge Discovery with Enrichr.* Curr Protoc, 2021. **1**(3): p. e90.
120. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
121. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. **34**(3): p. 267-73.
122. Colen, R., et al., *NCI Workshop Report: Clinical and Computational Requirements for Correlating Imaging Phenotypes with Genomics Signatures.* Transl Oncol, 2014. **7**(5): p. 556-69.
123. Bai, H.X., et al., *Imaging genomics in cancer research: limitations and promises.* Br J Radiol, 2016. **89**(1061): p. 20151030.
124. Duan, X.Y., et al., *Predictive significance of standardized uptake value parameters of FDG-PET in patients with non-small cell lung carcinoma.* Braz J Med Biol Res, 2015. **48**(3): p. 267-72.
125. Sunnetcioglu, A., et al., *Associations between the standardized uptake value of (18)F-FDG PET/CT and demographic, clinical, pathological, radiological factors in lung cancer.* Int J Clin Exp Med, 2015. **8**(9): p. 15794-800.
126. Qiu, X., et al., *Prognostic impact of maximum standardized uptake value on (18) F-FDG PET/CT imaging of the primary lung lesion on survival in advanced non-small cell lung cancer: A retrospective study.* Thorac Cancer, 2021. **12**(6): p. 845-853.
127. AbdElaal, A.A., et al., *Correlation of primary tumor metabolic parameters with clinical, histopathological and molecular characteristics in breast cancer patients at pre-operative staging FDG-PET/CT study.* Egyptian Journal of Radiology and Nuclear Medicine, 2021. **52**(1): p. 171.
128. Sengoz, T., et al., *Relationships of (18)F-FDG uptake by primary tumors with prognostic factors and molecular subtype in ductal breast cancer.* Rev Esp Med Nucl Imagen Mol (Engl Ed), 2021.
129. Soffer, S., et al., *PET-CT imaging features that differentiate between positive and negative EWSR1 translocation in Ewing sarcoma.* Nucl Med Commun, 2019.
130. Karski, E.E., et al., *Identification of Discrete Prognostic Groups in Ewing Sarcoma.* Pediatr Blood Cancer, 2016. **63**(1): p. 47-53.
131. Stahl, M., et al., *Risk of recurrence and survival after relapse in patients with Ewing sarcoma.* Pediatr Blood Cancer, 2011. **57**(4): p. 549-53.

132. Kuo, M.D. and N. Jamshidi, *Behind the numbers: Decoding molecular phenotypes with radiogenomics--guiding principles and technical considerations.* Radiology, 2014. **270**(2): p. 320-5.
133. Berghuis, D., et al., *Pro-inflammatory chemokine-chemokine receptor interactions within the Ewing sarcoma microenvironment determine CD8(+) T-lymphocyte infiltration and affect tumour progression.* J Pathol, 2011. **223**(3): p. 347-57.
134. Petrova, V., et al., *The hypoxic tumour microenvironment.* Oncogenesis, 2018. **7**(1): p. 10.
135. Skubitz, K.M., et al., *Effect of chemotherapy on cancer stem cells and tumor-associated macrophages in a prospective study of preoperative chemotherapy in soft tissue sarcoma.* J Transl Med, 2019. **17**(1): p. 130.
136. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles.* Nat Methods, 2015. **12**(5): p. 453-7.
137. Kumar, S., et al., *Tumor-infiltrating CD8(+) T cell antitumor efficacy and exhaustion: molecular insights.* Drug Discov Today, 2021. **26**(4): p. 951-967.
138. van Oost, S., et al., *Linking Immunity with Genomics in Sarcomas: Is Genomic Complexity an Immunogenic Trigger?* Biomedicines, 2021. **9**(8).
139. Hack, R.I., et al., *When SUV Matters: FDG PET/CT at Baseline Correlates with Survival in Soft Tissue and Ewing Sarcoma.* Life (Basel), 2021. **11**(9).
140. *Human Protein Atlas available from www.proteinatlas.org*.
141. Warburg, O., *On respiratory impairment in cancer cells.* Science, 1956. **124**(3215): p. 269-70.
142. Warburg, O., *On the origin of cancer cells.* Science, 1956. **123**(3191): p. 309-14.
143. Zumkeller, W. and S. Burdach, *The insulin-like growth factor system in normal and malignant hematopoietic cells.* Blood, 1999. **94**(11): p. 3653-7.
144. Sipol, A., et al., *MondoA drives malignancy in B-ALL through enhanced adaptation to metabolic stress.* Blood, 2022. **139**(8): p. 1184-1197.
145. Velasquez, L.M., et al., *Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies.* J Nucl Med, 2009. **50**(10): p. 1646-54.
146. Schafer, J.F., et al., *Simultaneous whole-body PET/MR imaging in comparison to PET/CT in pediatric oncology: initial results.* Radiology, 2014. **273**(1): p. 220-31.
147. Sachpekidis, C., et al., *Comparison of (18)F-FDG PET/CT and PET/MRI in patients with multiple myeloma.* Am J Nucl Med Mol Imaging, 2015. **5**(5): p. 469-78.
148. Drzezga, A., et al., *First clinical experience with integrated whole-body PET/MR: comparison to PET/CT in patients with oncologic diagnoses.* J Nucl Med, 2012. **53**(6): p. 845-55.
149. *GeneCards: The Human Gene Database available from www.genecards.org*. Weizmann Institute of Science.
150. Fagone, P., et al., *Identification of novel chemotherapeutic strategies for metastatic uveal melanoma.* Sci Rep, 2017. **7**: p. 44564.
151. Jin, Q., et al., *Novel function of FAXDC2 in megakaryopoiesis.* Blood Cancer J, 2016. **6**(9): p. e478.
152. Fedorova, M.S., et al., *NETO2 Is Deregulated in Breast, Prostate, and Colorectal Cancer and Participates in Cellular Signaling.* Front Genet, 2020. **11**: p. 594933.
153. Wang, X., et al., *Neuropilin and tolloid-like 2 regulates the progression of osteosarcoma.* Gene, 2021. **768**: p. 145292.
154. Jeong, D., et al., *ELOVL2: a novel tumor suppressor attenuating tamoxifen resistance in breast cancer.* Am J Cancer Res, 2021. **11**(6): p. 2568-2589.
155. Tanaka, K., et al., *ELOVL2 promotes cancer progression by inhibiting cell apoptosis in renal cell carcinoma.* Oncol Rep, 2022. **47**(2): p. 23.
156. Gimple, R.C., et al., *Glioma Stem Cell-Specific Superenhancer Promotes Polyunsaturated Fatty-Acid Synthesis to Support EGFR Signaling.* Cancer Discov, 2019. **9**(9): p. 1248-1267.
157. Sannino, G., et al., *Gene expression and immunohistochemical analyses identify SOX2 as major risk factor for overall survival and relapse in Ewing sarcoma patients.* EBioMedicine, 2019. **47**: p. 156-162.

158. Musa, J., et al., *MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis.* Cell Death Dis, 2017. **8**(6): p. e2895.
159. Xiong, Y.C., et al., *Overexpression of MYBL2 promotes proliferation and migration of non-small-cell lung cancer via upregulating NCAPH.* Mol Cell Biochem, 2020. **468**(1-2): p. 185-193.
160. Li, Q., et al., *MYBL2 disrupts the Hippo-YAP pathway and confers castration resistance and metastatic potential in prostate cancer.* Theranostics, 2021. **11**(12): p. 5794-5812.
161. Bayley, R., C. Ward, and P. Garcia, *MYBL2 amplification in breast cancer: Molecular mechanisms and therapeutic potential.* Biochim Biophys Acta Rev Cancer, 2020. **1874**(2): p. 188407.
162. Qi, G., et al., *CDCA8, targeted by MYBL2, promotes malignant progression and olaparib insensitivity in ovarian cancer.* Am J Cancer Res, 2021. **11**(2): p. 389-415.
163. Musa, J., et al., *Cooperation of cancer drivers with regulatory germline variants shapes clinical outcomes.* Nat Commun, 2019. **10**(1): p. 4128.
164. Jeppsson, S., S. Srinivasan, and B. Chandrasekharan, *Neuropeptide Y (NPY) promotes inflammation-induced tumorigenesis by enhancing epithelial cell proliferation.* Am J Physiol Gastrointest Liver Physiol, 2017. **312**(2): p. G103-g111.
165. Groneberg, D.A., et al., *Neuropeptide Y (NPY).* Pulm Pharmacol Ther, 2004. **17**(4): p. 173-80.
166. Wheway, J., H. Herzog, and F. Mackay, *NPY and receptors in immune and inflammatory diseases.* Curr Top Med Chem, 2007. **7**(17): p. 1743-52.
167. Chandrasekharan, B., B.G. Nezami, and S. Srinivasan, *Emerging neuropeptide targets in inflammation: NPY and VIP.* Am J Physiol Gastrointest Liver Physiol, 2013. **304**(11): p. G949-57.
168. Zhao, X., J. Liu, and I. Ahmad, *Differentiation of embryonic stem cells into retinal neurons.* Biochem Biophys Res Commun, 2002. **297**(2): p. 177-84.
169. Li, J., Y. Tian, and A. Wu, *Neuropeptide Y receptors: a promising target for cancer imaging and therapy.* Regen Biomater, 2015. **2**(3): p. 215-9.
170. Sun, F., et al., *Identification of hub genes in gastric cancer by integrated bioinformatics analysis.* Transl Cancer Res, 2021. **10**(6): p. 2831-2840.
171. Huang, Y.J., et al., *Why MUC16 mutations lead to a better prognosis: A study based on The Cancer Genome Atlas gastric cancer cohort.* World J Clin Cases, 2021. **9**(17): p. 4143-4158.
172. Dietrich, P., et al., *Molecular crosstalk between Y5 receptor and neuropeptide Y drives liver cancer.* J Clin Invest, 2020. **130**(5): p. 2509-2526.
173. Lv, X., et al., *Neuropeptide Y1 receptor inhibits cell growth through inactivating mitogen-activated protein kinase signal pathway in human hepatocellular carcinoma.* Med Oncol, 2016. **33**(7): p. 70.
174. Medeiros, P.J., et al., *Neuropeptide Y stimulates proliferation and migration in the 4T1 breast cancer cell line.* Int J Cancer, 2012. **131**(2): p. 276-86.
175. Liu, J., et al., *The Novel Methylation Biomarker NPY5R Sensitizes Breast Cancer Cells to Chemotherapy.* Front Cell Dev Biol, 2021. **9**: p. 798221.
176. Korner, M., B. Waser, and J.C. Reubi, *High expression of neuropeptide Y1 receptors in ewing sarcoma tumors.* Clin Cancer Res, 2008. **14**(16): p. 5043-9.
177. van Valen, F., E. Keck, and H. Jurgens, *Neuropeptide Y inhibits vasoactive intestinal peptide- and dopamine-induced cyclic AMP formation in human Ewing's sarcoma WE-68 cells.* FEBS Lett, 1989. **249**(2): p. 271-4.
178. van Valen, F., W. Winkelmann, and H. Jurgens, *Expression of functional Y1 receptors for neuropeptide Y in human Ewing's sarcoma cell lines.* J Cancer Res Clin Oncol, 1992. **118**(7): p. 529-36.
179. Kitlinska, J., *Neuropeptide Y in neural crest-derived tumors: effect on growth and vascularization.* Cancer Lett, 2007. **245**(1-2): p. 293-302.
180. Tilan, J.U., et al., *Hypoxia shifts activity of neuropeptide Y in Ewing sarcoma from growth-inhibitory to growth-promoting effects.* Oncotarget, 2013. **4**(12): p. 2487-501.

181. Tilan, J.U., et al., *Systemic levels of neuropeptide Y and dipeptidyl peptidase activity in patients with Ewing sarcoma--associations with tumor phenotype and survival.* Cancer, 2015. **121**(5): p. 697-707.
182. Tilan, J. and J. Kitlinska, *Neuropeptide Y (NPY) in tumor growth and progression: Lessons learned from pediatric oncology.* Neuropeptides, 2016. **55**: p. 55-66.
183. Lu, C., et al., *Dipeptidyl peptidases as survival factors in Ewing sarcoma family of tumors: implications for tumor biology and therapy.* J Biol Chem, 2011. **286**(31): p. 27494-505.
184. Hong, S.H., et al., *High neuropeptide Y release associates with Ewing sarcoma bone dissemination - in vivo model of site-specific metastases.* Oncotarget, 2015. **6**(9): p. 7151-65.
185. Lu, C., et al., *Hypoxia-activated neuropeptide Y/Y5 receptor/RhoA pathway triggers chromosomal instability and bone metastasis in Ewing sarcoma.* Nat Commun, 2022. **13**(1): p. 2323.
186. *R2: Genomics Analysis and Visualization Platform available from http://r2.amc.nl.*
187. Sauvageau, M. and G. Sauvageau, *Polycomb group genes: keeping stem cell activity in balance.* PLoS Biol, 2008. **6**(4): p. e113.
188. Posfai, E., et al., *Polycomb function during oogenesis is required for mouse embryonic development.* Genes Dev, 2012. **26**(9): p. 920-32.
189. Yan, Q., et al., *Emerging role of RNF2 in cancer: From bench to bedside.* J Cell Physiol, 2021. **236**(8): p. 5453-5465.
190. Zhang, Z., et al., *RNF2 ablation reprograms the tumor-immune microenvironment and stimulates durable NK and CD4(+) T-cell-dependent antitumor immunity.* Nat Cancer, 2021. **2**(10): p. 1018-1038.
191. Xie, D., et al., *Emerging Role of E2F Family in Cancer Stem Cells.* Front Oncol, 2021. **11**: p. 723137.
192. Bracken, A.P., et al., *E2F target genes: unraveling the biology.* Trends Biochem Sci, 2004. **29**(8): p. 409-17.
193. Johnson, D.G. and R. Schneider-Broussard, *Role of E2F in cell cycle control and cancer.* Front Biosci, 1998. **3**: p. d447-8.
194. Kent, L.N. and G. Leone, *The broken cycle: E2F dysfunction in cancer.* Nat Rev Cancer, 2019. **19**(6): p. 326-338.
195. Schwentner, R., et al., *EWS-FLI1 employs an E2F switch to drive target gene expression.* Nucleic Acids Res, 2015. **43**(5): p. 2780-9.
196. Bailey, K., et al., *Emerging novel agents for patients with advanced Ewing sarcoma: a report from the Children's Oncology Group (COG) New Agents for Ewing Sarcoma Task Force.* F1000Res, 2019. **8**.
197. Guenther, L.M., et al., *A Combination CDK4/6 and IGF1R Inhibitor Strategy for Ewing Sarcoma.* Clin Cancer Res, 2019. **25**(4): p. 1343-1357.
198. Dowless, M., et al., *Abemaciclib Is Active in Preclinical Models of Ewing Sarcoma via Multipronged Regulation of Cell Cycle, DNA Methylation, and Interferon Pathway Signaling.* Clin Cancer Res, 2018. **24**(23): p. 6028-6039.
199. Jiang, X., et al., *E2F1 promotes Warburg effect and cancer progression via upregulating ENO2 expression in Ewing sarcoma.* Mol Med Rep, 2022. **26**(1).
200. Luo, L., et al., *Up-regulation of Transcription Factor 3 Is Correlated With Poor Prognosis in Cervical Carcinoma.* Int J Gynecol Cancer, 2017. **27**(7): p. 1422-1430.
201. Patel, D., S. Chinaranagari, and J. Chaudhary, *Basic helix loop helix (bHLH) transcription factor 3 (TCF3, E2A) is regulated by androgens in prostate cancer cells.* Am J Cancer Res, 2015. **5**(11): p. 3407-21.
202. Li, C., et al., *Hypomethylation-associated up-regulation of TCF3 expression and recurrence in stage II and III colorectal cancer.* PLoS One, 2014. **9**(11): p. e112005.
203. Wu, C.I., et al., *Function of Wnt/β-catenin in counteracting Tcf3 repression through the Tcf3-β-catenin interaction.* Development, 2012. **139**(12): p. 2118-29.

204. Chen, W., et al., *Analysis of Stemness and Prognosis of Subtypes in Breast Cancer Using the Transcriptome Sequencing Data.* J Oncol, 2022. **2022**: p. 5694033.
205. Shibue, T. and R.A. Weinberg, *EMT, CSCs, and drug resistance: the mechanistic link and clinical implications.* Nat Rev Clin Oncol, 2017. **14**(10): p. 611-629.
206. Ge, Y., et al., *Stem Cell Lineage Infidelity Drives Wound Repair and Cancer.* Cell, 2017. **169**(4): p. 636-650.e14.
207. Iyalomhe, O. and M.D. Farwell, *Immune PET Imaging.* Radiol Clin North Am, 2021. **59**(5): p. 875-886.
208. Gotthardt, M., et al., *Imaging of inflammation by PET, conventional scintigraphy, and other imaging techniques.* J Nucl Med Technol, 2013. **41**(3): p. 157-69.
209. Chouaib, S., et al., *Hypoxic stress: obstacles and opportunities for innovative immunotherapy of cancer.* Oncogene, 2017. **36**(4): p. 439-445.
210. Knowles, H.J., et al., *Hypoxia and hypoglycaemia in Ewing's sarcoma and osteosarcoma: regulation and phenotypic effects of Hypoxia-Inducible Factor.* BMC Cancer, 2010. **10**: p. 372.