Epigenetic patterns and methylation-based models for robust outcome prediction in osteosarcoma

Joshua J. Bowers[1], Nikhil Ramavenkat[1 2], Christopher E. Lietz[1 3], Miya Hugaboom[1 4], Aditya Madupur[1 5], Edwin Choy[4 6], Gregory M. Cote[4 6], Dimitrios Spentzos[1 4 6]

1. Division of Orthopedic Oncology, Mass General Brigham Department of Orthopedic Surgery, Boston, MA
2. Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD.
3. Boston University Chobanian and Avedisian School of Medicine, Boston, MA
4. Harvard Medical School, Harvard University, Boston, MA
5. Kenneth P. Dietrich School of Arts and Sciences, University of Pittsburgh, Pittsburgh, PA
6. Division of Hematology/Oncology, Mass General Brigham Cancer Institute, Boston, MA

ABSTRACT

Osteosarcoma (OSA) is a relatively rare but aggressive bone malignancy that primarily affects children and adolescents and has not seen survival improvements in at least three decades. No clinical or pathological prognostic factor has been identified that can lead to improved outcomes. Further, the complex genomic architecture and genetic heterogeneity of OSA and the rare nature of the tumor have significantly hampered identification and validation of molecular predictors. Recent assessment of DNA methylation in OSA has associated patterns with clinically informative outcomes, yet models building on these findings that can be applied in the clinic have not been developed. To address this unmet need, we developed methylation-based models for individualized patient risk prognostication and treatment response prediction at the time of diagnosis. By applying a region-based method to assess methylation, we were able to greatly reduce dimensionality and soften CpG-level noise, yielding stable features for model construction. These models, based on variably methylated regions (VMRs), underwent comprehensive internal cross-validation, and when possible, external validation in a completely independent dataset that was generated using a different technology from the discovery dataset. Further, they were contextualized with an unsupervised exploration of the features, showing widespread global hypermethylation being associated with unfavorable outcomes. Additionally, survival-associated signal appeared to be concentrated within a smaller, more focal subset of features, while in contrast, the chemoresponse signal seemed to come from a larger, more diffuse feature set, potentially suggesting a shift in cell state. An examination of epigenetic age using a methylation-derived epigenetic clock was also shown to provide an orthogonal approach of investigating the methylation landscape with significant association to outcome. Finally, clustering phenotypes were compared between methylation and several other omics platforms (mRNA, miRNA, and copy number variation) and showed very little overlap with one another. Together, these findings establish a foundation for methylation-based outcome stratification at the time of diagnosis and underscore the need for continued investigation including validation in larger and prospective cohorts.

INTRODUCTION

Osteosarcoma (OSA) is the most common primary malignant bone tumor with a predisposition to affect children and young adults and remains a leading cause of cancer mortality in pediatric patients[1,2]. Despite the use of multi-agent MAP chemotherapy (methotrexate, doxorubicin, and cisplatin) coupled with surgical resection[3,4], 5-year survival rates have plateaued for at least three decades[5–9]. Outcome remains especially poor for patients who present upfront with metastatic disease, or those whose tumors are found to be chemoresistant upon primary treatment. Early and accurate prognostication is not yet possible as there are no established outcome-based biomarkers at the time of diagnosis. Pathology-based assessment of response to chemotherapy, which is the only known and widely accepted prognostic factor, can only be obtained following exposure to preoperative chemotherapy, at the time of surgical resection. Further, this histological marker has not resulted in new avenues for treatment that may improve the outcome of patients with poor prognosis, as shown in the negative results of the EURAMOS trial[8,10,11]. This is compounded by the high and often early metastatic potential of the disease. Historically, prior to the implementation of neoadjuvant or adjuvant chemotherapy, when amputation was the only treatment for localized tumors, about 80% of patients still developed later metastases. These factors highlight the need to develop molecular (as opposed to simply histologic or clinical) biomarkers that anticipate long-term outcome and may open new avenues to inform treatment strategies at the time of diagnosis.
The genome of OSA tumors is often dominated by complex structural rearrangements resulting in broad copy-number alterations (deletions, amplifications), numerous other structural variants, and an overall picture that has been called "genomic chaos" [12–17]. It is thought that these rearrangements are brought about by massive mutational events like chromothripsis and chromoplexy [13,18,19]. The genomic landscape is further complicated by whole-genome duplication, which is common in OSA, plausibly resulting from a stress response to significant loss of copy number and resulting loss of heterozygosity [20]. Notably, emerging evidence suggests that a large portion of the genomic complexity is established early in tumorigenesis rather than occurring from ongoing chromosomal instability[21]. In parallel, alterations in DNA methylation are detectable early and appear to co-occur with these architectural changes[22,23].

Within this unique genomic architecture, several characteristic individual recurrent mutations have been identified including alterations such as TP53, RB1, MYC, and ATRX [12,13,24–28]. However, none of them has shown consistent prognostic or predictive value apart from MYC which has only recently shown a convincing association with outcome and is gaining increased traction as a potential future therapeutic target[29–31].

Given the complex structural heterogeneity and the limited yield of single-gene biomarkers, recent expert consensus in the field has emphasized higher order biomarker work, with potential emphasis in the analysis of epigenetic regulation [32–34]. Early evidence has shown CpG site-specific methylation patterns are prognostic of outcome, and global CpG-based methylation patterns are also strongly predictive of response to chemotherapy[35,36]. Across many other cancers, methylation patterns often display a coupling of global hypomethylation and focal hypermethylation of specific oncogenic loci, a pattern that has been linked with disease progression[37,38]. In OSA, in contrast, the effects of global methylation may be less straightforward as apparent by the emerging strong association between global hypomethylation and positive outcomes[35]. Significant gaps still exist in understanding the role of the DNA methylome in OSA and its effect on tumor outcome.

Leveraging the NCI (National Cancer Institute) Therapeutically Applicable Research to Generate Effective Treatments (TARGET) OSA cohort, we conducted a new methylation-focused study. First, using a region-based methylation approach (variably methylated regions; VMRs), we derive unsupervised global subgroups to characterize genome-wide methylation patterns. Region-based approaches exploit the local concordance amongst adjacent CpGs to derive methylation features with greatly reduced dimensionality and improved robustness to CpG-level noise[39,40]. We subsequently build, for the first time, fully supervised and internally validated individualized prognostic and predictive models for long-term survival and response to chemotherapy. We then validate the reproducibility and generalizability of the patterns and supervised models using an independent cohort. Additionally, we investigate epigenetic age, its association with outcome, and whether it provides additional and complementary insights to the global prognostic model. Finally, taking advantage of the multi-omics data provided by the TARGET cohort, in a cross-platform analysis, we assessed whether our DNA methylation-derived prognostic phenotypes are redundant or complementary to phenotypes derived by transcriptomic, miRNA, and copy-number patterns.

RESULTS

Patient Cohorts and Datasets

We analyzed two independent, clinically annotated Osteosarcoma (OSA) DNA methylation datasets and their association with disease outcome. The first dataset comes from the NCI Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative, which is a publicly available resource hosted on the NCI Genomic Data Commons[41] (GDC, https://portal.gdc.cancer.gov/projects/TARGET-OS). This cohort includes 83 patients with comprehensive clinical annotations. Genome-wide array-based methylation profiling was performed on the TARGET samples using the Illumina Infinium HumanMethylation450 BeadChip array (450k array). The second dataset comprises 15 OSA samples from the Albert Einstein College of Medicine (AECM). Comparable clinical variables were also available for this cohort and genome-wide methylation profiling was performed using a HELP-tagging assay. In subsequent sections, the AECM dataset acts as a validation cohort for the findings derived from the TARGET dataset. Detailed information regarding both cohorts and assays have been previously reported[35,36,41].

Variably Methylated Regions Analysis Reveals a Clinically Relevant Methylation Profile in Osteosarcoma

Derivation of variably methylated regions

To reduce redundancy, multidimensionality, and noise inherent to single-site methylation analyses involving 450K CpG sites, we processed and analyzed the TARGET cohort methylome by identifying 4,427 regions of the genome with high inter-sample methylation variability and high intra-region concordance among CpG sites, termed variably methylated regions (VMRs, Supplementary Data 1)[40]. These VMRs were derived from methylation β-values generated from raw intensity files and processed using standard techniques as described in methods**.** This collection of VMRs is henceforth called the "global VMR profile". The profile is distributed across all chromosomes and appears to be overrepresented in chromosome 13 and underrepresented in chromosome 9 (Figure 1a) with each VMR averaging around 366 base pairs wide and encompasses approximately 5 CpG sites. Within each of the VMRs, CpG sites are aggregated into a single average methylation value. Additional characteristics of the VMR distribution are provided in Supplementary Data 1.

Unsupervised association of the global VMR profile with clinical outcomes

The global VMR profile was then analyzed using UMAP coupled with k-means clustering, resulting in two molecularly distinct clusters: cluster 1 (C1, unfavorable) and cluster 2 (C2, favorable) (Figure 1b). Samples in cluster 1 were predominately hypermethylated while samples in cluster 2 tended to be hypomethylated (Figure 1c). Global clusters were associated with recurrence-free survival (RFS, log-rank $p = 0.049$; median RFS: C1 26.7 mo., C2 104.7 mo.; Figure 1d) and showed a near-significant association with overall survival (OS, log-rank $p = 0.063$; median OS: C1 94.7 mo., C2 not reached; Figure 1f), suggesting a link between the hypermethylated subgroup and relatively poor outcome. After stratifying for metastasis status at

diagnosis (MetDx), the association remained and showed stronger statistical significance for both RFS (RFS stratified log-rank p = 0.0040; non-metastatic median RFS: C1 65.5 mo., C2 not reached; metastatic median RFS: C1 2.3 mo., C2 26.7 mo.; Figure 1e) and OS (OS stratified log-rank p = 0.0031; non-metastatic median OS: C1 not reached, C2 not reached; metastatic median OS: C1 9.5 mo., C2 53.6 mo.; Figure 1g). In addition, the global clusters showed significant association with response to neoadjuvant chemotherapy (chemoresponse) (χ^2 p = 0.0006, OR = 15.11, 95% CI = 3.22-70.84) in an analysis of 43 samples. Several important clinicopathologic variables including sex, age at diagnosis (AgeDx), mutations in several relevant oncogenes or tumor suppressor genes, and expression levels of TERT and ATRX were evaluated as potential confounders in relation to global clustering (Figure 1e); none of which showed a significant association.

Functional subset analysis of global VMR profile

We next investigated whether subgrouping by specific functional genomic regions, called resort types, provided further biologically informed insights into the effects of methylation on RFS. The distribution of VMRs by resort type categories can be found in Supplementary Data 2. Unsupervised UMAP coupled with DBSCAN clustering was then applied on each of the resort-type subgroups, where 2-cluster solutions were consistently found. Of the four resort-type subgroup analyses, only the open sea region VMRs gave a significant RFS association (Figures 2a-d). However, after stratifying for MetDx, the other three subgroup clustering analyses also become significant in relation to RFS (Figures 2e-h).

Enhancer subset analysis of global VMR profile

Next, we investigated a subset of the global VMRs profile that mapped to enhancer elements and their association with RFS. To identify tissue- and disease-specific enhancer regions relevant to osteosarcoma, we developed a catalog of enhancers from an independent H3K27ac ChIP-seq dataset, which we used as the basis for our analyses. Mapping these enhancers to the global VMR profile, we observed 225 enhancer-linked VMRs. Unsupervised UMAP coupled with DBSCAN clustering was then applied, and a two-cluster solution was observed. A trending, but non-significant RFS association (log-rank p = 0.14; Figure 2i) which became significant after stratification for MetDx (stratified log-rank p = 0.0093; Figure 2j), was noted for this clustering.
Additionally, the enhancer subgroup was further subdivided by resort type status, mirroring the analysis done on the undivided global VMR profile. Unsupervised clustering again yielded clear two-cluster solutions across all subgroups. Among enhancer-linked VMRs, those located in CpG island, shores, and open sea regions showed significant associations with RFS (log-rank p = 0.0088, 0.029, 0.0012, respectively; Figures 2k-m), while shelf-localized VMRs were observed to be insignificant, yet trending (log-rank p = 0.087; Figure 2n). After stratifying each of these analyses for MetDx, all regions-specific enhancer subgroups remained or became significant (stratified log-rank p = 0.050, 2.1×10^{-5}, 0.017, 0.031, respectively; Figures 2o-r).

Supervised Analysis of Variably Methylated Regions Reveals Strong Prognostic Risk Stratification

Univariable survival analysis of global VMR profile

We next sought to quantify the prognostic associations observed in our unsupervised analysis to develop an individualized prognostic model. We first performed Cox Proportional Hazards (CPH) Regression analysis of individual VMRs for association with survival. This analysis resulted in 61 VMRs associated with RFS and 1 associated with OS ($q < 0.05$, Supplementary Data 3). Due to fewer events and subsequently lower power, the OS stringency threshold was relaxed (FDR $q < 0.1$), which yielded 271 OS-associated VMRs (Supplementary Data 3). When correcting these analyses for MetDx, a higher number of VMRs (992 and 1464 VMRs, FDR $q < 0.05$, Supplementary Data 4) were observed to be significant.

Supervised risk prognostication

We employed a supervised CPH model to assign risk scores to each sample using a leave-one-out cross validation (LOOCV) approach (Supplementary Data 5). A signed average methylation variable was calculated per sample by weighing each univariately significant VMR's methylation value by the direction of its univariable Cox Regression coefficient, then averaging across those VMRs. Univariately significant VMRs and the signed average variable are re-calculated at each iteration in a fully cross-validated loop where the resulting values serve as input to the supervised CPH model (see details in Methods). The aggregated set of risk scores, defined by the signed average cross validation scheme, were then used as a fixed predictor to fit a final CPH model, resulting in a hazard ratio for RFS of 2.15 ($p = 3.09 \times 10^{-7}$, 95% CI = 1.60-2.87). When correcting the model for MetDx, the hazard ratio increased to 2.49 ($p = 1.57 \times 10^{-8}$, 95% CI = 1.81-3.42).
We then dichotomized risk scores at the median to define high- and low-risk groups. Kaplan-Meier survival analysis demonstrated significance stratification for both RFS (log-rank $p = 0.0052$, Figure 3a) and OS (log-rank $p = 0.015$, Figure 3g). The risk groups were further stratified for MetDx resulting in an increased association significance (RFS: stratified log-rank $p = 3.2 \times 10^{-4}$, Figure 3b; OS: stratified log-rank $p = 5.6 \times 10^{-4}$, Figure 3h). Additionally, we applied a time-dependent receiver operating characteristic (tdROC) analysis on the risk scores produced for both RFS (Figures 3c-d) and OS (Figures 3i-j) at 1- and 5-year time points. For RFS, the tdROC area under the curve (AUC) values were 0.765 and 0.648 at 1- and 5-year time points ($p = 3.3 \times 10^{-4}$, 0.043; 95% CI = 0.620-0.909, 0.505-0.791), while the AUC values for OS were 0.771 and 0.634, respectively ($p = 0.0072$, 0.081; 95% CI = 0.573-0.968, 0.483-0.785). When we corrected the risk scores for MetDx (Figures 3e–f, Figures 3k-l), resulting in improved model performance at both endpoints. At 1- and 5-year time points, the MetDx-corrected RFS model achieved AUC values of 0.810 and 0.721 ($p = 2.4 \times 10^{-5}$, 5.8×10^{-4}; 95% CI = 0.666-0.954, 0.595-0.847), and the MetDx-corrected OS model AUC values were 0.940 and 0.721 ($p = 6.4 \times 10^{-22}$, 2.3×10^{-5}; 95% CI = 0.850-1.00, 0.645-0.894).

Supervised Analysis of Variably Methylated Regions Strongly Predicts Response to Neoadjuvant Chemotherapy

Univariable chemoresponse analysis of global VMR profile

We next assessed the global profile's association with chemoresponse. Chemoresponse annotations were available for only a subset of the TARGET cohort, thus this analysis was restricted to 43

samples. In this univariable analysis, 952 VMRs were significantly associated with chemoresponse (FDR $q < 0.05$; Supplementary Data 6).

Supervised chemoresponse predictions

Treating chemoresponse as a binary endpoint, we trained and tested a single-predictor signed-average logistic regression model using a LOOCV approach, emulating the prognostic CPH regression model above (Figure 4a). Performance of the model was assessed by an ROC analysis resulting in an AUC = 0.770, ($p = 0.0012$, 95% CI = 0.616-0.906).
Given the complexity and broad coverage of the chemoresponse profile (comprising several hundred VMRs), we also tested two alternative methods: L1-penalized logistic regression (Figure 4b), and random forest on individual VMR features (Figure 4c), both showing largely comparable or stronger performance, AUC = 0.857, ($p = 2.0 \times 10^{-5}$, 95% CI = 0.725-0.957), and AUC 0.744, ($p = 0.0010$, 95% CI = 0.585-0.892), respectively.
Apart from the sparsity intrinsic to L1 regularization, no explicit univariable feature filtering was conducted for the primary chemoresponse analyses. A sensitivity analysis applying feature filtering at two stringencies (nominal $p < 0.005$ and 0.0025) resulted in negligible differences in model performance. This suggests that chemoresponse methylation markers may reflect a broader cell state as opposed to a narrow mechanism based on a small number of pathways. Comprehensive results for all models and feature filtering stringencies can be found in Supplementary Data 7.

Prognostic and Predictive Methylation Profiles and Models are Reproducible in an Independent Validation Dataset (AECM)

Mapping and derivation of AECM-VMR set

We tested the robustness and prognostic power of the methylation profiles defined on the TARGET dataset using a publicly available, independent dataset from the AECM. This cohort comprised of 15 OSA samples with methylation characterized by a HELP-tagging assay, a distinctly different methylation assay than the Illumina 450k array. We preprocessed the raw data using standard techniques (see Methods for details). Following the preprocessing steps, we mapped the 4,427 TARGET-derived VMRs on to the HELP-tagging probe set and retained VMRs overlapping with at least one probe. This yielded 3,388 cross-platform VMRs, henceforth called the AECM-VMR set (Supplementary Data 8). Using these AECM-VMRs together with the cohort's corresponding clinical variables, we treat the AECM dataset as an external, independent validation cohort for our findings.

Validation of survival analysis

Of the 61 VMRs associated with RFS in the TARGET dataset ($q < 0.05$), 51 mapped to the HELP-Tagging space (Figure 5a), 9 and 6 of which were univariably associated ($p < 0.10$) with 5-year EFS and OS, respectively, in the AECM dataset. Two-group semi-supervised hierarchical clustering using these 51 AECM-VMRs yielded a trend towards differences in 5-year OS (χ^2 $p = 0.10$, Figure 5a) and no statistically significant differences for chemoresponse, 5yr-EFS, or MetDx. In a fully supervised manner, we then tested a signed-average model, where the sign of the original TARGET CPH model coefficients was applied to the AECM methylation values for these 51

VMRs. Importantly, no information from the AECM dataset was used to train this model. The AECM dataset did not provide patient censoring information, and thus the prognostic model was tested using logistic regression on a binarized 5-year survival endpoint. The resulting model, trained on the TARGET cohort and applied to the AECM cohort, ROC analysis achieved an AUC of 0.860 and 0.907 for 5-year EFS and OS, respectively ($p = 0.028$, 0.0065; 95% CI = 0.640-1.00, 0.722-1.00; Figure 6c-d).

Validation of chemoresponse analysis

Of the 952 VMRs associated with chemoresponse in the TARGET dataset, 680 mapped to the HELP-Tagging space (Figure 5b), 24 of which were univariably associated ($p < 0.10$) with chemoresponse in the AECM dataset. Semi-supervised hierarchical cluster of these 680 VMRs yielded two groups with a significant difference in 5-year EFS (χ^2 $p = 0.007$, Figure 5b), while no statistically significant associations were seen with chemoresponse, 5yr-OS, or MetDx. Additionally, methylation patterns seem to strongly differentiate between the two main clusters with hypermethylation being associated with the relatively poor prognosis subgroup. Given the inherent challenges of cross-platform conversion, the very small number of chemoresponse (positive) events, the reduced sample size of the TARGET training set in chemoresponse analyses, and the large number of univariably-significant chemoresponse VMRs, validation of a fully supervised multi-VMR chemoresponse prediction model in the AECM cohort was not feasible.

Epigenetic Age Effectively Predicts Both Survival and Response to Chemotherapy

Horvath clock, epigenetic age, and association with outcome

Prior work within cancer biology links epigenetic age acceleration with increased cancer risk[42]. This prompted us to ask whether assessing epigenetic age via a predefined methylation clock could provide further insights into tumor behavior and aggressiveness. We used the previously published Horvath's epigenetic clock[43], a tissue-agnostic algorithm, to assess both survival and chemoresponse in the TARGET dataset and evaluated whether the resulting epigenetic-age stratification was independent of the global methylation clustering presented above (Figure 1).
We first performed unsupervised analysis on Horvath-clock CpG sites (312/353 retained after QC; see details in Methods) using UMAP coupled with k-means clustering. This resulted in two sample groups (Figure 6a), which did not significantly differ in terms of RFS (log-rank $p = 0.47$; Figure 6b) or chemoresponse (χ^2 $p = 0.14$).

Additionally, using the clock, we then calculated epigenetic age and subsequently ΔAge (epigenetic age - chronological age) for each sample (ΔAge distribution shown in Figure 6c). Epigenetic and chronological age showed no signs of association with one another. Overwhelmingly, the samples trended towards age acceleration, with only 6 of the 83 samples resulting in age deceleration. ΔAge was positively associated with chemoresponse (t-test $p = 0.0043$, median ΔAge difference = 25.49 mo., Figure 6d). When split into two groups by median ΔAge, survival analysis showed a significant association with RFS (log-rank $p = 0.012$, Figure 6e). We then trained a CPH model to evaluate the risk of recurrence with respect to ΔAge, resulting in a significant hazard ratio of 1.31 ($p = 0.0047$, 95% CI = 1.09-1.59).

Horvath clock vs global VMR overlap at CpG level

We then asked the question of whether the effect of ΔAge on survival could be associated with or confounding the main effect of global VMR profile, described in the previous sections. We examined the agreement between the median-split ΔAge groups and our global methylation clusters (Figure 1). Association analysis showed a statistically significant relationship between clusters (χ^2 $p = 2.8 \times 10^{-4}$), though the strength of agreement, after accounting for chance, was low (ARI = 0.168). This suggests the two approaches have some degree of overlap, but they capture largely distinct signals.
Finally, we introduced ΔAge as a covariate in the original supervised CPH model and observed that there was no effect on the prognostic value of the global VMR model, with the hazard ratio remaining unchanged and significant (HR = 2.15 to 2.12, $p = 3.09 \times 10^{-7}$ to 2.90×10^{-5}).

Multi-Omics Comparative Analysis Reveals Non-Redundant Prognostic Information in Relation to Methylation Effect.

Single-platform prognostic association analysis

In addition to the methylation data, several other omics platforms have been made available for the TARGET cohort, including mRNA sequencing (RNA-Seq), miRNA qRT-PCR (miRNA profiling), and copy number (CN) information. We evaluated whether these additional genomic and transcriptomic features offer outcome information that may be redundant or complementary to that of DNA methylation.

We first analyzed each platform separately for prognostic discrimination by performing hierarchical clustering on the TARGET cohort (Figures 7a-c). The RNA-seq partition (k = 2, feature filtering = 90%) was significantly associated with RFS (log-rank p = 0.033, Figure 7d), however, no strong association was noted with chemoresponse. Likewise, miRNA profiling (k = 2, feature filtering = 50%) resulted in a near significant association with RFS (log-rank p = 0.069, Figure 7f), and no association with chemoresponse. Segment-level CN data, analyzed using GISTIC2.0, resulted in 68 regions of significant amplification or deletion across samples. These regions, while not associated with RFS and chemoresponse under two-cluster solutions (Supplementary Data 9a), became strongly significant for both when considering higher-order clustering (RFS: three-cluster log-rank p = 0.0098 (Figure 7h), four-cluster log-rank p = 0.017 (Supplementary Data 9b); Chemoresponse: three-cluster χ^2 p = 0.019, four-cluster χ^2 p = 0.042).

Cross-platform concordance analysis for prognosis.

Finally, we assessed concordance across omics platforms to evaluate the extent of overlapping or complementary information (Figures 7e, g, and i). Methylation was clearly distinct and exhibited no agreement with the other three platforms (RNA-Seq ARI = 0.02, miRNA profiling ARI = 0.01, CN ARI = -0.01). Similarly, CN showed minimal concordance with RNA-Seq and miRNA profiling (ARI = 0.01, ARI = -0.09). And finally, RNA-Seq and miRNA profiling demonstrated slightly more, but still low levels of concordance (ARI = 0.12), indicating partial but not complete redundancy.

DISCUSSION

Treatment outcomes in osteosarcoma have reached a plateau for the last 30 years, since the introduction of combined modality treatment with chemotherapy and surgery, with up to 30-40% of patients experiencing subsequent relapse following treatment of primary tumor. Management of osteosarcoma would greatly benefit from the development of molecular biomarkers predicting long-term outcomes at the point of diagnosis. The only widely accepted prognostic factor—pathologically assessed tumor necrosis—requires months of chemotherapy before assessment potentially missing an early window for improved therapeutic development. Furthermore, pathologic necrosis alone has not led to improved treatment outcomes when used in the large international randomized EURAMOS trial[8,10,11]. The complex structural makeup and profound genomic heterogeneity have obfuscated a clear path toward translational progress in biomarker development. Furthermore, there is now expert consensus in the field, that single gene and single mutation-based outcome biomarkers are unlikely to be useful in OSA [32–34], with c Myc being the only recently recognized exception to this consensus[29–31].

Although recent work by our group effectively tied DNA methylation patterns to outcome in OSA[35], these findings have yet to be translated into robust prognostic and predictive models. In this study, we define models to predict long-term survival and treatment response while simultaneously using updated region-based methylation profiles (VMR profiles) that are more resistant to potential single-CpG noise and have greatly reduced dimensionality and therefore higher statistical robustness. We validate these models internally and also independently, when possible, based on independent data availability. We also perform complementary analyses including an epigenetic age acceleration estimate providing unique insights into the biology of aggressive OSA. Finally, we perform multi-platform assessment across RNA-seq, miRNA expression, and copy number variation data, to assess if DNA methylation provides overlapping or orthogonal prognostic information to alternative omics profiles. Together, our results establish methylation-based models as practical and promising clinical tools for risk stratification at the time of diagnosis and provide insights into the possible role of DNA methylation in modulating outcome in OSA.

We have developed methylation-based models for individualized patient risk prognostication and treatment response prediction at the time of diagnosis, representing the first time such predictive modeling has been successfully applied in this disease. Both models have been internally validated using cross-validation techniques, resulting in robust performance quantified by ROC and tdROC analyses. These metrics demonstrate strong and significant discriminatory ability across all outcome variables (RFS, OS, and chemoresponse). Additionally, we performed independent validation of survival-based models showcasing their generalizability to new cohorts. Notably, the two cohorts used in this analysis provided methylation data that was generated from distinct assay technologies, namely a methylation microarray versus a HELP tagging assay. This further underscores the reproducibility of our findings, given that the performance of our models was robust despite the technical differences and the substantial amount of attrition from cross-assay mapping.

While a fully independent statistical validation was not possible in the chemoresponse model due to dataset limitations, an unsupervised look into the validation cohort's chemoresponse-associated VMR profile shows a picture reminiscent of the TARGET chemoresponse profile with global

hyper methylation associated with suboptimal chemoresponse. Thus far, genome-wide DNA methylation studies in OSA have been largely associative and lacking clearly defined feature sets suitable for downstream validation. Our models set the foundation for outcome-based prognostication and prediction in OSA, filling a much-needed translational gap in the field.

Unsupervised analysis of TARGET VMR profile provides potential insights into the biological context for the prognostic and predictive capabilities of our models. Genome-wide VMR analysis reveals global clusters that are significantly associated with RFS, OS and chemoresponse. Consistent with our prior findings assays with a CpG based analysis of the TARGET cohort, global VMR clusters show an obvious dichotomy between global hyper- and hypomethylation, with the hypermethylated subgroup being associated with unfavorable outcomes. Notably this seems to contrast the typical paradigm within many other cancers where global hypomethylation is linked to unfavorable outcomes, owing to the improper maintenance of methylation state over successive replications[37,44]. While global hypermethylation being associated with unfavorable outcomes is far from common, it has also been shown to be associated with survival and drug resistance in a few other cancers[45]. In some of these other cancer types, demethylating agents have been used effectively in sensitizing resistant cells, highlighting a potential significant clinical implication of this epigenetic state[46,47]. Additionally, *in vitro* treatment of OSA cells with demethylating agents has been shown to inhibit growth by reactivating epigenetically silenced tumor suppressor genes but the possible interplay between these drugs and global hypermethylation in OSA remains to be elucidated[22,48–50]. Notably, several such agents (e.g., decitabine and azacytidine) are already FDA-approved for other hematologic tumors, specifically myelodysplastic syndrome and some leukemias, and therefore represent promising candidates for relatively accelerated therapeutic repurposing in OSA.

Furthermore, univariable associations between methylation features and chemoresponse and survival outcomes provide additional insights into future therapeutic applications. The chemoresponse-associated VMR profile comprised a substantial number of largely concordant features. The diffuse nature of the epigenetic signature raises the possibility that chemoresponse is modulated not by discrete loci, but rather by coordinated changes in cell identity, potentially resulting from the downstream effects of chromothripsis and signifying a discrete global "cell state". Considering what we know about the sensitization of chemoresistant cells in cancers with similar global epigenetic motifs, it is plausible that application of demethylating agents may hold promise for synergy with chemotherapy or restoring chemosensitivity in these states. In contrast, survival analysis produced a much smaller VMR subset, which may point to specific loci that influence long-term outcomes, and thus may require more targeted therapies.

The presence of overt metastasis at the time of diagnosis (MetDx), is a well-established predictor of patient outcome in all solid tumors, including OSA, and serves a crucial role in patient assessment at the point of diagnosis. The evaluation of MetDx showed no significant association with the global VMR clusters, suggesting that they capture some independent biological signal. To further assess potential confounding, we performed stratified and multivariate analysis for MetDx and we observed that methylation profiles remained significant and indeed showed an even stronger effect size. Thus, incorporating MetDx into our prognostication model resulted in improvements in performance.

Analysis of the VMRs that make up the global profile revealed that, while it spans all chromosomes, it shows a non-uniform distribution across the genome, with considerable VMR enrichment and depletion on chromosomes 13 and 9, respectively. Chromosome 13 harbors major tumor-suppressor genes (e.g., RB1, BRCA2) implicated in OSA oncogenesis. Additionally, copy number analysis in the TARGET dataset revealed two focal deletion peaks on chromosome 9 (9p24.3 and 9p21.3), potentially accounting for a proportion of the reduced methylation variability. This non-uniform VMR landscape showcases the unique nature of the OSA methylome and suggests that methylation effects align with structural alterations and common functional loci, rather than arising from indiscriminate genomic variation.

Resort-type differentiation of global and enhancer-linked VMRs offer insight into the spatial localization and potential functional context of the prognostic signal. Resort-type survival analysis showed a stronger signal in open sea regions, though after MetDx stratification all other genomic contexts demonstrated significance as well. At the enhancer level, all regions except shelves showed significant discrimination. After stratification for MetDx, all enhancer-linked VMR regions remained or became significant. Enhancer-linked VMRs located in CpG islands showed increased discriminatory power in comparison to their enhancer-agnostic CpG island VMR counterparts, plausibly reflecting the increased regulatory potential of enhancers in close proximity to gene promoters. Overall, these findings highlight two key insights into the localization of prognostic signals. First, open sea regions, which are often overlooked in functional annotations, consistently provide strong survival associations in both global and enhancer subsets. Second, enhancer status provides additional regulatory context, especially in gene dense regions.

One of the challenges with methylation data is its enormous dimensionality, with microarrays covering hundreds of thousands (and currently millions) of CpGs, which necessitates aggressive false discovery rate (FDR) correction. Notably, human DNA methylation exhibits an abundance of redundant information amongst contiguous CpG sites. By leveraging this redundancy in combination with variation analysis, we can account for the vast majority of the methylation variability using a considerably lower number of features, called variably methylated regions (VMRs). In addition to having a considerably lower dimensionality requiring less stringent FDR correction VMRs also result in models less vulnerable to single CpG-level noise, potentially optimizing their generalizability.

We also present the first ever epigenetic age analysis in OSA, as an orthogonal approach to investigating the methylation landscape. Epigenetic clocks estimate biological age based on methylation patterns at predefined CpG sites and the resulting epigenetic age values have been linked with increased cancer risk[42]. The TARGET dataset samples overwhelmingly trended toward age acceleration, plausibly resulting from the significant dysregulation of the methylome coinciding with oncogenesis. This assertion is further supported by the lack of clear association between biological and chronological age. Cox regression analysis linked ΔAge (biological – chronological) to RFS with a significant hazard ratio, indicating that epigenetic age acceleration may underlie tumor aggressiveness. Grouping samples by median ΔAge revealed significant differences in both RFS and chemoresponse, further showcasing epigenetic age's association with clinical outcomes. Epigenetic clock CpGs did not overlap with the global VMR profiles, and when comparing these groups with the phenotypes detected by global methylation (Figure 1), a weak

(though statistically significant) agreement was noted, suggesting that while the two approaches have some degree of overlap, they capture largely distinct biological signals.

Our comparisons across multiple omics platforms revealed that methylation-based clustering captures a distinct biological signal that is non-redundant with other molecular layers. To assess this, we independently clustered each platform (RNA-seq, miRNA profiling, and CN information) and subsequently compared resulting phenotypes. Even though each of these omics' platforms was associated with survival, ARI analysis revealed little concordance between them. Most specifically also, methylation phenotypes were distinct from all others indicating that global methylation profiling provides a largely unique signal, independent of transcriptomic, miRNA, and copy number profiles. This independence highlights the potential for complementary modeling approaches that integrate multiple molecular layers to enhance prognostic accuracy and biological insight, and to this end, recent studies have shown promise in this respect[29,51]. Nonetheless, the question of whether and to what extent these different molecular platforms offer overlapping versus orthogonal information has not yet been formally addressed and our current study provides a first, though by no means definitive, formal analysis of this question.

Some constraints of our study should be noted. Given the rare nature of OSA, sample size is a perpetual issue. While this was partially solved with the development of the NCI TARGET multi-omics OSA cohort, external validation still requires an additional sizeable dataset. To this end, our independent validation dataset was small with an n=15. Beyond sample size, the TARGET dataset lacks certain annotations for histological subtypes, preventing the investigation of associations with underlying signal, and provides chemotherapy response information for only half of the samples. In terms of controlling for potential confounders, our study still controlled for the most powerful clinical prognostic factor, the presence of metastases at the time of diagnosis. Future studies should also control for the effect of clinical prognostic variables, which have been previously analyzed in published prognostic nomograms[52,53], as well as new molecular markers such as c Myc[29–31].

With respect to our use of VMR instead of CpG based analysis, while this offers important advantages as described above, VMRs summarize methylation signal across contiguous CpGs, which can potentially overgeneralize and mask important site-level effects. Moreover, the requirement for VMRs to cover at least two CpG sites can result in the exclusion of important single-site CpGs and introduce selection bias towards gene dense genomic regions.

The two cohorts that formed the basis of our study were previously generated using two different assay technologies, the Illumina microarray and HELP tagging assays. Future studies will also need to focus on standardizing the assay that may be appropriate for clinical application. This could conceivably be a custom array with a fixed CpG content. Additional newly available options such as enzymatic methyl sequencing (EM sequencing) or Illumina 5-base sequencing, offer important advantages compared to both conventional bisulfite sequencing, and microarrays and may have to be evaluated as well.
Future studies should include the development of a large, independent validation cohort. Importantly, this would also build towards the calibration of current and future predictive and prognostic models and help to establish specific model cutoffs to define clinical risk, which can then be tested prospectively. Further, building on our findings that RNA-seq, miRNA, and copy

number profiling yield largely non-overlapping yet prognostic signals, integrated modeling approaches that combine different molecular layers could enhance both clustering resolution and model predictive accuracy, reflecting a more comprehensive view of osteosarcoma biology. Additionally, further investigation is needed into the relationship between epigenetic age acceleration and OSA tumor behavior, and the mechanisms that underpin it. Finally, future experimental studies should focus on the testing demethylating agents in OSA, especially in the context of enhancing chemoresponse and drug sensitization.

METHODS

Study Cohorts and Data Acquisition

Genome-wide DNA methylation data from 83 osteosarcoma (OSA) patients was obtained from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative[54]. This initiative's publicly available dataset was hosted via the NCI Genomic Data Commons[41] (GDC, https://portal.gdc.cancer.gov/projects/TARGET-OS). Paired RNA sequencing (RNA-seq), miRNA qRT-PCR (miRNA profiling), and segment-level copy number (CN) data were available for 75 of the 83 methylation samples. The TARGET OSA cohort served as the discovery dataset for all primary analyses.

An independent validation dataset was obtained from the Albert Einstein College of Medicine (AECM)[36]. The AECM cohort consisted of 15 OSA patients with attached methylation information as generated by a HELP-Tagging assay[55]. This publicly available dataset was downloaded from the Gene Expression Omnibus (GSE59200). Comparable clinical annotations were available for this cohort, enabling the replication of the primary analyses derived from the TARGET discovery dataset.

Additional information regarding both datasets has been previously reported[35]. All data analyzed in this study was publicly available through respective sources. Ethics approvals and informed consent procedures for data generation have been described previously[35]. All analyses in the present study were performed in accordance with relevant guidelines and regulations.

Methylation Data Preprocessing

Methylation profiling of the TARGET cohort was conducted using Illumina Infinium HumanMethylation450 BeadChip (450k array). Raw IDAT files were preprocessed using the minfi package in R (v1.52.1)[56]. Standard preprocessing included normal–exponential out-of-band (NOOB) correction, which uses control probes for background and dye-bias adjustment, was performed[57]. This was followed by functional normalization, leveraging the principal components of control probes to perform between-array correction and reduce unwanted technical variation[58]. Furthermore, quality control measures were performed, including the removal of probes containing common SNP, probes targeting sex chromosomes, probes targeting cross-reactional sites[59], and probes where expression doesn't differ significantly from background noise (>10% of samples with failed probes). CpG annotations were obtained from the Illumina 450k hg19 annotation file (using IlluminaHumanMethylation450kanno.ilmn12.hg19 annotation package in R). Preprocessed methylation intensities were converted to β values and subsequently transformed into M-values, which more closely approximate the assumptions of parametric statistical models[60]. Unless otherwise stated, all analyses were conducted using M values.

$$\beta \text{ value} = \frac{Methylated\ Intensities}{(Methylated + \ Unmethylated\ Intensities)}$$

$$M\ value = \ \ln\left(\frac{\beta}{1-\beta}\right)$$

Methylation profiles for the AECM validation cohort were generated using the HELP-tagging assay, which produces paired signal intensities from digestion with a methylation-sensitive enzyme (HpaII) and a methylation-insensitive enzyme (MspI). For each probe, the ratio of HpaII to MspI

intensities can be visualized as a point in two-dimensional space, and the angle between this point and the x-axis reflects the degree of methylation[55]. Angles range from 90° (fully unmethylated) to 0° (fully methylated) and were scaled to fall between 0 and 1, serving as pseudo-β values. These values were subsequently logit-transformed into pseudo-M values to facilitate comparability with the TARGET 450k dataset.

Variably Methylated Region (VMR) Derivation

To reduce noise and redundancy, single-CpG measurements (β values) were aggregated into variably methylated regions (VMRs) using DMRcate (v3.2.1)[40]. Regions were defined based on genomic proximity (λ=500) and then filtered to retain those with high inter-sample variability and within-region concordance. Methylation values within said regions were aggregated into a single value by averaging β values across all CpG sites. CpG island context (island, shore, shelf, open sea), termed resort type, were annotated using the Illumina 450k hg19 annotation file.
AECM HELP-tagging methylation probes were then aligned to the TARGET-defined VMRs. Probes that overlapped or fell within 1,000 base pairs of the VMR's boundaries were averaged into a single aggregate methylation value for each VMR, which we collectively called the AECM VMR set.

Enhancer Annotations

Due to their highly tissue-specific nature[61–63], we sought to identify osteosarcoma-relevant enhancers. To do this, we leveraged H3K27ac ChIP-seq data, a histone marker strongly associated with active enhancers[64]. We analyzed publicly available H3K27ac profiles from localized osteosarcoma bone tumors (GSE234997), generated by a research group at Texas A&M[65]. Utilizing these independent profiles, we applied BEDtools (v2.31.1) to identify replicated peaks consistently observed across all three samples, which we designated as osteosarcoma-specific enhancer regions[66]. VMRs overlapping these replicated H3K27ac peaks were classified as enhancer linked. For downstream analyses, enhancer-linked VMRs were further stratified by resort type.

Unsupervised Clustering Analysis

Unsupervised clustering analysis was performed on several profiles using standard workflows. For the global VMR profile in the TARGET cohort, clustering was performed using Uniform Manifold Approximation and Projection (UMAP, umap-learn v0.5.7, n_neighbors = 8 and min_dist = 0.01) in combination with the k-means clustering algorithm (scikit-learn v1.6.1, n_clusters = 2) applied to the UMAP embeddings[67].
To assess whether resort type context influenced clustering, VMRs were stratified into four groups (island, shore, shelf, and open sea) according to Illumina 450k hg19 annotations. VMRs were designated island, shore, shelf, or open sea if 90% or more of its CpG sites belong to the corresponding resort type. Within each subgroup, unsupervised clustering was performed using dimensionality reduction (UMAP) followed by density-based spatial clustering of applications with noise (DBSCAN, v1.2.2)[68]. DBSCAN parameters varied based on subgroup.
VMRs overlapping replicated H3K27ac peaks (see Section 4) were analyzed separately using the same techniques as the resort-type-specific VMR profiles (UMAP + DBSCAN). For consistency,

enhancer-linked VMRs were also stratified by resort type context, and the same clustering approach was applied within each subgroup.

Supervised Predictive and Prognostic Analyses

A univariable survival analysis was conducted on the TARGET global VMR profile testing for associations with RFS using a Cox proportion hazards (CPH) regression model (lifelines v0.28.0)[69,70]. P-values were adjusted for multiple testing error using Benjamini-Hochberg false discover rate (BH FDR)[71] correction and were considered significant at $q < 0.05$.
VMRs that were significantly associated with RFS were used to calculate a signed-average value, as we and others have previously reported[72–77]. The signed average is calculated for each sample by taking the sign of the direction of the univariable association with RFS and multiplying this across the methylation values for that VMR and then subsequently calculating the average across the given sample. To avoid bias and overfitting, all CPH prognostication models were evaluated using a leave-one-out cross validation (LOOCV) framework[78], calculating a risk score for the left-out sample at each iteration. Univariable survival analyses and signed-average scores are re-calculated at each iteration. This supervised framework was conducted with and without the utilization of metastasis at diagnosis (MetDx) as a covariate during fitting and application of the model within the LOOCV loop. Risk groups were defined by splitting samples based on median risk score. Additionally, model performance was evaluated using the time-dependent receiver operating characteristics (tdROC) method[79].
Chemoresponse status was available for a subset of the TARGET cohort and was treated as a binary endpoint (percent necrosis >90%). For patients with available chemoresponse data, an univariable analysis was conducted on the TARGET global VMR profile testing for associations with response to chemotherapy (chemoresponse) by using logistic regression. P-values were adjusted for multiple testing error using BH FDR correction and were considered significant at $q < 0.05$.
Three supervised models were trained and tested under LOOCV including a signed average logistic regression, L1-penalized logistic regression (LASSO), and random forest models. For LASSO and random forest models, individual VMRs were used as features, replacing the single signed average value used in the previous models. Model performance was assessed using ROC analysis, with results quantified by area under the curve (AUC), and supplemented by accuracy-based metrics (available in supplement). Apart from the sparsity intrinsic to L1 regularization, no explicit univariable feature filtering was conducted for the primary chemoresponse analyses, however, a sensitivity analysis was conducted with nominal p-values of 0.005 and 0.0025. Nominal p-values were chosen over FDR-corrected p-values for the chemoresponse analyses due to the extreme variability in the number of significantly associated VMRs between perturbations.
Supervised models trained on the TARGET cohort were applied to the AECM validation cohort. Signed average scores were recalculated for the AECM samples using the direction of the TARGET-derived logistic regression coefficients. LASSO and random forest models were retrained in the TARGET cohort and evaluated in the AECM cohort. Model performance was once again assessed and quantified by ROC and AUC analyses, respectively. Importantly, the AECM dataset did not provide patient censoring information, and thus the prognostic model was tested using logistic regression on binarized 5-year survival endpoints.

Epigenetic Clock Analysis

Epigenetic age was estimated with a pan-tissue methylation clock originally described by Horvath[43] and implemented via pyaging[80]. ΔAge was defined as predicted epigenetic age – chronological age at diagnosis. All epigenetic age and Δage values were z-score normalized. Horvath CpG sites were analyzed via unsupervised clustering using UMAP dimension reduction paired with k-means clustering (k = 2). Association between Δage and chemoresponse was analyzed via t-test. Survival comparison between median-split Δage groups was done via Kaplan-Meier analysis with significance assessed via log-rank test. Agreement between median-split Δage groupings and global VMR clusters was assessed using χ^2 test, and relative strength was interpreted using an Adjusted Rand Index (ARI) score. Finally, risk of recurrence with respect to Δage was assessed using a CPH model[69,70].

Comparative analysis of multiple omics layers

RNA-seq counts for the TARGET cohort were normalized using the regularized log (rlog) transformation implemented in DESeq2 (v1.46.0)[81]. The rlog transformation stabilizes variance across the dynamic range of expression and was applied prior to downstream clustering. miRNA expression data were generated by qRT-PCR and normalized using the $2^{-\Delta Ct}$ method relative to the reference control. Segment-level copy number data was interpreted using GISTIC2.0[82], and results were further log-transformed and z-score normalized. Normalized gene and miRNA expression as well as CN information were used directly in downstream analyses. For both RNA-seq and miRNA expression profiles, features were ranked and filtered by relative variability using the median absolute deviation (MAD) for RNA-seq and standard variance for miRNA profiling. Unsupervised hierarchical clustering was performed using Euclidean distance and Ward linkage methods. Additionally, all clustering methods were evaluated at 2-, 3-, and 4-cluster partitions.
To assess similarity among clustering solutions across all three platforms (methylation, RNA-seq, miRNA profiling, and genomic CN), we calculated the adjusted rand index (ARI) between each pair[83–85]. The ARI measures agreement between two clustering approaches while correcting for random chance. This value ranges from -1 (complete discordance) to 1 (perfect concordance). Where appropriate, the significance of the concordance was evaluated using a permutation test.

General Statistical Analyses

Survival outcomes were analyzed using Kaplan-Meier (KM) survival analysis[86]. Survival differences between two-cluster solutions were evaluated using log-rank tests, with multi-group log-rank tests being applied when k > 2 clusters were compared[87]. Median survival times, hazard ratios, and p-values were then reported. Hazard ratios were estimated using CPH models[69,70]. Binary associations (e.g., chemoresponse) were tested using χ^2 or Fisher's exact tests as appropriate. Where applicable, two-sided 95% confidence intervals are reported for effect-size estimates, hazard ratios, odds ratios, and AUCs. Multiple testing error was addressed using BH FDR method, which was applied to both survival and chemoresponse analyses.
Model performance was evaluated using time-dependent ROC[79] analysis for survival outcomes and standard ROC[88] analysis for binary endpoints, with performance quantified by AUC scores.

Reproducibility and Code Availability

All analyses were conducted using **R (v4.4.3)** and **Python (v3.11.6).** In R, commonly used packages included *minfi*[56], *DMRcate*[40], *DESeq2*[81], *timeROC*[89], and *survival*[90,91]. In Python, core packages included *scikit-learn*[92], *scipy*[93], and *lifelines*[69]. A complete list of 43 packages and version numbers is provided in Supplementary Table 10. Illumina 450k methylation array probe annotations were obtained from the IlluminaHumanMethylation450kanno.ilmn12.hg19 package. Where applicable, scripts and workflows were run with fixed seeds to ensure reportability of clustering and modeling.

Acknowledgements:

The results published here are in part based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (https://ocg.cancer.gov/programs/target) initiative, phs000218. The data used for this analysis are available at https://portal.gdc.cancer.gov/projects. Supported by the Amy Chase McMahon Sarcoma Research Fund and the Casper Colson philanthropic donation to Dimitrios Spentzos and the Sarcoma Program at the MGH Cancer Center.

The authors wish to acknowledge Martin J. Aryee at Dana Farber Cancer Institute, Department of Data Science for his valuable scientific commentary on our analyses.

Competing interests:

The authors declare no competing interests.

Corresponding author:

Correspondence to Dimitrios Spentzos.

REFERENCES

1. Botter, S. M., Neri, D. & Fuchs, B. Recent advances in osteosarcoma. *Curr Opin Pharmacol* **16**, 15–23 (2014).

2. Kun-Peng, Z., Chun-Lin, Z. & Xiao-Long, M. Antisense lncRNA FOXF1-AS1 Promotes Migration and Invasion of Osteosarcoma Cells Through the FOXF1/MMP-2/-9 Pathway. *Int J Biol Sci* **13**, 1180–1191 (2017).

3. Rosen, G. *et al.* Primary osteogenic sarcoma: the rationale for preoperative chemotherapy and delayed surgery. *Cancer* **43**, 2163–2177 (1979).

4. Link, M. P. *et al.* The effect of adjuvant chemotherapy on relapse-free survival in patients with osteosarcoma of the extremity. *N Engl J Med* **314**, 1600–1606 (1986).

5. Bacci, G. *et al.* Neoadjuvant chemotherapy for high-grade central osteosarcoma of the extremity. Histologic response to preoperative chemotherapy correlates with histologic subtype of the tumor. *Cancer* **97**, 3068–3075 (2003).

6. Goorin, A. M. *et al.* Presurgical chemotherapy compared with immediate surgery and adjuvant chemotherapy for nonmetastatic osteosarcoma: Pediatric Oncology Group Study POG-8651. *J Clin Oncol* **21**, 1574–1580 (2003).

7. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2012. *CA Cancer J Clin* **62**, 10–29 (2012).

8. Marina, N. M. *et al.* Comparison of MAPIE versus MAP in patients with a poor response to preoperative chemotherapy for newly diagnosed high-grade osteosarcoma (EURAMOS-1): an open-label, international, randomised controlled trial. *Lancet Oncol* **17**, 1396–1408 (2016).

9. Smeland, S. *et al.* Survival and prognosis with osteosarcoma: outcomes in more than 2000 patients in the EURAMOS-1 (European and American Osteosarcoma Study) cohort. *Eur J Cancer* **109**, 36–50 (2019).

10. Whelan, J. S. *et al.* EURAMOS-1, an international randomised study for osteosarcoma: results from pre-randomisation treatment. *Ann Oncol* **26**, 407–414 (2015).

11. Bielack, S. S. *et al.* Methotrexate, Doxorubicin, and Cisplatin (MAP) Plus Maintenance Pegylated Interferon Alfa-2b Versus MAP Alone in Patients With Resectable High-Grade Osteosarcoma and Good Histologic Response to Preoperative MAP: First Results of the EURAMOS-1 Good Response Randomized Controlled Trial. *J Clin Oncol* **33**, 2279–2287 (2015).

12. Lorenz, S. *et al.* Unscrambling the genomic chaos of osteosarcoma reveals extensive transcript fusion, recurrent rearrangements and frequent novel TP53 aberrations. *Oncotarget* **7**, 5273–5288 (2016).

13. Chen, X. *et al.* Recurrent Somatic Structural Variations Contribute to Tumorigenesis in Pediatric Osteosarcoma. *Cell Rep* **7**, 104–112 (2014).

14. Behjati, S. *et al.* Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. *Nat Commun* **8**, 15936 (2017).

15. Mirabello, L. *et al.* Frequency of Pathogenic Germline Variants in Cancer-Susceptibility Genes in Patients With Osteosarcoma. *JAMA Oncol* **6**, 724–734 (2020).

16. Yamaguchi, T. *et al.* Allelotype Analysis in Osteosarcomas: Frequent Allele Loss on 3q, 13q, 17p, and 18q. *Cancer Res* **52**, 2419 (1992).

17. Kruzelock, R. P., Murphy, E. C., Strong, L. C., Naylor, S. L. & Hansen, M. F. Localization of a novel tumor suppressor locus on human chromosome 3q important in osteosarcoma tumorigenesis. *Cancer Res* **57**, 106–109 (1997).

18. Cortés-Ciriano, I. *et al.* Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet* **52**, 331–341 (2020).

19. Espejo Valle-Inclan, J. *et al.* Ongoing chromothripsis underpins osteosarcoma genome complexity and clonal evolution. *Cell* **188**, 352-370.e22 (2025).

20. Wu, C.-C. *et al.* Immuno-genomic landscape of osteosarcoma. *Nat Commun* **11**, 1008 (2020).

21. Rajan, S. *et al.* Structurally Complex Osteosarcoma Genomes Exhibit Limited Heterogeneity within Individual Tumors and across Evolutionary Time. *Cancer Research Communications* **3**, 564–575 (2023).

22. Kresse, S. H. *et al.* Integrative analysis reveals relationships of genetic and epigenetic alterations in osteosarcoma. *PLoS One* **7**, e48262 (2012).

23. Zhang, Y. *et al.* Global impact of somatic structural variation on the DNA methylome of human cancers. *Genome Biol* **20**, 209 (2019).

24. Wong, F. L. *et al.* Cancer incidence after retinoblastoma. Radiation dose and sarcoma risk. *JAMA* **278**, 1262–1267 (1997).

25. Hawkins, M. M., Draper, G. J. & Kingston, J. E. Incidence of second primary tumours among childhood cancer survivors. *Br J Cancer* **56**, 339–347 (1987).

26. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med* **373**, 2336–2346 (2015).

27. Perry, J. A. *et al.* Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. *Proceedings of the National Academy of Sciences* **111**, (2014).

28. Sayles, L. C. *et al.* Genome-Informed Targeted Therapy for Osteosarcoma. *Cancer Discov* **9**, 46–63 (2019).

29. Jiang, Y. *et al.* Multi-omics analysis identifies osteosarcoma subtypes with distinct prognosis indicating stratified treatment. *Nat Commun* **13**, 7207 (2022).

30. Marinoff, A. E. *et al.* Clinical Targeted Next-Generation Panel Sequencing Reveals *MYC* Amplification Is a Poor Prognostic Factor in Osteosarcoma. *JCO Precis Oncol* https://doi.org/10.1200/PO.22.00334 (2023) doi:10.1200/PO.22.00334.

31. Nagy, M. R. *et al. MYC* amplification and MYC protein expression are poor prognostic markers in pediatric and young adult osteosarcoma. *Cancer* **131**, (2025).

32. Bishop, M. W., Janeway, K. A. & Gorlick, R. Future directions in the treatment of osteosarcoma. *Curr Opin Pediatr* **28**, 26–33 (2016).

33. DuBois, S. G., Corson, L. B., Stegmaier, K. & Janeway, K. A. Ushering in the next generation of precision trials for pediatric cancer. *Science (1979)* **363**, 1175–1181 (2019).

34. Roberts, R. D. *et al.* Provocative questions in osteosarcoma basic and translational biology: A report from the Children's Oncology Group. *Cancer* **125**, 3514–3525 (2019).

35. Lietz, C. E. *et al.* Genome-wide DNA methylation patterns reveal clinically relevant predictive and prognostic subtypes in human osteosarcoma. *Commun Biol* **5**, 213 (2022).

36. Rosenblum, J. M. *et al.* Predictive properties of DNA methylation patterns in primary tumor samples for osteosarcoma relapse status. *Epigenetics* **10**, 31–39 (2015).

37. Johnstone, S. E., Gladyshev, V. N., Aryee, M. J. & Bernstein, B. E. Epigenetic clocks, aging, and cancer. *Science (1979)* **378**, 1276–1277 (2022).

38. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1**, 239–59 (2009).

39. Jaffe, A. E., Feinberg, A. P., Irizarry, R. A. & Leek, J. T. Significance analysis and statistical dissection of variably methylated regions. *Biostatistics* **13**, 166–178 (2012).

40. Peters, T. J. *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**, 6 (2015).

41. Heath, A. P. *et al.* The NCI Genomic Data Commons. *Nat Genet* **53**, 257–262 (2021).

42. Kresovich, J. K. *et al.* Methylation-Based Biological Age and Breast Cancer Risk. *JNCI: Journal of the National Cancer Institute* **111**, 1051–1058 (2019).

43. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, 3156 (2013).

44. Li, J. *et al.* The Prognostic Value of Global DNA Hypomethylation in Cancer: A Meta-Analysis. *PLoS One* **9**, e106290 (2014).

45. Chan, D. W. *et al.* Genome-wide DNA methylome analysis identifies methylation signatures associated with survival and drug resistance of ovarian cancers. *Clin Epigenetics* **13**, 142 (2021).

46. Li, X. *et al.* Promoter Methylation of the MGRN1 Gene Predicts Prognosis and Response to Chemotherapy of High-Grade Serous Ovarian Cancer Patients. *Front Oncol* **11**, (2021).

47. Feng, L., Huang, Y., Zhang, W. & Li, L. LAMA3 DNA methylation and transcriptome changes associated with chemotherapy resistance in ovarian cancer. *J Ovarian Res* **14**, 67 (2021).

48. Asano, N. *et al.* Epigenetic reprogramming underlies efficacy of DNA demethylation therapy in osteosarcomas. *Sci Rep* **9**, 20360 (2019).

49. Osuna, M. A. L. *et al.* Activation of Estrogen Receptor Alpha by Decitabine Inhibits Osteosarcoma Growth and Metastasis. *Cancer Res* https://doi.org/0008-5472.CAN-18-1255 [pii] 10.1158/0008-5472.CAN-18-1255 [doi] (2019) doi:0008-5472.CAN-18-1255 [pii] 10.1158/0008-5472.CAN-18-1255 [doi].

50. Chaiyawat, P. *et al.* Expression profiling of DNA methyl transferase I (DNMT1) and efficacy of a DNA-hypomethylating agent (decitabine) in combination with chemotherapy in osteosarcoma. *J Bone Oncol* **25**, 100321 (2020).

51. Tang, S., Roberts, R. D., Cheng, L. & Li, L. Osteosarcoma Multi-Omics Landscape and Subtypes. *Cancers (Basel)* **15**, 4970 (2023).

52. Ilcisin, L. A., Ma, C., Janeway, K. A., DuBois, S. G. & Shulman, D. S. Derivation and validation of risk groups in patients with osteosarcoma utilizing regression tree analysis. *Pediatr Blood Cancer* **68**, e28834 (2021).

53. Ogura, K. *et al.* Development and external validation of nomograms predicting distant metastases and overall survival after neoadjuvant chemotherapy and surgery for patients with nonmetastatic osteosarcoma: A multi-institutional study. *Cancer* **121**, 3844–3852 (2015).

54. Meltzer, P. S. *et al.* Abstract LB-307: Translational and mechanistic implications of osteosarcoma genomics: A TARGET report. in *Cancer Research* vol. 80 LB-307 (2020).

55. Suzuki, M. *et al.* Optimized design and data analysis of tag-based cytosine methylation assays. *Genome Biol* **11**, R36 (2010).

56. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

57. Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* **41**, e90–e90 (2013).

58. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol* **15**, 503 (2014).

59. Chen, Y. *et al.* Cross-Reactive DNA Microarray Probes Lead to False Discovery of Autosomal Sex-Associated DNA Methylation. *The American Journal of Human Genetics* **91**, 762–764 (2012).

60. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).

61. Benner, P. & Vingron, M. Quantifying the tissue-specific regulatory information within enhancer DNA sequences. *NAR Genom Bioinform* **3**, (2021).

62. Borsari, B. *et al.* Enhancers with tissue-specific activity are enriched in intronic regions. *Genome Res* **31**, 1325–1336 (2021).

63. Pennacchio, L. A., Loots, G. G., Nobrega, M. A. & Ovcharenko, I. Predicting tissue-specific enhancers in the human genome. *Genome Res* **17**, 201–211 (2007).

64. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107**, 21931–21936 (2010).

65. Singh, I., Rainusso, N., Kurenbekova, L., Muruganandham, A. & Yustein, J. Intrinsic epigenetic state of primary osteosarcoma drives metastasis [ChIP-seq]. Preprint at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE234997 (2023).

66. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

67. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J Open Source Softw* **3**, 861 (2018).

68. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)* 226–231 (AAAI Press, Portland, OR, USA, 1996).

69. Davidson-Pilon, C. lifelines: survival analysis in Python. *J Open Source Softw* **4**, 1317 (2019).

70. Prentice, R. L. Introduction to Cox (1972) Regression Models and Life-Tables. in 519–526 (1992). doi:10.1007/978-1-4612-4380-9_36.

71. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).

72. Hill, K. E. *et al.* An imprinted non-coding genomic cluster at 14q32 defines clinically relevant molecular subtypes in osteosarcoma across multiple independent datasets. *J Hematol Oncol* **10**, 107 (2017).

73. Lietz, C. E. *et al.* MicroRNA-mRNA networks define translatable molecular outcome phenotypes in osteosarcoma. *Sci Rep* **10**, (2020).

74. Haibe-Kains, B., Desmedt, C., Sotiriou, C. & Bontempi, G. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics* **24**, 2200–2208 (2008).

75. Haibe-Kains, B. *et al.* A three-gene model to robustly identify breast cancer molecular subtypes. *J Natl Cancer Inst* **104**, 311–325 (2012).

76. Haibe-Kains, B. *et al.* Comparison of prognostic gene expression signatures for breast cancer. *BMC Genomics* **9**, 394 (2008).

77. Haibe-Kains, B. *et al.* A fuzzy gene expression-based computational approach improves breast cancer prognostication. *Genome Biol* **11**, R18 (2010).

78. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J R Stat Soc Series B Stat Methodol* **36**, 111–133 (1974).

79. Heagerty, P. J., Lumley, T. & Pepe, M. S. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics* **56**, 337–344 (2000).

80. de Lima Camillo, L. P. <tt>pyaging</tt> : a Python-based compendium of GPU-optimized aging clocks. *Bioinformatics* **40**, (2024).

81. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

82. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).

83. Rand, W. M. Objective Criteria for the Evaluation of Clustering Methods. *J Am Stat Assoc* **66**, 846–850 (1971).

84. Hubert, L. & Arabie, P. Comparing partitions. *J Classif* **2**, 193–218 (1985).

85. Steinley, D. Properties of the Hubert-Arable Adjusted Rand Index. *Psychol Methods* **9**, 386–396 (2004).

86. Kaplan, E. L. & Meier, P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc* **53**, 457 (1958).

87. Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* **50**, 163–70 (1966).

88. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).

89. Blanche, P., Dartigues, J. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* **32**, 5381–5397 (2013).

90. Terry M. Therneau & Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. (Springer, New York, 2000).

91. Therneau, T. M. A Package for Survival Analysis in R. Preprint at https://CRAN.R-project.org/package=survival (2024).

92. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

93. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272 (2020).

FIGURES AND LEGENDS:

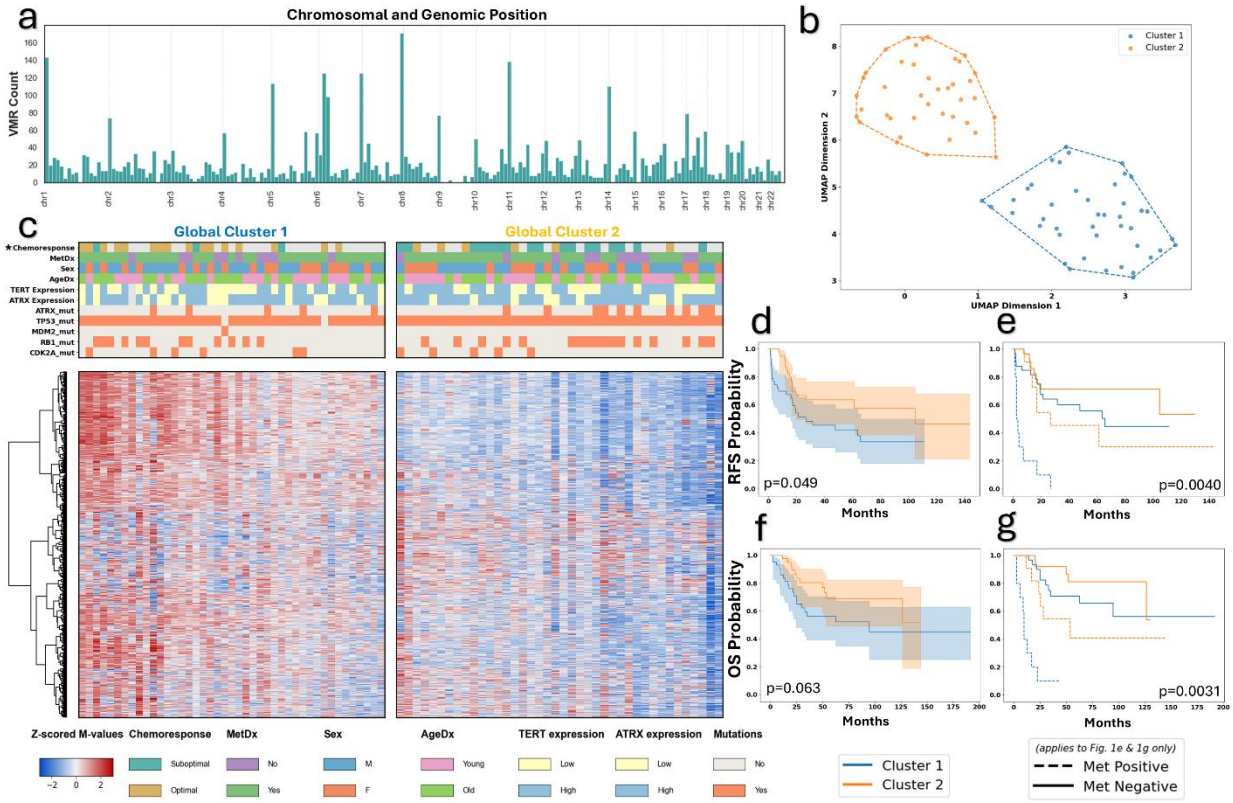

Figure 1: Unsupervised clustering analysis of the TARGET dataset

Figure 1a: Distribution of global VMR profile by genomic coordinates. Figure 1b: Unsupervised UMAP projection of TARGET samples by global VMR profile with k-means clustering. Figure 1c: VMR heatmap of global clusters sorted by average methylation. Color bars display clinical and molecular variable distribution. Asterisk denotes the only significant association (χ^2 p = 0.0006). Figure 1d: RFS analysis of 2-group global VMR UMAP clusters (log-rank p = 0.049; median RFS = 26.7 mo., 104.7 mo.). Figure 1e: MetDx stratified RFS analysis of 2-group global VMR UMAP clusters (stratified log-rank p = 0.0040; non-metastatic: median RFS = 65.5 mo., not reached; metastatic: median RFS = 2.3 mo., 26.7 mo.). Figure 1f: OS analysis of 2-group global VMR UMAP clusters (log-rank p= 0.063; median OS = 94.7 mo., not reached). Figure 1g: MetDx stratified OS analysis of 2-group global VMR UMAP clusters (stratified log-rank p = 0.0031; non-metastatic: median OS = not reached, not reached; metastatic: median OS = 9.5 mo., 53.6 mo.).

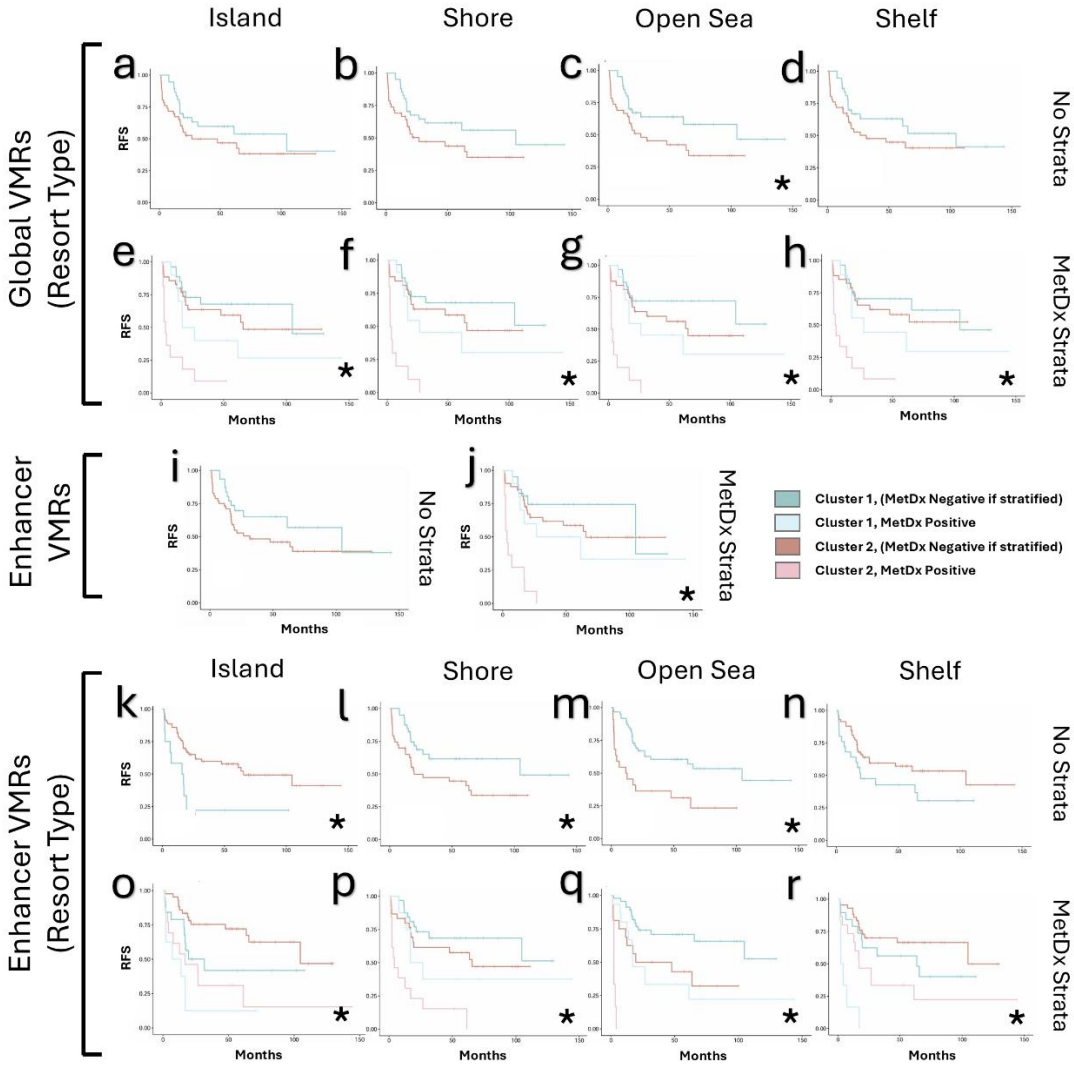

Figure 2: Functional group analysis in the TARGET dataset

Figure 2a-d: RFS analysis of global profile, across resort subgroups (log-rank p = 0.16, 0.065, 0.042, 0.16). Figures 2e-h: MetDx stratified RFS analysis of global profile, across resort subgroups (multi-group log-rank p = 0.040, 0.0070, 0.0040, 0.050). Figure 2i: RFS analysis of enhancer-associated VMRs (log-rank p = 0.14). Figure 2j: MetDx stratified RFS analysis of enhancer-associated VMRs (multi-group log-rank p = 0.00938). Figure 2k-n: RFS analysis of enhancer-associated VMRs, across resort subgroups (log-rank p = 0.0088, 0.029, 0.0012, 0.087). Figure 2o-r: MetDx stratified RFS analysis of enhancer-associated VMRs, across resort subgroups (multi-group log-rank p = 0.050, 2.1×10^{-5}, 0.017, 0.031)

*Note: * denote p <= 0.05*

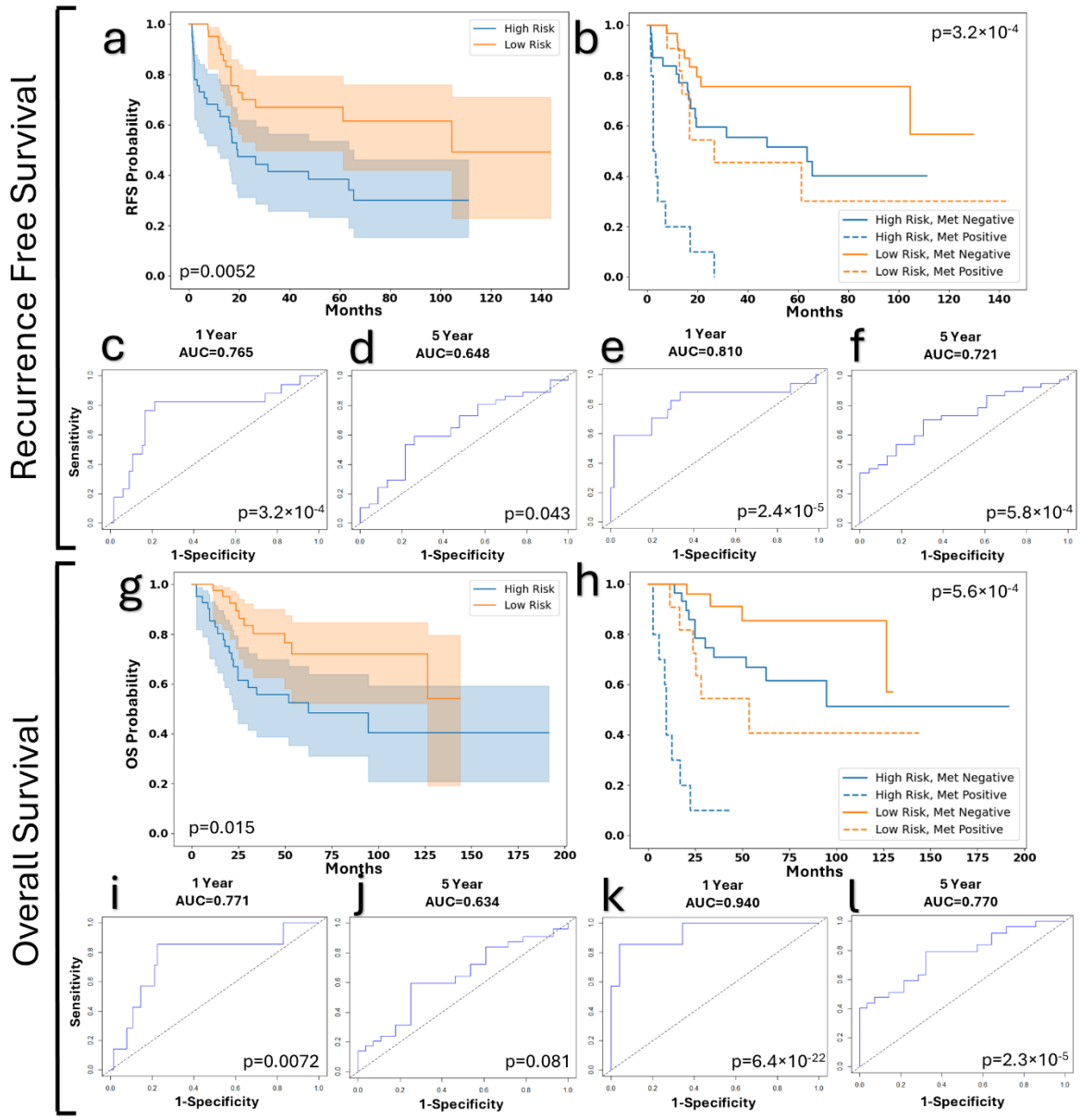

Figure 3: Supervised prognostic survival analysis in the TARGET dataset

Figure 3a: RFS analysis of signed-average methylation risk groups (log-rank p = 0.0052; Median RFS: 19.5 mo. vs. 104.7 mo.). Figure 3b: MetDx stratified RFS analysis of signed-average methylation risk groups (multi-group log-rank p = 3.2×10^{-4}; non-metastatic: median RFS = 63.5 mo., not reached; metastatic: median RFS = 2.3 mo., 26.7 mo.). Figure 3c-d: tdROC analysis of RFS-derived signed-average risk predictions at 1- and 5-year time points (AUC = 0.765, 0.648; p = 3.3 × 10-4, 0.043). Figure 3e-f: tdROC analysis of RFS-derived MetDx corrected signed-average risk predictions at 1- and 5-year time points (AUC = 0.810, 0.721; p = 2.4×10^{-5}, 5.8×10^{-4}). Figure 3g: OS analysis of signed-average risk groups (log-rank p = 0.015; median OS =

62.7 mo. vs. not reached). Figure 3h: MetDx stratified analysis of signed-average risk groups (multi-group log-rank p =5.6 × 10^{-4}; non-metastatic: median OS = not reached, not reached; metastatic: median OS = 9.53 mo., 53.62 mo.). Figures 3i-j: tdROC analysis of OS-derived signed-average risk predictions at 1- and 5-year time points (AUC = 0.771, 0.634; p = 0.0072, 0.081). Figures 3k-f: tdROC analysis of OS-derived MetDx stratified signed-average risk predictions at 1- and 5-year time points (AUC = 0.940, 0.721; p = 6.4 × 10^{-22}, 2.3 × 10^{-5}).

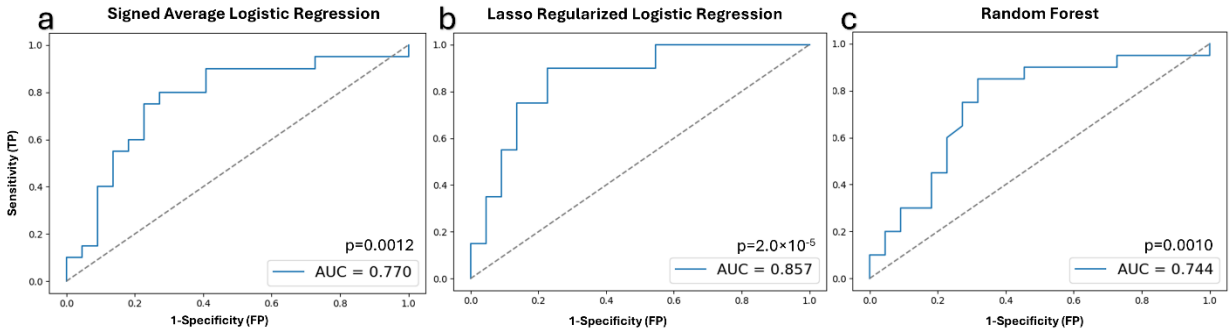

Figure 4: Supervised predictive analysis for Chemoresponse in the TARGET dataset

Figure 4a-c: ROC analysis of (a) signed-average, (b) regularized logistic regression, (c) random forest CR prediction model (AUC=0.770, 0.857, 0.744; p = 0.0012, 2.0 × 10^{-5}, 0.0010)

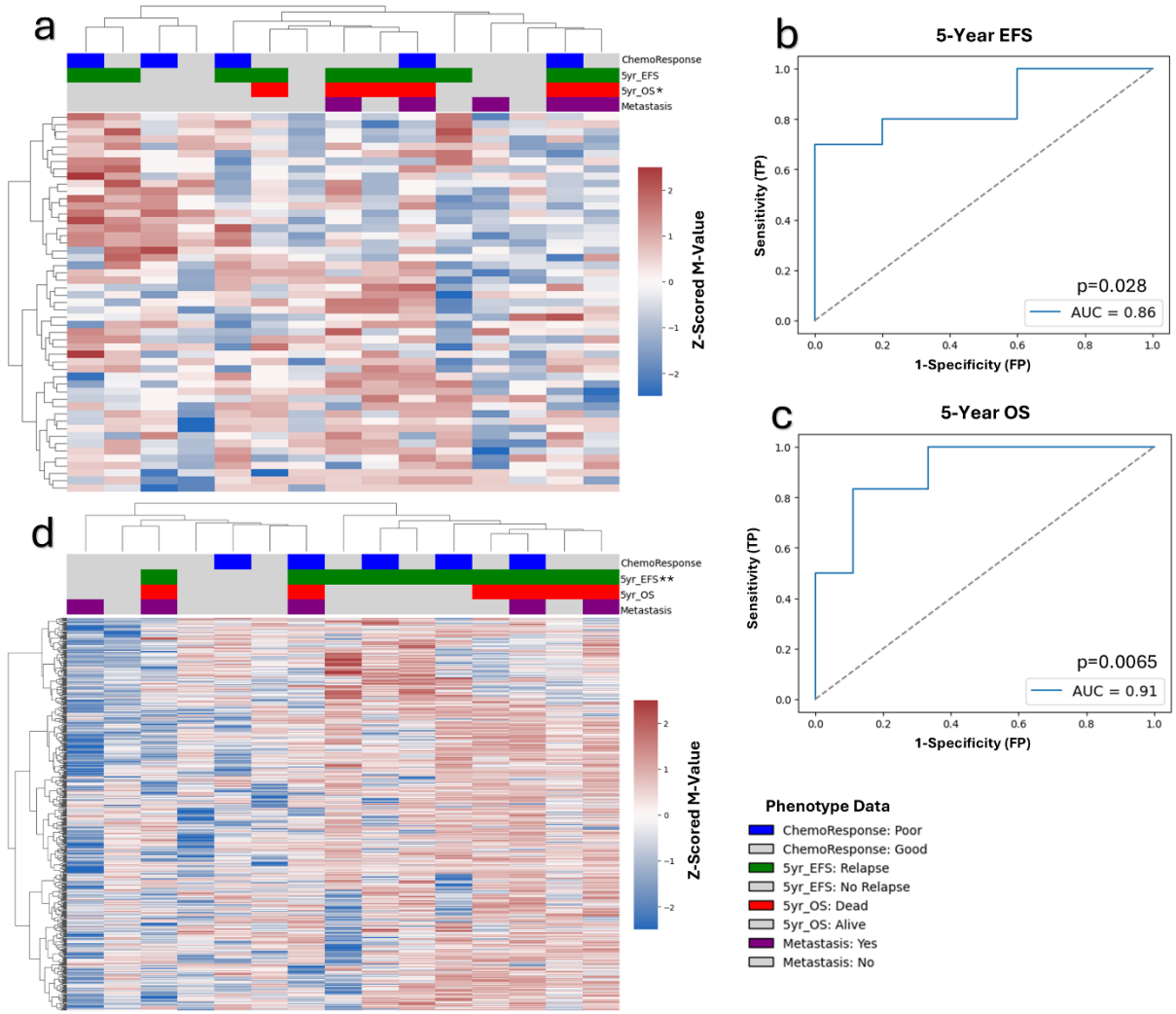

Figure 5: Independent validation analysis in the AECM dataset

Figure 5a: Unsupervised clustering and heatmap of RFS-associated VMRs. Figures 5b-c: ROC analyses of prognostic signed average model trained on TARGET cohort and applied to AECM dataset for 5-year EFS and OS (AUC = 0.86, 0.91; p = 0.028, 0.0065). Figure 5d: Unsupervised clustering and heatmap of CR-associated VMRs

*Note: * denotes p <= 0.1; ** denotes p <= 0.01.*

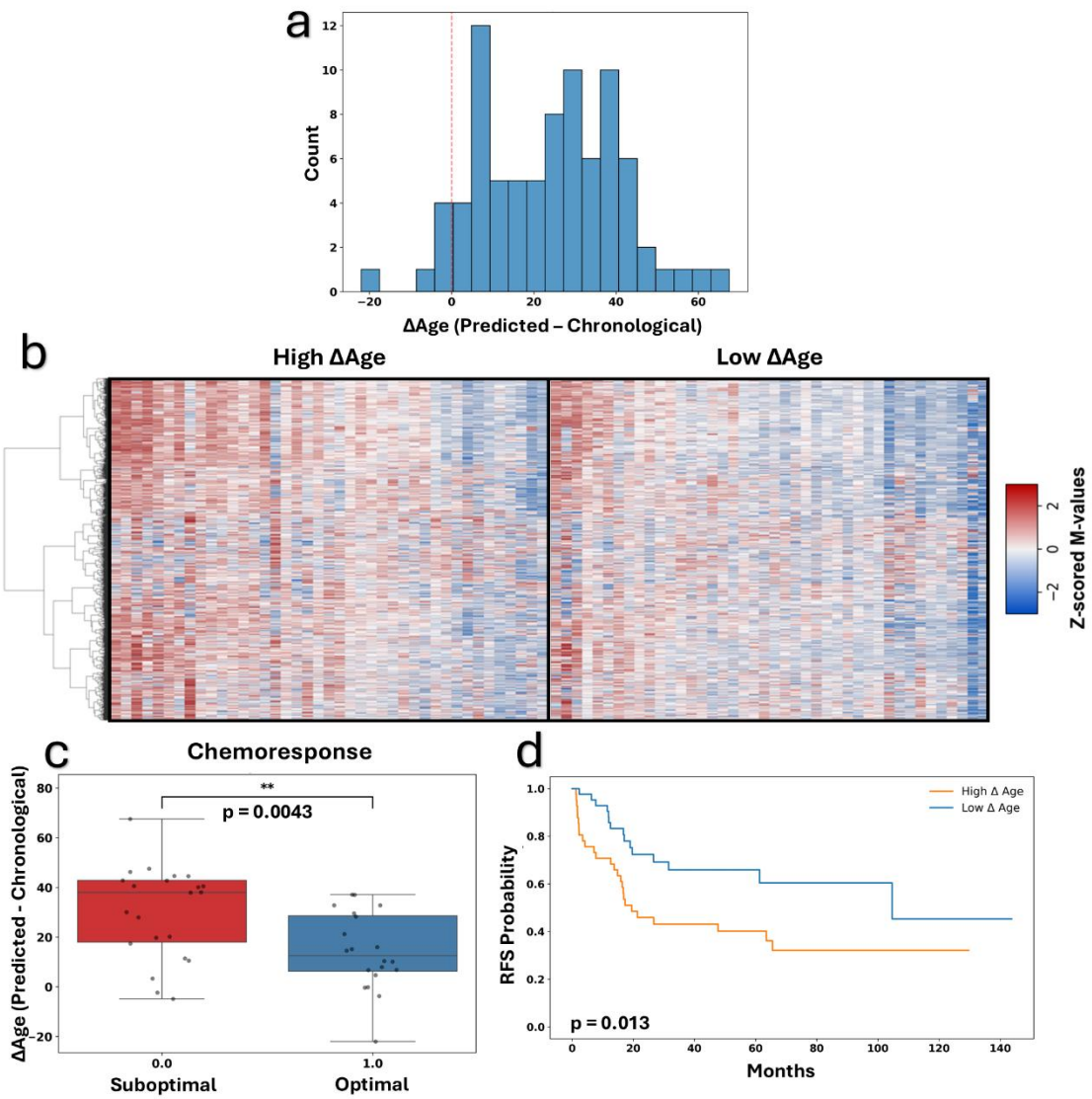

Figure 6: Epigenetic Clock Analysis in the TARGET dataset

Figure 6a: Distribution of ΔAge across cohort as predicted by epigenetic clock (ΔAge=Predicted-Chronological). Figure 6b: VMR heatmap of median-split ΔAge groups sorted by average methylation (Mean: High ΔAge = 0.009, Low ΔAge = -0.582; t-test $p = 4.41 \times 10^{-4}$). Figure 6c: Association analysis between ΔAge and chemoresponse (t-test p = 0.0043, median ΔAge difference between Suboptimal and Optimal treatment response = 25.49 mo.). Figure 6d: RFS analysis of the cohort split at median ΔAge (log-rank p = 0.013).

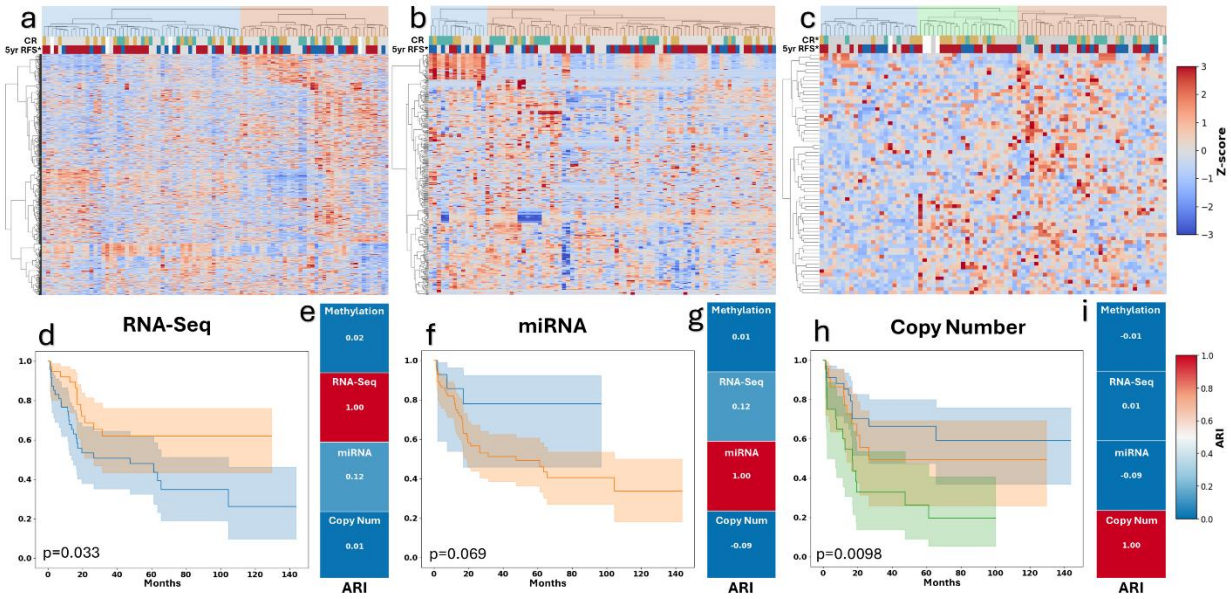

Figure 7: Cross-omics comparison analysis of the TARGET dataset

Figure 7a-c: Unsupervised hierarchical clustering and heatmaps of RNA-seq (a), miRNA-RT-PCR (b), and segment-level copy number (c) platforms. Color bars display chemoresponse (green/brown: top) and 5-year RFS (blue/red: bottom) distribution for each sample. Asterisk denotes association p <= 0.05. Figure 7d-e: RFS (d) and concordance anlaysis (e) for two-group RNA-seq clustering (log-rank p = 0.037; RNA-Seq Adjusted Rand Index (ARI) concordance: Methylation = 0.02, miRNA = 0.12, CN = 0.01). Figure 7f-g: RFS (f) and concordance (g) analysis for two-group miRNA clustering (log rank p = 0.069; miRNA ARI concordance: Methylation = 0.01, RNA-seq = 0.12, CN = -0.09). Figure 7h-i: RFS (h) and concordance (i) analysis for three-group RNA-seq clustering (multi-group log-rank p = 0.0098; CN ARI concordance: Methylation = -0.01, RNA-seq = 0.01, miRNA = -0.09).