

Journal homepage: <https://modernpathology.org/>

## Research Article

# DNA Methylation Profiling of Salivary Gland Tumors Supports and Expands Conventional Classification

Philipp Jurmeister<sup>a,b,\*</sup>, Maximilian Leitheiser<sup>a</sup>, Alexander Arnold<sup>c</sup>, Emma Payá Capilla<sup>a</sup>, Liliana H. Mochmann<sup>a</sup>, Yauheniya Zhdanovic<sup>a</sup>, Konstanze Schleich<sup>a</sup>, Nina Jung<sup>a</sup>, Edgar Calderon Chimal<sup>a</sup>, Andreas Jung<sup>a,b</sup>, Jörg Kumbrink<sup>a</sup>, Patrick Harter<sup>b,d</sup>, Niklas Prenißl<sup>c</sup>, Sefer Elezkurtaj<sup>c</sup>, Luka Brcic<sup>e</sup>, Nikolaus Deigendesch<sup>f</sup>, Stephan Frank<sup>f</sup>, Jürgen Hench<sup>f</sup>, Sebastian Försch<sup>g</sup>, Gerben Breimer<sup>h</sup>, Ilse van Engen van Grunsven<sup>i</sup>, Gerben Lassche<sup>j,k</sup>, Carla van Herpen<sup>j,k</sup>, Fang Zhou<sup>l</sup>, Matija Snuderl<sup>m</sup>, Abbas Agaimy<sup>m</sup>, Klaus-Robert Müller<sup>n,o,p,q</sup>, Andreas von Deimling<sup>r,s</sup>, David Capper<sup>t,u</sup>, Frederick Klauschen<sup>a,b,q</sup>, Stephan Ihrler<sup>v</sup>

<sup>a</sup> Institute of Pathology, Ludwig-Maximilians-Universität München, Munich, Germany; <sup>b</sup> German Cancer Consortium (DKTK), Partner Site Munich, and German Cancer Research Center (DKFZ), Heidelberg, Germany; <sup>c</sup> Institute of Pathology, Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany; <sup>d</sup> Institute of Neuropathology, Ludwig-Maximilians-Universität München, Munich, Germany; <sup>e</sup> Diagnostic and Research Institute of Pathology, Medical University of Graz, Graz, Austria; <sup>f</sup> Department of Pathology, Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Basel, Switzerland; <sup>g</sup> Institute of Pathology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany; <sup>h</sup> Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands; <sup>i</sup> Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>j</sup> Department of Medical Oncology, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>k</sup> Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>l</sup> Department of Pathology, New York University Langone Health, School of Medicine, New York, New York; <sup>m</sup> Institute of Pathology, Friedrich-Alexander-University Erlangen-Nürnberg, University Hospital Erlangen, Erlangen, Germany; <sup>n</sup> Machine Learning Group, Department of Software Engineering and Theoretical Computer Science, Technical University of Berlin, Berlin, Germany; <sup>o</sup> Department of Artificial Intelligence, Korea University, Seoul, South Korea; <sup>p</sup> Max Planck Institute for Informatics, Saarbrücken, Germany; <sup>q</sup> BIFOLD—Berlin Institute for the Foundations of Learning and Data, Berlin, Germany; <sup>r</sup> Department of Neuropathology, University Hospital Heidelberg, Heidelberg, Germany; <sup>s</sup> Clinical Cooperation Unit Neuropathology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany; <sup>t</sup> Department of Neuropathology, Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany; <sup>u</sup> German Cancer Consortium (DKTK), Partner Site Berlin, and German Cancer Research Center (DKFZ), Heidelberg, Germany; <sup>v</sup> DERMPATH München, Munich, Germany

## ARTICLE INFO

### Article history:

Received 5 April 2024

Revised 5 August 2024

Accepted 13 September 2024

Available online 25 September 2024

### Keywords:

salivary gland tumors

DNA methylation

machine learning

diagnostic biomarkers

## ABSTRACT

Tumors of the major and minor salivary glands histologically encompass a diverse and partly overlapping spectrum of frequent diagnostically challenging neoplasms. Despite recent advances in molecular testing and the identification of tumor-specific mutations or gene fusions, there is an unmet need to identify additional diagnostic biomarkers for entities lacking specific alterations. In this study, we collected a comprehensive cohort of 363 cases encompassing 20 different salivary gland tumor entities and explored the potential of DNA methylation to classify these tumors. We were able to show that most entities show specific epigenetic signatures and present a machine learning algorithm that achieved a mean balanced accuracy of 0.991. Of note, we showed that cribriform adenocarcinoma is epigenetically distinct from classical polymorphous adenocarcinoma, which could support risk stratification of these tumors. Myoepithelioma and pleomorphic adenoma form a uniform epigenetic class, supporting the theory of a single entity with a broad but continuous morphologic spectrum. Furthermore, we identified a histomorphologically heterogeneous but

\* Corresponding author.

E-mail address: [Philipp.jurmeister@med.uni-muenchen.de](mailto:Philipp.jurmeister@med.uni-muenchen.de) (P. Jurmeister).



epigenetically distinct class that could represent a novel tumor entity. In conclusion, our study provides a comprehensive resource of the DNA methylation landscape of salivary gland tumors. Our data provide novel insight into disputed entities and show the potential of DNA methylation to identify new tumor classes. Furthermore, in future, our machine learning classifier could support the histopathologic diagnosis of salivary gland tumors.

© 2024 THE AUTHORS. Published by Elsevier Inc. on behalf of the United States & Canadian Academy of Pathology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

The rarity and remarkable diversity of salivary gland tumors pose significant diagnostic challenges for surgical pathologists. The recently revised World Health Organization (WHO) Classification of Head and Neck Tumors encompasses 15 benign and 21 malignant epithelial neoplasms.<sup>1</sup> Traditionally, histopathologic diagnosis of this group of tumors is based on morphology and limited immunohistochemical expression profiles. However, it is well recognized that many tumor entities encompass a wide differentiation pattern, and hence, frequently significant histomorphologic and immunohistochemical overlaps exist between different tumor entities.<sup>2</sup> Furthermore, the distinction of several benign salivary tumor entities from certain low-grade malignant counterparts—such as myoepithelioma and myoepithelial carcinoma—relies on subtle changes (eg, focal invasion) or rather subjective characteristics (eg, limited nuclear atypia or low proliferation).<sup>3</sup> Additionally, our recent work demonstrated that diagnosis of tumors derived from the minor salivary glands is significantly more difficult than that derived from the major glands.<sup>4</sup>

Over the last few years, molecular profiling has been introduced as an additional approach for improved tumor classification in the salivary glands. This has led to the discovery of numerous gene fusion events and mutations that are frequently specific to certain tumor entities.<sup>5</sup> The diagnostic value of these alterations varies with their frequency. Although >90% of secretory carcinomas harbor highly specific *ETV6::NTRK3* fusions, *MYB/MYBL1::NFIB* translocations were only found in approximately 60% of adenoid cystic carcinomas.<sup>6,7</sup> For some entities, such as myoepithelial carcinoma, no diagnostic molecular alterations have been identified to date.

In recent years, DNA methylation has emerged as a further approach for tumor classification and has been initially established for brain and soft tissue tumors.<sup>8–11</sup> DNA methylation occurs when a methyl group is added to cytosine nucleotides at guanine cytosine dinucleotides of the DNA. This process can result in gene repression, especially when it takes place in the gene's promoter region. DNA methylation plays a direct role in regulating the expression profile of various cell types based on their intended functions, resulting in highly tissue-specific global DNA methylation profiles.<sup>12</sup> Because these signatures are generally stable throughout the progression of a tumor, they can be used to classify neoplasms according to their cell of origin. Owing to the complexity of global DNA methylation profiles, machine learning algorithms are trained to identify entity-specific epigenetic profiles and can then be used to classify new cases.<sup>8</sup>

In this study, we provide a reference of the DNA methylation landscape of salivary gland tumors by profiling a comprehensive cohort encompassing 20 different benign and malignant salivary gland tumor entities. On this basis, we demonstrate the feasibility of machine learning–based epigenetic classification of this diagnostically challenging tumor type.

## Methods

### Patients and Samples

For the salivary gland tumor cohort, formalin-fixed paraffin-embedded tissues from 363 cases were mainly retrieved from the consultation archive of DERMPATH München (S.I.; 205 cases); additional cases were acquired from the archives of the Institutes of Pathology of the Charité—Universitätsmedizin Berlin, Ludwig-Maximilians-Universität München, University Hospital Mainz, University Hospital Erlangen, University Hospital Graz, New York University Langone Health, University Medical Center Utrecht, and the Radboud University Medical Center Nijmegen. Central reevaluation was performed by an expert salivary gland tumor pathologist (S.I.).

The cohort consisted of 174 male and 189 female patients with a median age of 58 years (range, 19–91 years). A total of 243 cases (66.9%) were located in the major salivary glands, including 216 cases (59.5%) from the parotid, 21 cases (5.8%) from the submandibular, and 6 cases (1.7%) from the sublingual gland. Approximately 120 cases (33.1%) affected the minor salivary glands and were located in the oral cavity (81; 22.3%), nasal cavity and paranasal sinus (37; 10.2%), and lungs (2; 0.6%).

Generally, only tumors with an unequivocal diagnosis were included based on histomorphology, immunohistology, and, as far as necessary and available, molecular studies. Additional molecular data were available for 137 cases (37.7%), including somatic mutational data for 86 cases (23.7%) and gene fusion data for 66 cases (18.2%). Of note, only cases with positive results in molecular studies were included in the entities mucoepidermoid carcinoma, secretory carcinoma, microsecretory adenocarcinoma, basal cell adenoma, and clear cell carcinoma. In other entities with a lower prevalence of typical molecular alterations, we also included a minority of cases with negative molecular tests (eg, adenoid cystic carcinoma and epithelial-myoepithelial carcinoma). Carcinomas ex pleomorphic adenoma were not included in this study. The frequency of the different entities included in this study is shown in [Supplementary Figure S1](#).

Entities with low tumor cell content (eg, because of intense lymphocytic infiltration) were not included in this study. Furthermore, similar to previous DNA methylation classification studies, we excluded some exceedingly rare tumor entities for which <5 cases could be collected.

In addition to the salivary gland cohort, we compiled a cohort of clinically relevant differential diagnoses. This data set contains cases of 32 clear cell renal cell carcinomas, 17 squamous cell carcinomas of the skin, and 64 squamous cell carcinomas of the upper aerodigestive tract. For 16 of these cases, formalin-fixed paraffin-embedded tissues were retrieved from the Institute of Pathology of the Ludwig-Maximilians-Universität München. For the other cases, raw DNA methylation data were retrieved from the Cancer Genome Atlas through the Genomic Data Commons Data Portal and the Gene Expression Omnibus repositories

GSE105288, GSE61441, GSE92482, GSE171994, GSE140686, and GSE67097.<sup>11,13-16</sup> Detailed information for the cases of this cohort is listed in [Supplementary Table S1](#).

#### Nucleic Acid Extraction

At least 2 unstained slices with a thickness between 2 and 10  $\mu\text{m}$  were used for nucleic acid extraction. Representative tumor areas were identified using light microscopy of consecutive hematoxylin and eosin-stained sections. Macrodissection was performed using sterile surgical blades to achieve tumor cell contents of at least 50%. Semiautomated nucleic acid extraction was performed using Maxwell RSC Blood DNA kits (Promega) and an adjusted protocol on the Maxwell RSC 16 machine (Promega). Concentrations were measured on a Qubit fluorometer (ThermoFisher).

#### DNA Methylation Analysis

Extracted DNA between 100 and 500 ng was used as an input for bisulfite conversion using the EpiTect Fast Bisulfite Conversion Kit (Qiagen), according to the manufacturer's instructions. DNA methylation analysis was performed using Infinium MethylationEPIC BeadChip arrays (Illumina) following standard protocols. Arrays were scanned on an iScan (Illumina) or NextSeq 550 device (Illumina).

#### Copy Number Analysis

Copy number data were extracted from raw DNA methylation data using a modified version of the conumee package.<sup>17</sup> Focal gains or losses were identified using a previously described method, which identifies short alterations encompassing <3 million base pairs.<sup>18</sup> Copy number heat maps were generated using the copynumber package.<sup>19</sup> To assess the overall copy number alteration burden of individual cases, we calculated a chromosomal alteration index (CAI), which we defined as the sum of the length of all altered segments divided by the total genome length.

#### Next-Generation Sequencing Panel

For targeted mutational and gene fusion profiling, 50 ng of DNA or RNA was processed using the Archer Variantplex nNGM 2.0 or the Archer FusionPlex Pan Solid Tumor v2 panel. Libraries were sequenced on a NextSeq 550 device (Illumina), and data were processed using the Archer Suite Analysis software version 7.1.3.

#### DNA Methylation Data Preprocessing

Statistical analysis was performed with R (version 4.3.3). DNA methylation data were processed using a modified version of the minfi package.<sup>20</sup> The *pfilter* and *dasen* functions from the watermelon package were used to filter poor-quality cases.<sup>21</sup> Probes with known cross-reactivity or association with sex chromosomes or single-nucleotide polymorphisms were filtered using the annotation provided by Zhou et al.<sup>22-24</sup> The probes were further reduced to the intersection of available probes in EPIC and 450k Illumina arrays to allow processing of data obtained with both platforms.

#### t-Distributed Stochastic Neighbor Embedding

The 10,000 most variant 5'-cytosine-phosphate-guanine-3 on the salivary gland data set were used as input for t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction using the Rtsne package. To determine the optimal number of principal components (PCs) and the appropriate perplexity for t-SNE analysis, we generated t-SNE plots for a range of input PCs (2, 3, 4, ... 200) and perplexities (1, 1.25, 1.5, ... 40). The resulting time-lapse videos of these t-SNE plots are available in Supplementary Files S1 and S2. Our analysis revealed that the data converge around 20 input PCs, and the plots remain quite similar up to approximately 100 PCs, indicating that the relevant biological information is captured within this range. Beyond this point, as the number of PCs increases, the samples become more dispersed, and previously stable groups become less distinct. Occasionally, individual data points may temporarily appear closer to other groups, but they return to their original groups within a few iterations. These fluctuations suggest the introduction of technical noise. Thus, we limited the number of PCs to 40 to retain the most biologically relevant information. Regarding perplexity, the data converge at approximately 5, and the plots remain consistent for perplexities between 10 and 20, with no significant changes observed at higher values. Therefore, we selected a perplexity of 15. The final t-SNE was generated using the following parameters: input PCs = 40, iterations = 2000, perplexity = 15, and  $\eta = n(\text{cases})/12 = 30.25$ .

#### Development and Evaluation of Machine Learning Classifier

Following a well-established approach, we developed a 2-component classification model that predicts calibrated tumor entity scores from DNA methylation profiles.<sup>25</sup> Specifically, a support vector machine (SVM) classification model with radial basis function kernel (using R package *e1071*) was trained for the prediction of raw class scores, and a multinomial logistic regression model regularized with a ridge penalty (using R package *glmnet*) was trained for their calibration.<sup>26</sup>

The data set used for classifier development and evaluation comprised the salivary gland tumor cohort ( $n = 363$ ) containing 20 classes and additional samples ( $n = 113$ ) for 3 classes of clinically relevant differential diagnoses. The stability and performance of the training procedure were evaluated in a 10 times repeated 10-fold cross-validation with stratified folds on this data set. The final production model was then obtained by repeating the procedure on the full data set.

In the cross-validation loop, the respective training set was first reduced to its 10,000 most variant features. Optimal hyperparameters for the SVM model were then selected by minimizing logloss in another nested 10-fold cross-validation, performing a grid search over the parameter space spanned by  $\gamma = 10^{(-5, \dots, 5)}/10,000$  and  $C = 10^{(-10, \dots, 10)}$ . An SVM model was retrained on the full training set with the selected hyperparameters. The calibration model was developed on the inner cross-validation SVM scores corresponding to the optimal hyperparameters. Model training and hyperparameter selection were performed as previously described for soft tissue classification (with 100 times repeated resampling and training data proportion of 0.7), thus optimizing the calibrated scores for a rejection threshold of 0.9.<sup>11</sup> Finally, the respective test set was reduced to the 10,000 features selected on the training data, and calibrated scores were generated by applying the SVM and *glmnet* models.

## Results

### DNA Methylation Identifies Distinct Epigenetic Salivary Gland Tumor Classes

We first established a comprehensive reference cohort of high-quality DNA methylation data for 363 cases, covering 20 different salivary gland tumor entities and normal salivary gland tissue as a control class. A detailed list of all samples and associated metadata is available in [Supplementary Table S2](#). t-SNE analysis and unsupervised clustering of the 20,000 most variant 5'-cytosine-phosphate-guanine-3 sites were used to identify distinct epigenetic classes ([Fig. 1](#)). We did not observe any apparent differences for tumors originating from different primary sites, and potential bias caused by technical batch effects, such as position on the array, could be excluded ([Supplementary Fig. S2](#)).

The majority of the observed epigenetic classes aligned with the conventional histopathologic classification. Interestingly, upon visual inspection, the overall grouping of different entities in the t-SNE plot seems to reflect their general differentiation pattern, segregating tumors into categories with glandular/ductal, basaloid/myoepithelial, and indeterminate differentiation. However, we observed several important divergent findings:

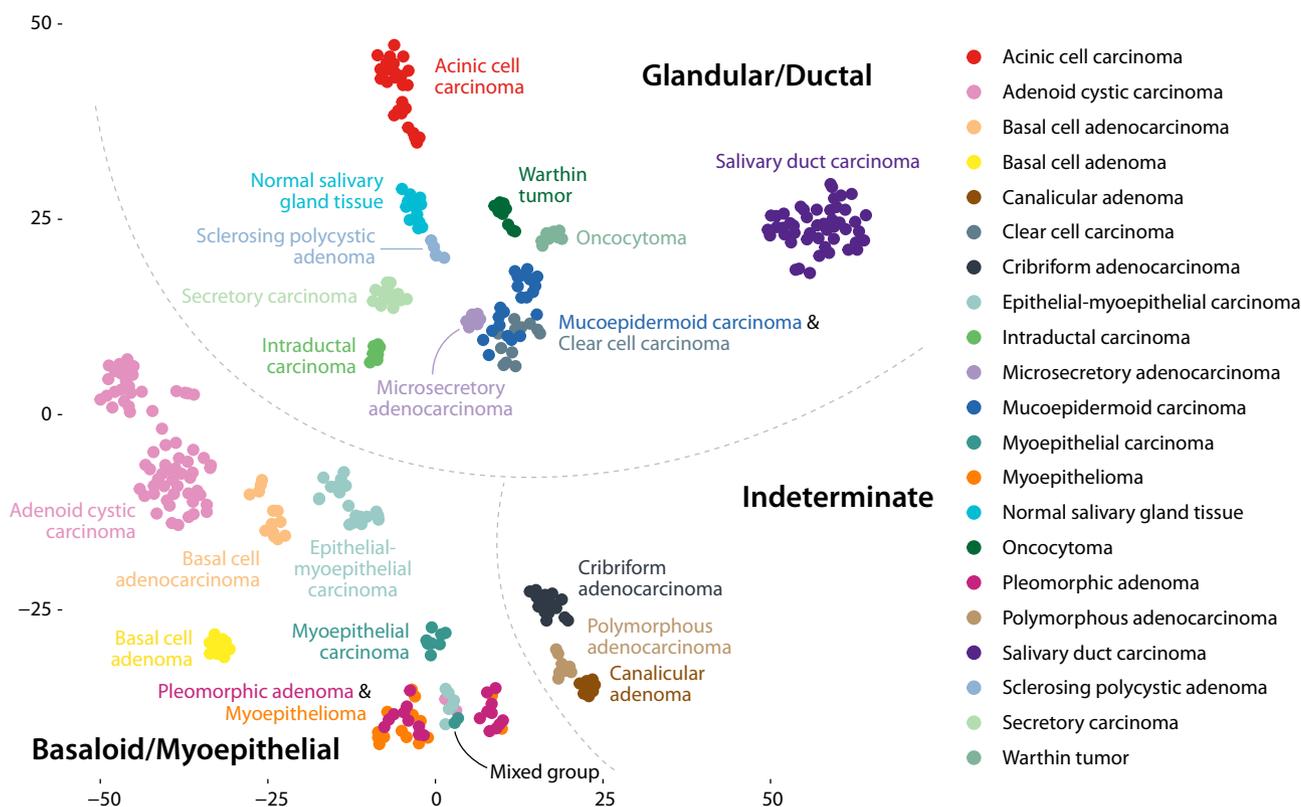
1. According to the current WHO classification, cribriform adenocarcinoma is regarded as a subgroup of polymorphous adenocarcinoma. However, we observed that cribriform adenocarcinoma aggregated in a separate group, that was adjacent, but clearly distinct from classical polymorphous adenocarcinoma.

2. Pleomorphic adenoma and myoepithelioma demonstrated a shared DNA methylation profile. Although these 2 entities split into 2 distinct epigenetic classes, the separation was not in line with histopathologic criteria and resulted in 2 mixed myoepithelioma and pleomorphic adenoma classes. Fusion data on *HMGA2* or *PLAG1* were not available.
3. We identified an additional, mixed, yet clearly distinct epigenetic class consisting of 10 cases. These cases had, on conventional criteria, initially been diagnosed as epithelial-myoepithelial carcinoma (6 cases), myoepithelial carcinoma (2 cases), and adenoid cystic carcinoma (2 cases). This group was closely related to the 2 pleomorphic adenoma/myoepithelioma groups and distinct from all other entities.

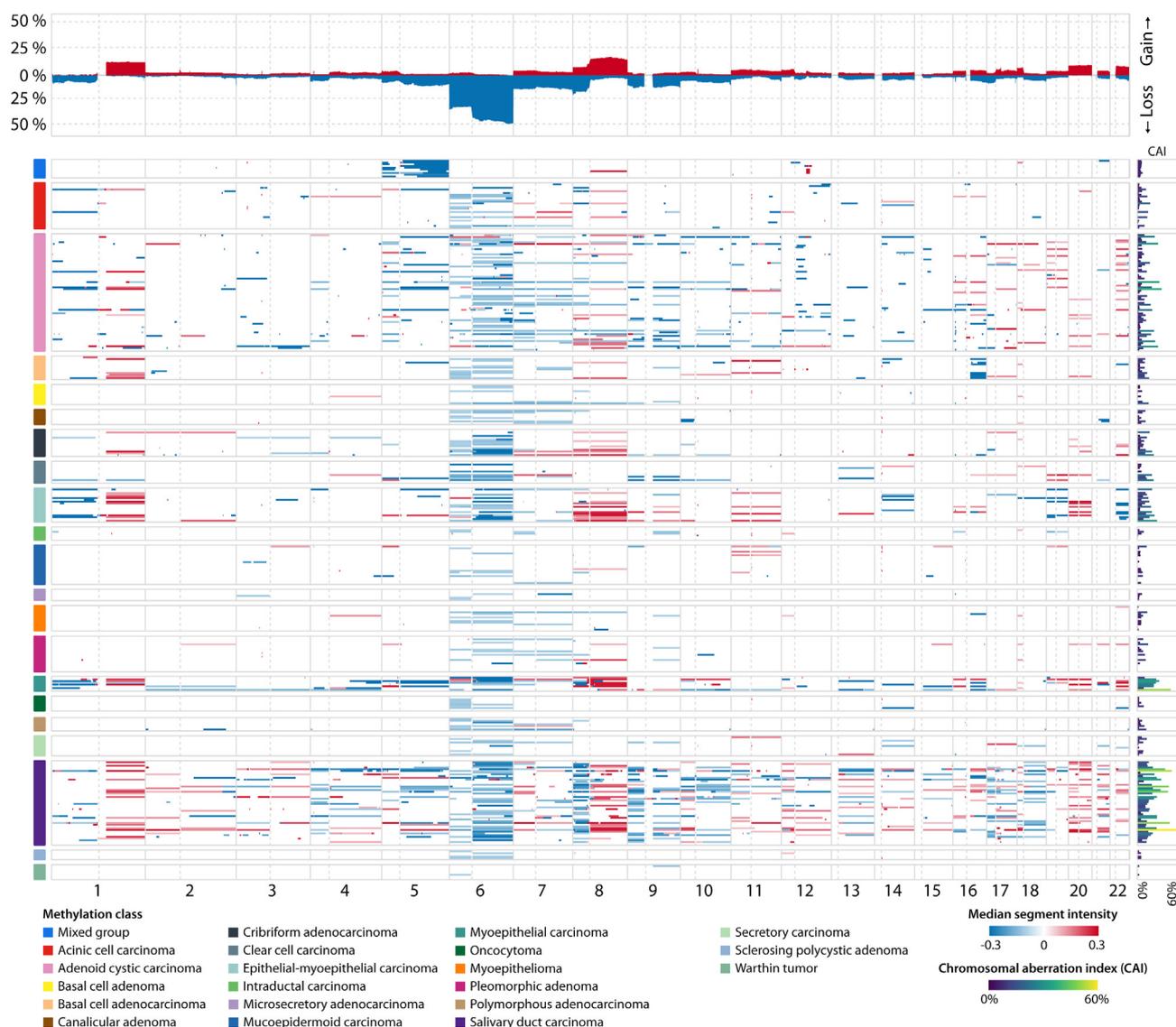
Furthermore, the group of adenoid cystic carcinoma separated into 2 distinct classes, but we observed no association of the 2 classes with histologic growth pattern, grading, or molecular alterations (eg, *MYB* fusion) that were covered by the next-generation sequencing panel used in this study. There was no clear differentiation between mucoepidermoid carcinoma and hyalinizing clear cell carcinoma. This was irrespective of the morphologic appearance (eg, clear cell variant of mucoepidermoid carcinoma) and gene fusion status.

### Correlation of DNA Methylation Classes With Additional Mutational Data

For further assessment of the molecular background of the different DNA methylation classes, the epigenetic classification



**Figure 1.** t-Distributed stochastic neighbor embedding of the salivary gland cohort. Conventional tumor entities according to the World Health Organization classification are annotated by different colors. Upon visual inspection, it seems that classes with common differentiation (basaloid/myoepithelial, glandular/ductal, and indeterminate) tend to group together.



**Figure 2.**

Analysis of global copy number alterations derived from DNA methylation data on the salivary gland tumor cohort. The summary copy number plot on the top of the figure shows the frequency of numeric chromosomal alterations at the respective site of the chromosome (chr). Losses are depicted as negative (blue) and gains as positive (red) deviations from the baseline. The heatmap below shows the detected copy number alterations of each case ordered by methylation class. The barplot on the right illustrates the chromosomal alteration index (CAI), a per-case metric defined as the percentage of the overall genome with numeric copy number alterations.

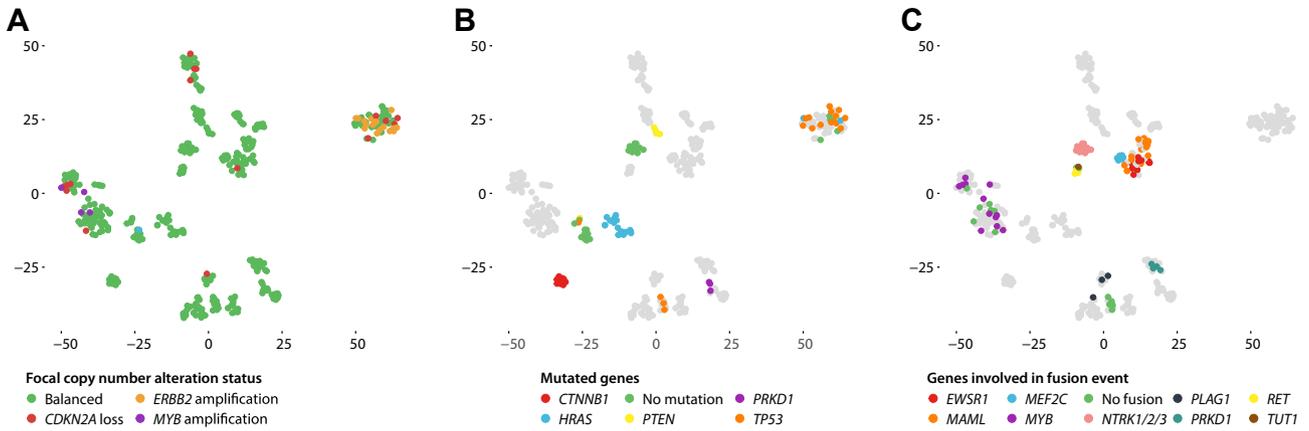
was correlated with available additional molecular data, including copy number profiles, derived from raw DNA methylation data, and panel-based somatic mutation and fusion analysis.

The results from global copy number analysis are shown in [Figure 2](#). Overall, loss of chromosome 6q was the most common chromosomal alteration, with a frequency of up to 50%. Summary copy number profiles for each DNA methylation class are shown in [Supplementary Figure S3](#). Of note, summary copy number profiles of pleomorphic adenoma and myoepithelioma were very similar with comparable rates of losses on chromosomes 6 and 7 and limited number of additional alterations.

As described in the “Methods” section, we calculated a CAI, defined as the fraction of the total genome with chromosomal gain or loss. The overall median CAI across all methylation classes was 5.8%. CAI was low in benign (eg, Warthin tumor; median CAI, 0.02%) and low-grade malignant tumor entities (eg, microsecretory adenocarcinoma; median CAI, 2.0%) but considerably

higher in high-grade malignancies (eg, salivary duct carcinoma; median CAI, 18.8%).

We also screened for focal copy number alterations, defined as gains or losses encompassing <3 million base pairs. As expected, *ERBB2* amplification was common in salivary duct carcinoma (16/48; 33%) and not present in other entities. Loss of tumor suppressor gene *CDKN2A* was detected in 16 of 350 cases (4.6%) but was not entity specific, occurring in adenoid cystic carcinoma, salivary duct carcinoma, acinic cell carcinoma, hyalinizing clear cell carcinoma, and myoepithelial carcinoma. *MYB* amplification was exclusively observed in adenoid cystic carcinoma (4/26; 15%). In t-SNE analysis, none of these alterations defined specific subgroups within their methylation class ([Fig. 3A](#)). There was no case of a positive mutation, which was inconsistent with the conventional tumor diagnosis. Similar to copy number alterations, we observed no epigenetic subgroups associated with recurrent somatic mutations or gene fusion events ([Fig. 3B, C](#)).



**Figure 3.**

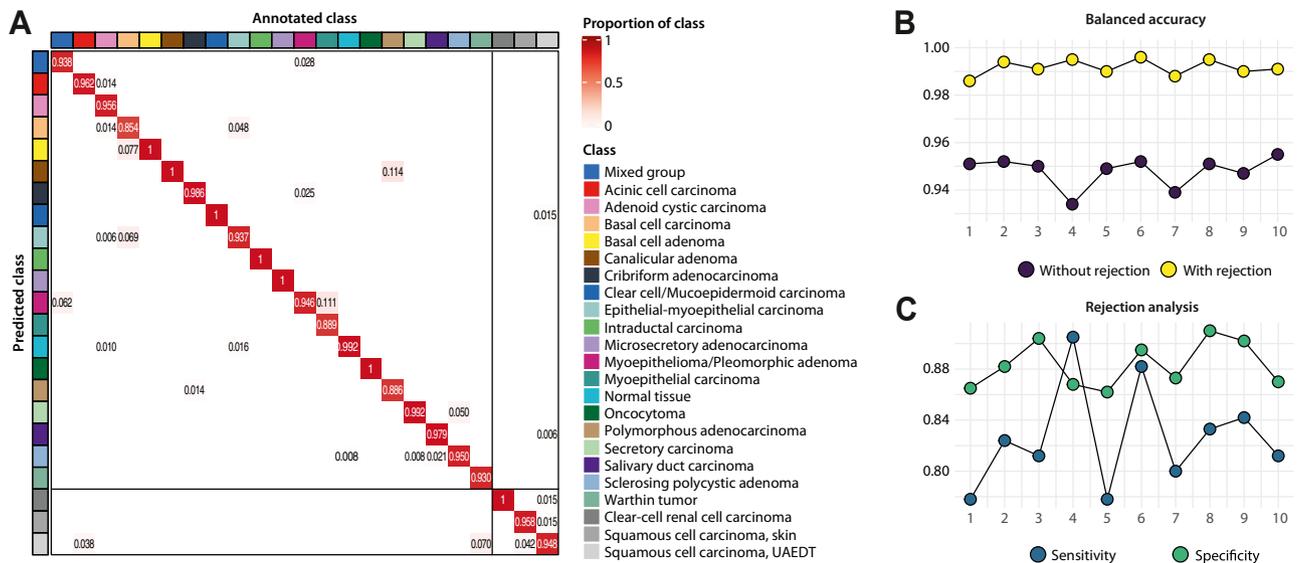
Correlation of epigenetic classes with focal copy number alterations, somatic mutations, and gene fusions on the salivary gland tumor cohort. (A) t-Distributed stochastic neighbor embedding (t-SNE) of the full cohort. Different focal copy number alterations are depicted by different colors. (B) t-SNE plot showing the most frequent and diagnostically relevant somatic mutations. (C) t-SNE plot with diagnostically relevant gene fusion events highlighted by different colors.

**Machine Learning–Based Diagnostic Classifier Accurately Predicts DNA Methylation Classes**

For reliable and rapid classification of diagnostic cases, we developed a machine learning classifier that predicts the salivary gland tumor entity of tissue samples from their DNA methylation profiles and provides calibrated confidence scores for these predictions. Taking into account the findings described above, we defined 20 DNA salivary gland methylation classes as entities for classifier prediction (Fig. 4A). Additionally, we extended the data set by cases of clear cell renal cell carcinoma (n = 32), squamous cell carcinoma of the skin (n = 17), and squamous cell carcinoma of the upper aerodigestive tract (n = 64) as clinically relevant non-salivary gland entities. These tumors are either prone to

metastasize to the salivary glands or can histomorphologically mimic primary salivary gland tumors.

Performance and stability of the training procedure were evaluated in a 10 times repeated 10-fold cross-validation on the full data set. The classifiers achieved a high balanced cross-validation accuracy in each repetition (Fig. 4B) with a mean of 0.95 (SD = 0.004) across repetitions. The pooled confusion matrix of the corresponding predictions is shown in Figure 4A. Rejection of samples with calibrated scores <0.9 further improved the balanced accuracies (Fig. 4B) to near-perfect classification, resulting in a mean of 0.991 (SD = 0.003). The rejection analysis showed both high sensitivity and specificity in detecting incorrect classifications (Fig. 4B) with means of 0.827 (SD = 0.041) and 0.883 (SD = 0.018), respectively.



**Figure 4.**

Results from machine learning classifier development and evaluation on the salivary gland tumor cohort extended by additional differential diagnoses. (A) Pooled confusion matrix summarizing the results from the 10 times repeated 10-fold nested cross-validation. Columns represent the annotated classes, whereas rows correspond to the classes predicted by the classifier. The numerical values indicate the proportion of the predicted class within the annotated class. (B) Connected dot plot displaying the balanced accuracy across 10 repetitions of the 10-fold nested cross-validation, with and without rejection of cases with a calibrated score <0.9. (C) Connected dot plot illustrating the sensitivity and specificity of rejection across the 10 repetitions.

## Discussion

In this study, we—to the best of our knowledge—for the first time defined the DNA methylation landscape of salivary gland tumor entities. We demonstrate that epigenetic profiling recapitulates in the majority of entities by the conventional histomorphologic classification according to the WHO and present a machine learning algorithm that can be used for salivary gland tumor classification. However, we also observed several important discrepancies, which might contribute to currently ongoing controversial discussion in the head and neck pathologists' community.

First, we observed that cribriform adenocarcinoma harbors a stand-alone DNA methylation profile, clearly distinct from that of polymorphous adenocarcinoma. In the current WHO classification, cribriform adenocarcinoma is classified as a subtype of polymorphous adenocarcinoma because of substantial histomorphologic overlap. However, there is an ongoing dispute about whether cribriform adenocarcinoma might represent a distinct tumor entity.<sup>27</sup> This view is supported by the observation that cribriform adenocarcinoma shows distinct morphologic features, has a strong predilection for the base of the tongue, shows a high rate of *PRKD* gene family fusions (instead of mutations in polymorphous adenocarcinoma), and is more likely associated with lymph node metastases.<sup>28–30</sup> However, the inability of these features to clearly differentiate the 2 entities, along with the existence of borderline cases, has led to the current definition in the WHO classification. In our opinion, our finding adds novel and substantial evidence that cribriform adenocarcinoma might represent a distinct tumor entity. Regardless of the classification as a separate entity or a subtype, DNA methylation might be a novel approach to distinguish cribriform from polymorphous adenocarcinoma and help identify patients with a higher risk of developing lymph node metastases.

Second, for pleomorphic adenoma and myoepithelioma, 2 related but overlapping groups were generated in t-SNE analysis. Thus, both defined entities could not be separated by differences in their DNA methylation profiles. Interestingly, pleomorphic adenoma is defined as a biphasic tumor consisting of ductal and myoepithelial cells with a highly variable potency for epithelial-mesenchymal transdifferentiation. Myoepithelioma is primarily composed of myoepithelial cells and contains only few ductal structures with a suggested cutoff of <5% of the tumor area.<sup>31</sup> However, this cutoff appears arbitrary, and in clinical practice, distinguishing these 2 entities can be somewhat subjective. It has repeatedly been postulated that pleomorphic adenoma and myoepithelioma represent a pure histomorphologic continuum, ranging from classical pleomorphic adenoma over myoepithelial-rich pleomorphic adenoma to myoepithelioma.<sup>32</sup> This is supported by the fact that *PLAG1* gene fusions occur in a subset of both tumors. Several authors suggest that pleomorphic adenoma and myoepithelioma should be regarded as a single tumor entity with a broad and continuous morphologic spectrum, which is additionally supported by our result.

Finally, we identified a so far unclear, yet unique epigenetic class of 10 cases that, by conventional means, were initially diagnosed as epithelial-myoeplithelial, adenoid cystic, or myoepithelial carcinoma. Additional in-depth molecular analyses are required to assess if this epigenetic class actually represents a new distinct tumor entity with dominant myoepithelial differentiation. Because our aims for this study were to describe the general epigenetic landscape of salivary gland tumors and assess the general feasibility of DNA methylation-based classification, the

detailed histologic and molecular characterization of this “mixed group” is out of the scope of this study and will be addressed in future work.

A minor limitation of our study is that a few tumor entities (eg, lymphadenoma and lymphoepithelial carcinoma) could not be analyzed because of a major reactive (lymphocytic) cell component and that some very rare entities could yet not be included because of the lack of sufficient case numbers. Following the implementation and broader use of our classifier, we aim to include these tumor entities in a future updated version.

Interestingly, there was no clear differentiation between mucoepidermoid carcinoma and hyalinizing clear cell carcinoma. It is well known that mucoepidermoid carcinoma (especially clear cell type) and hyalinizing clear cell carcinoma can show considerable morphologic overlap. However, both entities are characterized by distinct and highly recurrent fusion events, namely *CRTC1/3::MAML2* in mucoepidermoid carcinomas and *EWSR1::ATF1* (rarely *EWSR1::CREM*) in hyalinizing clear cell carcinomas.<sup>33,34</sup> The morphologic similarity of the 2 entities and their very similar DNA methylation profiles could, for unclear reasons, indicate shared or closely related cells of origin. However, with additional cases tested in the future, the classification algorithm may learn minor epigenetic differences more effectively, potentially enabling a clearer separation.

To facilitate the clinical application of DNA methylation-based diagnosis of salivary gland tumors, we trained and validated a machine learning-based classifier that predicts salivary gland tumor entities with confidence scores. Additionally, squamous cell and clear cell renal cell carcinomas can also be predicted by the classifier, covering the most relevant non-salivary gland entities in a clinical diagnostic setting. Overall, the classifier achieved an excellent mean balanced accuracy of 0.991 for samples with high confidence on our data, underscoring the feasibility of incorporating this classifier into a controlled experimental diagnostic setting.

In conclusion, our study provides a comprehensive resource of DNA methylation data for salivary gland tumors. Our data suggest that cribriform adenocarcinoma may represent a distinct tumor entity, potentially separate from polymorphous adenocarcinoma. Conversely, pleomorphic adenoma and myoepithelioma could be regarded as a common entity with a broad differentiation capacity. Similar to earlier research in other tumor types, such as brain tumors and sarcomas, we further show that DNA methylation could identify potentially novel tumor entities; the existence of an additional “mixed tumor group” has to be verified by further in-depth molecular studies. Finally, we demonstrate that DNA methylation-based classification of salivary gland tumors is a promising tool that could, in the future, complement diagnostics and help solve challenging cases that are in the moment resolved insufficiently by conventional histomorphology, immunohistochemistry, and molecular profiling.

### Acknowledgments

The authors gratefully acknowledge the excellent technical assistance of Peggy Wolkenstein, Ines Koch, Daniel Teichmann, and Carola Geiler. The results are in part based on data generated by The Cancer Genome Atlas Research Network: <https://www.cancer.gov/tcga>.

### Author Contributions

P.J., S.I., and D.C. conceptualized this study. P.J., S.I., and M.L. wrote the original draft of the manuscript. A.v.D., F.K., and D.C.

provided the resources. P.J., S.I., M.L., J.K., and A.J. performed formal analysis. P.J. and M.L. developed the software. P.J., A. Arnold, E.P.C., L.H.M., Y.Z., K.S., N.J., E.C.C., A.J., J.K., P.H., N.P., S.E., L.B., N.D., S. Frank, J.H., S. Försch, G.B., I.v.E.v.G., G.L., C.v.H., M.S., F.Z., A. Agaimy, and S.I. performed data curation. All authors reviewed and edited the manuscript. P.J. and M.L. generated visualization. P.J. and S.I. administered the project. P.J. and A.v.D. acquired funding.

#### Data Availability

Raw DNA methylation data measured in the course of this study were deposited in Gene Expression Omnibus (GEO) under the accession GSE243075. All other raw DNA methylation data used in this study are available from the previously published GEO repositories GSE105288, GSE61441, GSE92482, GSE171994, GSE140686, and GSE67097 and the Cancer Genome Atlas.

#### Funding

This study was supported by an Illumina research grant to A.v.D. P.J. was participant in the Berlin Institute of Health—Charité Digital Clinician Scientist Program funded by the Charité—Universitätsmedizin Berlin, the Berlin Institute of Health, and the German Research Foundation (DFG) and is further supported by the Medical & Clinician Scientist Program of the Ludwig-Maximilians-Universität München.

#### Declaration of Competing Interest

D.C. and A.v.D. are listed as inventors on the patent application “DNA methylation—based method for classifying tumor species” (PCT/EP2016/055337) filed by Deutsches Krebsforschungszentrum Stiftung des öffentlichen Rechts and Ruprecht-Karls-Universität Heidelberg. All other authors report no relevant conflicts of interest.

#### Ethics Approval and Consent to Participate

This study was approved by ethics committee of the Ludwig-Maximilians-Universität München (23-0832) and performed in accordance with the Declaration of Helsinki.

#### Supplementary Material

The online version contains supplementary material available at <https://doi.org/10.1016/j.modpat.2024.100625>.

#### References

1. WHO Classification of Tumours Editorial Board. *WHO Classification of Tumours. Head and Neck Tumours*. [Internet; beta version ahead of print]. International Agency for Research on Cancer; 2022.
2. Skálová A, Hycza MD, Leivo I. Update from the 5th edition of the World Health Organization classification of head and neck tumors: salivary glands. *Head Neck Pathol*. 2022;16(1):40–53.
3. Xu B, Mneimneh W, Torrence DE, et al. Misinterpreted myoepithelial carcinoma of salivary gland: a challenging and potentially significant pitfall. *Am J Surg Pathol*. 2019;43(5):601–609.
4. Ihrler S, Agaimy A, Guntinas-Lichius O, et al. Why is the histomorphological diagnosis of tumours of minor salivary glands much more difficult? *Histopathology*. 2021;79(5):779–790.
5. Skálová A, Hycza MD, Vaneček T, Baněčková M, Leivo I. Fusion-positive salivary gland carcinomas. *Genes Chromosomes Cancer*. 2022;61(5):228–243.
6. Skálová A, Vaneček T, Sima R, et al. Mammary analogue secretory carcinoma of salivary glands, containing the ETV6-NTRK3 fusion gene: a hitherto undescribed salivary gland tumor entity. *Am J Surg Pathol*. 2010;34(5):599–608.

7. Fujii K, Murase T, Beppu S, et al. MYB, MYBL1, MYBL2 and NFIB gene alterations and MYC overexpression in salivary gland adenoid cystic carcinoma. *Histopathology*. 2017;71(5):823–834.
8. Capper D, Jones DTW, Sill M, et al. DNA methylation-based classification of central nervous system tumours. *Nature*. 2018;555(7697):469–474.
9. Jurmeister P, Glöf S, Roller R, et al. DNA methylation-based classification of sinonasal tumors. *Nat Commun*. 2022;13(1):7148.
10. Jurmeister P, Bockmayr M, Seegerer P, et al. Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. *Sci Transl Med*. 2019;11(509):eaaw8513.
11. Koelsche C, Schrimpf D, Stichel D, et al. Sarcoma classification by DNA methylation profiling. *Nat Commun*. 2021;12(1):498.
12. Løkk K, Modhukur V, Rajashekar B, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol*. 2014;15(4):3248.
13. Nam HY, Chandrashekar DS, Kundu A, et al. Integrative epigenetic and gene expression analysis of renal tumor progression to metastasis. *Mol Cancer Res*. 2019;17(1):84–96.
14. Wei JH, Haddad A, Wu KJ, et al. A CpG-methylation-based assay to predict survival in clear cell renal cell carcinoma. *Nat Commun*. 2015;6(1):8699.
15. Leitheiser M, Capper D, Seegerer P, et al. Machine learning models predict the primary sites of head and neck squamous cell carcinoma metastases based on DNA methylation. *J Pathol*. 2022;256(4):378–387.
16. Vandiver AR, Irizarry RA, Hansen KD, et al. Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. *Genome Biol*. 2015;16(1):80.
17. Hovestadt V, Zapatka M, Conuee: Enhanced copy-number variation analysis using Illumina DNA methylation arrays. R package version 1.9.0. Accessed October 11, 2024. <http://bioconductor.org/packages/conuee/>
18. Jurmeister P, Wrede N, Hoffmann I, et al. Mucosal melanomas of different anatomic sites share a common global DNA methylation profile with cutaneous melanoma but show location-dependent patterns of genetic and epigenetic alterations. *J Pathol*. 2022;256(1):61–70.
19. Nilsen G, Liestøl K, Loo PV, et al. Copynumber: efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics*. 2012;13(1):591.
20. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–1369.
21. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14(1):293.
22. Peters TJ, Buckley MJ, Chen Y, Smyth GK, Goodnow CC, Clark SJ. Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate. *Nucleic Acids Res*. 2021;49(19):e109.
23. Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*. 2015;8(1):6.
24. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res*. 2017;45(4):e22.
25. Maros ME, Capper D, Jones DTW, et al. Machine learning workflows to estimate class probabilities for precision cancer diagnostics on DNA methylation microarray data. *Nat Protoc*. 2020;15(2):479–512.
26. Tay JK, Narasimhan B, Hastie T. Elastic net regularization paths for all generalized linear models. *J Stat Softw*. 2023;106(1):1–31.
27. Xu B, Barbieri AL, Bishop JA, et al. Histologic classification and molecular signature of polymorphous adenocarcinoma (PAC) and cribriform adenocarcinoma of salivary gland (CASG). *Am J Surg Pathol*. 2020;44(4):545–552.
28. Skálová A, Sima R, Kaspírková-Nemcová J, et al. Cribriform adenocarcinoma of minor salivary gland origin principally affecting the tongue: characterization of new entity. *Am J Surg Pathol*. 2011;35(8):1168–1176.
29. Michael M, Skálová A, Simpson RH, et al. Cribriform adenocarcinoma of the tongue: a hitherto unrecognized type of adenocarcinoma characteristically occurring in the tongue. *Histopathology*. 1999;35(6):495–501.
30. Skálová A, Gnepp DR, Lewis JS, et al. Newly described entities in salivary gland pathology. *Am J Surg Pathol*. 2017;41(8):e33–e47.
31. WHO Classification of Tumours Editorial Board. *Head and Neck Tumours* [Internet; beta version ahead of print] International Agency for Research on Cancer; 2022 [cited 2024 March 3]. (*WHO Classification of Tumours series, 5th ed.*; vol. 9). <https://tumourclassification.iarc.who.int/chapters/52>
32. Pustaszzeri M, Braunschweig R, Mihaescu A. Pleomorphic adenoma with predominant plasmocytoid myoepithelial cells: a diagnostic pitfall in aspiration cytology. Case report and review of the literature. *Diagn Cytopathol*. 2009;37(1):56–60.
33. Antonescu CR, Katabi N, Zhang L, et al. EWSR1-ATF1 fusion is a novel and consistent finding in hyalinizing clear-cell carcinoma of salivary gland. *Genes Chromosomes Cancer*. 2011;50(7):559–570.
34. Jee KJ, Persson M, Heikinheimo K, et al. Genomic profiles and CRT1–MAML2 fusion distinguish different subtypes of mucoepidermoid carcinoma. *Mod Pathol*. 2013;26(2):213–222.