A Poor Prognosis Subtype of HNSCC Is Consistently Observed across Methylome, Transcriptome, and miRNome Analysis

Alain C. Jung, Sylvie Job, Sonia Ledrappier, Christine Macabre, Joseph Abecassis, Aurélien de Reyniès, and Bohdan Wasylyk

Abstract

Purpose: Distant metastasis after treatment is observed in about 20% of squamous cell carcinoma of the head and neck (HNSCC). In the absence of any validated robust biomarker, patients at higher risk for metastasis cannot be provided with tailored therapy. To identify prognostic HNSCC molecular subgroups and potential biomarkers, we have conducted genome-wide integrated analysis of four omic sets of data.

Experimental Design: Using state-of-the-art technologies, a core set of 45 metastasizing and 55 nonmetastasizing human papillomavirus (HPV)-unrelated HNSCC patient samples were analyzed at four different levels: gene expression (transcriptome), DNA methylation (methylome), DNA copy number (genome), and microRNA (miRNA) expression (miRNome). Molecular subgroups were identified by a model-based clustering analysis. Their clinical relevance was evaluated by survival analysis, and functional significance by pathway enrichment analysis.

Results: Patient subgroups selected by transcriptome, methylome, or miRNome integrated analysis are associated with shorter metastasis-free survival (MFS). A common subgroup, R1, selected by all three omic approaches, is statistically more significantly associated with MFS than any of the single omic-selected subgroups. R1 and non-R1 samples display similar DNA copy number landscapes, but more frequent chromosomal aberrations are observed in the R1 cluster (especially loss at 13q14.2-3). R1 tumors are characterized by alterations of pathways involved in cell–cell adhesion, extracellular matrix (ECM), epithelial-to-mesenchymal transition (EMT), immune response, and apoptosis.

Conclusions: Integration of data across several omic profiles leads to better selection of patients at higher risk, identification of relevant molecular pathways of metastasis, and potential to discover biomarkers and drug targets. Clin Cancer Res; 19(15); 4174–84. ©2013 AACR.

Introduction

Squamous cell carcinoma of the head and neck (HNSCC) is the sixth most frequent malignancy in the world, with more than 600,000 new cases diagnosed each year. HNSCC has various etiologic factors [tobacco smoking with alcohol consumption, human papillomavirus (HPV) infection, etc.], heterogeneous pathologic and clinical features, and diverse outcome (1). HNSCC is managed with surgery and/or chemoradiotherapy. Locally advanced tumors can be treated with neo-adjuvant chemotherapy to avoid the development of distant metastasis (2). However, these treatments are associated with high-grade toxicities. Despite recent improvements in treatment, HNSCC prognosis remains poor, with less than 50% of patients remaining alive after 5 years (2). In the absence of suitable biomarkers, therapeutic decisions are solely based on tumor localization (oral cavity, hypopharynx, and oropharynx) and staging, which includes the evaluation of tumor size (T), lymph node involvement (N), and the presence of distant metastases (M). However, lesions with similar pathologic features can differ in clinical outcome. Prognostic biomarkers are
required for the stratification of patients with high metastasis risk and choice of therapy. In this study, we used a cohort of 100 patients with HPV-unrelated HNSCC with similar treatment, but with different outcome for the development of distant metastasis within 3 years after treatment. We profiled patient samples with four different genome-wide profiling technologies as a comprehensive approach to identifying clinically useful markers.

Single biomarkers are usually insufficient to predict therapeutic response and combinatorial markers are being developed through high-resolution profiling at the "omic" level (3, 4). For example, molecular profiling at the transcriptome level has identified five subtypes of breast cancer associated with different prognosis and treatment response (5). Newer developments in high-throughput technologies have led to the analysis of tumors at different omic levels, that include DNA copy number (genome), DNA methylation (methylome), and microRNA (miRNA) expression (miRNome). There are many examples in the literature of genome-wide profiling efforts of HNSCC, at the individual levels of the transcriptome (6), DNA copy number (7), miRNA expression (8, 9), DNA methylation (10), and whole genome sequencing (refs. 11, 12; see Supplementary Table S1 for a more extensive compilation). Gene expression signatures have been reported that correlate with adverse prognosis and with higher risk for distant metastasis (13–15). However, these signatures have little in common and need further functional and clinical validation. Studies of DNA methylation have identified deregulated genes (16, 17), and analyses of the expression of particular miRNAs have uncovered single miRNAs whose up- or downregulation correlates with prognosis (18, 19). To extend the analysis of HNSCC at the omic level, we have compared whole genome analysis of the same core 100 HNSCC patient samples at four different levels: gene expression (transcriptome), DNA methylation (methylome), DNA copy number (genome), and miRNA expression (miRNome). Comparing the results of several omic profiles should help to distinguish robust molecular subgroups that are identified by different profiling techniques, and could single out profiling techniques that give more reliable predictions. Furthermore, combining analyses at different omic levels of the same samples could potentially identify important pathways that are deregulated by different mechanisms and that cannot be identified by any single omic analysis (20, 21).

We found that unsupervised clustering of the methylome, transcriptome, and miRNome profiles leads to the identification of a common robust subgroup of patients with high propensity for distant metastasis, whereas DNA copy number analysis is less effective. Pathway analysis highlighted the importance of cell adhesion and epithelial-to-mesenchymal transition (EMT) in distinguishing the common poor prognosis group at the functional level. Deregulated components of these pathways are potentially interesting candidates for both biomarker and mechanistic studies.

Materials and Methods

Patient characteristics and tumor samples

Hundred samples were profiled, including 89 for transcriptome, 84 for methylome, 64 for miRNome, and 88 for genome analyses. Their clinical, phenotypic, and molecular characteristics are summarized in Table 1 and detailed in the Supplementary Materials and Methods and Supplementary Table S2. Of note, 45 of 100 patients developed distant metastasis, mainly to lung, bone, and liver, as the first recurrence event within 3 years after treatment (M) and 55 of 100 patients followed during at least 3 years did not develop metastasis (NM). The median follow-up period for the cohort was 61.5 months (7–189 months). One hundred and one additional samples were used for methylome validation (Supplementary Materials and Methods and Supplementary Table S3). The samples did not display detectable HPV DNA and E6/E7 mRNA. HPV-related tumors were excluded. The Unio Internationale Contra Cancrum (UIICC) TNM system was used for staging (22).

Datasets

Methylation. Illumina Human Methylation 27K array data were obtained for 84 samples and 450K array data for 101 samples (Supplementary Tables S2 and S3; Integrum Gen). Common probes were used for analysis.

Transcriptome. HG-U133-plus-2.0 Affymetrix array data were obtained for 89 samples (14) and a public Affymetrix expression dataset was also used [Gene Expression Omnibus (GEO) website, GSE10300; ref. 23].

miRNA. miRNA data were obtained for 64 samples by the Illumina sequencing process (ISeq2000 [IntegraGen]).

Genomic. Eighty-eight were analyzed with Illumina HumanCNV370-Quad v3.0 chips, containing 373,397 probes (IntegraGen).
Survival curves were calculated according to the Kaplan–Meier method and differences between curves were assessed using the log-rank test.

Survival analysis.

Table 1. Characteristics of patients for the four "omic" analyses

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>88</td>
</tr>
<tr>
<td>Female</td>
<td>12</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>&gt;45 y</td>
<td>83</td>
</tr>
<tr>
<td>≤45 y</td>
<td>17</td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
</tr>
<tr>
<td>Hypopharynx</td>
<td>37</td>
</tr>
<tr>
<td>Oropharynx</td>
<td>49</td>
</tr>
<tr>
<td>Gums, floor of the mouth, oral cavity</td>
<td>11</td>
</tr>
<tr>
<td>Other part of the tongue</td>
<td>3</td>
</tr>
<tr>
<td>TNM N</td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>20</td>
</tr>
<tr>
<td>N1</td>
<td>18</td>
</tr>
<tr>
<td>N2</td>
<td>53</td>
</tr>
<tr>
<td>N3</td>
<td>9</td>
</tr>
<tr>
<td>TNM stage</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>9</td>
</tr>
<tr>
<td>III</td>
<td>27</td>
</tr>
<tr>
<td>IV</td>
<td>64</td>
</tr>
<tr>
<td>Cell differentiation</td>
<td></td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>1</td>
</tr>
<tr>
<td>Low</td>
<td>32</td>
</tr>
<tr>
<td>Medium</td>
<td>45</td>
</tr>
<tr>
<td>High</td>
<td>22</td>
</tr>
</tbody>
</table>

Details about the datasets, data preprocessing, and the overlap between samples used for each of the omics can be found in the Supplementary Material and Methods and Supplementary Fig. S1.

Omic analysis

Unsupervised analysis. Three methods were used to find unsupervised clusters within each omics: the recursively partitioned mixture model (RPMM) method (24), and two consensus clustering methods (25, 26). There is a strong association between the three methods for the four omics (Fisher exact P values from 2.10E–04 to 2.49E–19; see Supplementary Material and Methods). Only the results obtained with the RPMM method are described in the article.

Differential analysis. Several differential analysis methods were used: moderate t test for transcriptome and genome data, Wilcoxon test for miRNome and methylome data, and likelihood ratio tests based on a negative binomial model (R package edgeR) for miRNome data.

Signaling pathway analysis. Four methods were used to compare sample groups with 17,306 gene sets collected from Kyoto Encyclopedia of Genes and Genomes (KEGG), Molecular Signatures Database (MSigDB), Stanford Microarray Database (SMD), gene ontology (GO), or BioCarta [gene-set analysis (GSA), Global test, significance analysis of microarray extended to gene-set analysis (SAM-GS), and a Tuckey approach].

Survival analysis. Survival curves were calculated according to the Kaplan–Meier method and differences between curves were assessed using the log-rank test.

Classifier building. A centroid-based predictor was built on our series for the methylation and expression data, using genes differentially methylated (respectively expressed) between the studied subgroups (Wilcoxon test, P < 1E–3 for the methylation data; moderate t test, P < 1E–5 for the transcriptome data). The distance metric was adapted to each data type (Manhattan distance for methylation data and dqda distance for expression data).

See the Supplementary Materials and Methods for more details about the omics analysis and the validation of omic data by alternative molecular biology technologies.

Accession number E-MTAB-1328.

Results

A multi-omics unsupervised analysis identifies a common group

To identify molecular subgroups in HNSCC, we analyzed 100 HPV-negative tumor samples from distinct patients (see Table 1 for patient demographics) using four different high-content "omic" analyses: DNA methylation (methylome), mRNA expression (transcriptome), miRNA expression (miRNome), and DNA copy number (genome; see Materials and Methods).

Unsupervised clustering analyses were conducted on each omic dataset. The RPMM algorithm was used because it is suitable for both β-distributions (methylation data) and Gaussian distributions (other data), and it gives an optimal number of clusters using a recursive-partitioning algorithm (24). The clustering gave three groups for the methylome (Me.1-.3), the transcriptome (T.1-.3), and the miRNome (Mi.1-.3) and four groups for the genome (G.1-.4; Fig. 1). Heatmaps were generated to visualize molecular patterns specific for each cluster of samples. Distinct patterns are observed for each sample cluster at the methylome, transcriptome, and miRNome levels, but not at the genome level.

To determine whether the clusters have common tumor samples, we created contingency tables, with the methylome cluster partition as the reference (Fig. 2). Similar results were obtained when the transcriptome or miRNome clusters were used as the reference (data not shown). There is a strong association between the patient samples in the clusters for the methylome, transcriptome, and miRNome, and especially between Me.1, T.1, and Mi.1+.2. Statistically significant associations were found when comparing the methylome to the transcriptome (Fisher exact test; P = 5.76E–10), the methylome to the miRNome (Fisher exact test; P = 9.68E–6), and the transcriptome to the miRNome (Fisher exact test; P = 7.08E–9). There was a close association between Me.1 and T.1 (12 of 15 samples), Me.1 and Mi.1+.2 (12 of 17), and also Mi.1+.2 and T.1 (13 of 18). The associations between the other groups were weaker. In summary, unsupervised analysis identifies a group of patients, R1, which are coclustered in three of four omics approaches (Me.1, T.1, and Mi.1+.2).

Various controls were analyzed to ensure that these results were not driven by random noise or batch effects.
The different omic analyses were conducted on the same core set of tumors; however, some samples could not be analyzed with all four omic platforms due to progressive exhaustion of biologic material (see Supplementary Fig. S1 for a Venn diagram). To rule out the possibility that missing data could have biased the results, we repeated the analysis using the 60 samples with complete datasets. Similar results were obtained, with a new R1 subgroup \( (n = 10) \) that included 10 of 11 original R1 samples.

We compared the current study with our previous transcriptome analysis of an overlapping set of 97 samples \((14)\) that identified 4 nonsupervised clusters \((C1-4)\) and a 4-gene predictor. There is a good correlation between the previous clusters \((C1-4)\) and both the new clusters \((T.1-3; \text{Fisher exact test}; P = 1.98E−13)\) and the R1 subgroup \((\text{Fisher exact test}; P = 2.49E−08; \text{see Supplementary Table S3})\). C3 overlaps significantly with T.1 as well as the subgroup R1. The 4-gene predictor correctly predicts 7 of 8 R1 samples in the 82 common samples that were analyzed by transcriptomics \((\text{Fisher exact test}; P = 1.82E−02)\). This comparison shows that the current analysis agrees with our previous transcriptomic study.

The Me.1, T.1, and the Mi.1+2 clusters and their common subgroup R1 have poor metastasis-free survivals

We investigated whether the closely associated clusters \((\text{Me.1, T.1, and Mi.1+2})\) and the common subgroup R1 define a clinically relevant patient population. The Kaplan–Meier curves for metastasis-free survival \((\text{MFS})\) of these clusters versus the other patients were plotted (Fig. 3).

Figure 1. Dendograms and heatmaps of the unsupervised clusters for the methylome, transcriptome miRNome, and genome, restricted to genes having an ANOVA \( P \leq 5E−2\). Blue, low level; red, high level; white, mean level. The scale for each dataset is shown.
of the three closely associated clusters, namely Me.1, T.1, and Mi.1+2, was found to have poorer prognosis with significant log-rank $P$ values (Me.1 vs. Me.2+3, $P = 7.97E^{-03}$; T.1 vs. T.2+3, $P = 7.24E^{-03}$; Mi.1+2 vs. Mi.3+4, $P = 8.49E^{-04}$; Fig. 3A–C). The common subgroup R1, composed of 11 patients, was also found to have a poorer prognosis, with the strongest prognostic discrimination (log-rank test; $P = 5.37E^{-04}$; Fig. 3D). The three closely associated clusters and R1 also correlated with poorer overall survival (OS), disease-specific survival (DSS), and disease-free survival (DFS; Supplementary Table S4). Loco-regional–free survival could not be reliably analyzed, because of the small number of patients with this relapse in the study group.

Validation of the link between R1 samples and metastatic relapse on independent datasets

The association of R1 subgroup molecular characteristics to poorer prognosis was tested on two independent datasets.
Table 2. Multivariate Cox regression analysis of the influence of R1 status and of the main clinical variables on the MFS, OS, DFS, and DSS.

<table>
<thead>
<tr>
<th>Variable</th>
<th>MFS HR (95% CI)</th>
<th>OS HR (95% CI)</th>
<th>DFS HR (95% CI)</th>
<th>DSS HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age: &lt;58 y vs. ≥58 y</td>
<td>2.19 (1.16–4.15)</td>
<td>1.61E-02</td>
<td>1.10 (0.66–1.81)</td>
<td>NS</td>
</tr>
<tr>
<td>Gender: male vs. female</td>
<td>0.51 (0.25–1.04)</td>
<td>0.31E-02</td>
<td>0.52 (0.29–0.95)</td>
<td>NS</td>
</tr>
<tr>
<td>Tumor localization: hypoph vs. others</td>
<td>0.57 (0.38–0.86)</td>
<td>8.58E-04</td>
<td>0.52 (0.29–0.95)</td>
<td>NS</td>
</tr>
<tr>
<td>TNM stage: I vs. IV</td>
<td>2.82 (1.63–4.93)</td>
<td>1.30E-02</td>
<td>5.45 (1.54–19.37)</td>
<td>3.31E-07</td>
</tr>
<tr>
<td>Treatment: chemorad vs. rad</td>
<td>1.31E-05</td>
<td>8.58E-04</td>
<td>0.52 (0.29–0.95)</td>
<td>NS</td>
</tr>
<tr>
<td>Comorbidities: none vs. cardiovascular</td>
<td>0.74 (0.42–1.34)</td>
<td>0.68E-02</td>
<td>0.75 (0.42–1.34)</td>
<td>NS</td>
</tr>
<tr>
<td>R1 status: R1 vs. non-R1</td>
<td>5.13 (1.86–12.68)</td>
<td>3.91E-04</td>
<td>5.13 (1.86–12.68)</td>
<td>3.91E-04</td>
</tr>
</tbody>
</table>

Abbreviations: cardio-vascular, cardio-vascular; chemorad vs. rad, postoperative chemoradiotherapy vs. postoperative radiotherapy; hypoph, hypopharynx; tumor diff, tumor differentiation; well/mod vs. poor/no, well/moderately differentiated vs. poorly/non.

Clinical characterization of the R1 subgroup

To further characterize the poor prognosis subgroup R1, we investigated its association with different clinical criteria (Supplementary Table S6) and its prognostic value in a multivariate analysis of survival (Table 2). The R1 subgroup is associated with tumor differentiation (Fisher exact test; \( P = 8.58E-3 \)) as well as metastasis (Fisher exact test; \( P = 1.14E-2 \); Supplementary Table S6). Importantly, in a multivariate Cox regression analysis, the R1 subgroup remains an independent prognostic factor for MFS when additional potential confounding factors are considered: patient age, gender, cardio-vascular comorbidities, treatment, lymph node involvement (N), and tumor size, stage, localization, and differentiation (Wald test; \( P = 3.91E-04 \)). Similar results were obtained with OS, DFS, and DSS (Table 2). These results indicate that the R1 subgroup has clinical relevance.

Functional characterization of the R1 subgroup

R1 was studied at the functional level using four different methods of pathway enrichment analysis for the methylome, transcriptome, and miRNome data (see Materials and Methods). For each dataset, the results were aggregated, ranked, and categorized into three levels (+++; +/++; +/+/+; Supplementary Table S7). There is good agreement between the three datasets. Cell adhesion, extracellular matrix (ECM), and EMT are the most highly represented, followed by immunity, and then apoptosis, the Wnt pathway, cell motility, mitosis, and finally the cytoskeleton. The
order and type of pathways that are enriched agree with established knowledge of the importance of these pathways for the metastatic process. Heatmaps for individual genes in the top pathways illustrate differences in expression between the R1 and non-R1 groups (Fig. 4). ECM is not shown because of the redundancy with cell adhesion. ECM is related to cell adhesion, and, as might be expected, 19 of 44 genes (moderate t test; \( P < 1E^{-05} \)) are common between the two terms. Some notable changes in R1 are: downregulation of desmosome components that are involved in cell adhesion and EMT (DSG3, DSC3, DCS2, and JUP), overexpression of vimentin (moderate t test; \( P = 0.048 \)), and downregulation of keratins and E-cadherin involved in EMT, downregulation of BIK, and upregulation of PRAME that are respectively pro- and anti-apoptotic factors, and overexpression of MYB and decreased expression of Ly96 and TICAM1 that are involved in immunity. There are also notable differences in the expression of Myc family members in R1, including overexpression of MYCN (moderate t test; \( P = 6.87E^{-12} \) and fold change = 1.90), downregulation of MXD1 (moderate t test; \( P = 6.79E^{-8} \) and fold change = 2.69), and downregulation of MYC (moderate t test; \( P = 1.97E^{-04} \) and fold change = 1.34).

**Potential interactions between genes, miRNAs, and CpG dinucleotides that are significantly associated with the R1 subtype**

The good concordance between each omic of the pathways that differ between R1 and non-R1 (see above),
suggests that there are global functional connections between "upstream regulators" (methylobe and miRNome) and "downstream targets" (transcriptome). At the individual gene level, 17.9% of the upregulated genes (moderate t test; q ≤ 0.05 and fold change > 1) are hypomethylated (Wilcoxon test; q ≤ 0.05 and fold change > 0.1). This includes genes in the apoptosis pathway (GAS2, EDAR). Conversely, 7.9% of the downregulated genes are hypermethylated. Two examples are the KRT16 or KRT6B genes that are involved in the maintenance of epithelial characteristics (EMT pathway).

At the miRNA level, 291 connections were found between miRNAs that are differentially regulated between R1 and non-R1 and predicted targets that are anticorrelated. Examples include desmosome components (DSC2, DSC3, and DSG3) in the adhesion pathway. These results show that there are potential interactions between genes, miRNAs, and CpG dinucleotides associated with the R1 subgroup.

Validation of high-throughput measures with alternative methodologies

To validate the high-throughput analysis, we used alternative methodologies for selected examples (see Supplementary Materials and Methods and Supplementary Figs. S4–S6). As expected from the omic analyses, AIM2 (absent in melanoma 2), KRT16 (keratin 16), DSG3 (desmoglein 3), and miRNA (miR)-1 were found to be downregulated in R1 samples, and SFRP1 (secreted frizzled-related protein 1), Col9A3 (collagen type IX α 3), and MiR-345 were upregulated (Supplementary Fig. S4). As expected from the methylation scores (Supplementary Fig. S5A), GRP55 (G protein–coupled receptor 55; cg20287234 probe region) was found to be more highly methylated in the R1 samples (Supplementary Figs. S5A and S6) and IHH (Indian hedgehog; cg25908985 probe region) in non-R1 (Supplementary Fig. S5C). In conclusion, there is a very good correlation between the high-throughput and the alternative approaches.

The R1 subgroup has higher levels of genomic aberrations, especially at 13q14.2–3

Unsupervised clustering of the genomic data, based on patterns of genomic amplification or deletion, did not identify the R1 subgroup (Figs. 1 and 2). Likewise, inspection of pangenomic plots, which illustrate the distribution of gains and losses across the genome in the R1 and non-R1 samples, indicates that the overall patterns (landscapes) are similar in the two groups (Supplementary Fig. S7). We studied additional genomic features of R1, in an attempt to identify alterations that are characteristic of this clinically significant subgroup. We found that the R1 samples have a higher overall aberration rate relative to the other samples (moderate t test; P = 5.1E–06), and there are more significant losses than gains (see P values for the green troughs in Supplementary Fig. S8). The most significant loss was in the 13q region (see dotted area). The sensitivity and the specificity of individual probes to identify the two groups were determined (Supplementary Table S8). Nine hundred and two probes, corresponding to 26 genes, gave a sensitivity and a specificity greater than 0.75. Of the 26 genes, 22 were found to be localized in the 13q14.2 and 13q14.3 regions. In addition, 10 of these 22 genes were found to be differentially expressed between R1 and the other samples (Supplementary Table S9). Among them, the tumor-suppressor gene RB1 is significantly downregulated in R1 (moderate t test; P = 4.87E–3). These results show that there are higher levels of genomic aberrations in the R1 group, with particularly significant losses in the region 13q14.2–3.

Discussion

We have conducted unsupervised analyses of four omic datasets, generated with a patient cohort with nonmetastatic and metastatic HNSCC tumors. This analysis led to the identification of robust molecular clusters at the methylome (Me.1–Me.3), transcriptome (T.1–T.3), and miRNome (Mi.1–Mi.3) levels, but not at the genome level. Interestingly, from the clinical aspect, patients in the Me.1, T.1, and Mi.1+2 clusters display shorter MFS. Contingency analysis of the three clusters identified a common subgroup of patients, R1, who display poorer prognosis in terms of metastatic relapse, and with a higher statistical significance than Me.1, T.1, or Mi.1+2 alone. R1 was shown to be a robust subgroup, in that it was validated on independent datasets from methylation and gene expression analysis, and different patients from either our collection and from another study (23), respectively. Furthermore, R1 kept its prognostic value in a Cox regression model that took into account tumor differentiation as a potential confounder. At the functional level, R1 tumors are associated with biologic pathway alterations related to the metastatic process, including cell adhesion, ECM, EMT, the immune response, and apoptosis. The genomic data were also informative. Despite the fact that the landscape of chromosomal aberrations does not differ between R1 and non-R1 tumors, the overall rate of aberration was found to be higher in R1 samples. Particularly, noteworthy is the loss of genetic material in the 13q14 region.

Our results indicate that coherent and complementary information can be gained from the use of several different omic analyses on the same samples. Previous omic studies have increased our knowledge, but have generated distinct findings that are difficult to unify and to formulate global transcending understanding. Most omic studies have used gene expression analysis. An early study used 60 HNSCC samples to identify four molecular subtypes, one of which displayed shorter recurrence-free and global survival and relevant molecular changes [overexpression of TGF-α1 and activation of the EGFR receptor (EGFR) pathway (6)]. More recently, analysis of 40 HNSCC samples uncovered a 75-gene signature that predicts adverse clinical outcome and that comprises genes linked to EMT, NF-kB signaling, and cell–cell adhesion (27). Comparison of 11 metastasizing (M) with 8 nonmetastasizing (NM) HNSCC found 150 differential-expressed genes but no evidence for a metastasis signature (13). Our 98-tumor sample study (52 NM vs. 46 M) found 614 genes with significantly altered expression that...
are associated with development, differentiation, adhesion, and motility, and a 4-gene predictor that was able to discriminate NM from M patients with a 77% success rate (14). Recently, a study of 63 pretreatment tumor biopsies generated a gene prognostic signature associated with survival in locally advanced patients that had features of EMT, undifferentiation, and overexpression of genes involved in cell adhesion, NF-kB, and integrin signaling (15). Combined with our current study, an emerging theme seems to be the identification of common altered pathways, including cell adhesion, ECM, EMT, and the immune response, rather than similar single genes in the signatures. This indicates that one key issue in the acquisition of metastatic properties is the way tumor cells interact with and modify their environment.

Analysis of DNA methylation in HNSCC specimens has been mainly hypothesis driven, and has led to the identification of genes whose expression is deregulated at the epigenetic level. Many of these genes were already known to be involved in important functions, including apoptosis, the cell cycle, cell migration, and invasion (for reviews see refs. 28, 29). Interestingly, a recent analysis has shown that a protein involved in cell adhesion (cadherin 11) is specifically methylated in metastases (30).

miRNA expression profiling studies have mainly compared tumor with normal tissue, using either fresh or paraffin-embedded material (18, 31). These studies have identified miRNAs that are differentially expressed (in particular, let-7d, miR-155, miR-205, miR-375, and the miR106b-25 cluster) and associated with increased locoregional recurrence and poorer prognosis. These results could be connected to ours, in that they uncover miRNAs linked to the TGF-β–related EMT [miR106b-25 targets Smad7 (32)] and the DNA damage response [miR106b-25 targets p21 (33)]. In addition, let-7d and miR-155 are significantly differentially expressed between the poorer prognosis R1 and the non-R1 patients (Supplementary Table S10) and low let-7d levels are linked to metastasis and shorter MFS (reviewed in ref. 36). A consequence of EMT is thought to be the acquisition of a cancer stem cell phenotype, with abilities to escape apoptosis, resist therapy, and seed in an unfavorable environment (reviewed in ref. 37). Interestingly, we found a significant enrichment for pathways involved in stem cell maintenance and differentiation, and notably an increase in Tet1 expression in R1 (1.8-fold; $P = 1.99E - 12; \Delta q = 8.06E - 10$). Tet1 is involved in the balance between pluripotency and lineage commitment of cells and it plays a role in embryonic stem cell maintenance and inner cell mass cell specification (38). An important consideration is that these changes are uncovered by cross-analyzing three independent omic datasets, with high statistical significance, which reinforces the importance of these pathways in the acquisition of aggressive tumor features. It also underlines the biologic relevance of these mechanisms as potential drug targets.

The non-R1 group is less ideally defined than the R1 group, as it includes samples that have missing data. Moreover, it includes patients with poor MFS. There are a number of potential reasons for which these "non-R1 M" patients were not recovered in the R1 clusters. They include: (i) metastasis progression in the R1 cluster involves molecular anomalies and aberrations that occur frequently enough to be uncovered with statistical significance by our study. The "non-R1 M" tumors may take other less frequent roads to metastasis progression that might remain beyond the detection potential of our analysis, from a statistical point of view; (ii) the "non-R1 M" tumors may contain a low proportion of aggressive tumor cells, below the level of detection; (iii) the "non-R1 M" tumors may no longer contain cells that can form metastasis, as they have already left the primary tumor for other distant sites (bone marrow for instance). Nevertheless, our approach robustly identifies signaling pathways that are deregulated in patients with poor prognosis. These pathways have been repeatedly shown to relate to metastatic progression in the literature.

The R1 subgroup is associated with changes that reflect decreased cell–cell adhesion, increased EMT, decreased apoptosis, and decreased innate immunity with the altered immune response. These changes are compatible with the more aggressive prometastatic phenotype of the R1 subgroup. The innate immunity response has an important role in solid cancers and is a promising new area to develop cancer therapies (34). Immune cell-type differences in the tumor microenvironment are known to affect progression, through effects on cell migration and intravasation (reviewed in ref. 35). Loss of cell–cell adhesion and increased EMT has been shown to affect cell movement and tumor progression, and to correlate with increased metastasis and shorter MFS (reviewed in ref. 36). A consequence of EMT is thought to be the acquisition of a cancer stem cell phenotype, with abilities to escape apoptosis, resist therapy, and seed in an unfavorable environment (reviewed in ref. 37). Interestingly, we found a significant enrichment for pathways involved in stem cell maintenance and differentiation, and notably an increase in Tet1 expression in R1 (1.8-fold; $P = 1.99E - 12; \Delta q = 8.06E - 10$). Tet1 is involved in the balance between pluripotency and lineage commitment of cells and it plays a role in embryonic stem cell maintenance and inner cell mass cell specification (38). An important consideration is that these changes are uncovered by cross-analyzing three independent omic datasets, with high statistical significance, which reinforces the importance of these pathways in the acquisition of aggressive tumor features. It also underlines the biologic relevance of these mechanisms as potential drug targets.

We have identified particularly robust individual changes between the R1 and non-R1 patients, by validation with alternative techniques. These include downregulation in R1 of AIMP2, KRT16, DSG3, and miR-1 and upregulation of SFRP1, Col9A, and miR-345, and hypermethylation in R1 of GPR35 and hypomethylation of IHH. We have compared these findings with previous studies. AIMP2 was first identified by its absence in melanoma and has tumor-suppressor
functions. For example, restoration of AIM2 induces G1–M cell-cycle arrest and promotes invasion of colorectal cancer cells (39). KRT16 is one of the keratin intermediate filament proteins that are responsible for the structural integrity of epithelial cells. Decreased expression of KRT16 may be associated with loss of the epithelial cell phenotype during tumorigenesis. Interestingly, loss of KRT16 is associated with the transition between melanoma tumor progression and metastasis (40). DSG3 is a component of intercellular desmosome junctions that are involved in cell–cell adhesion. Reduced expression of DSG3 is correlated with poor prognosis (see review in ref. 41). There is considerable evidence for a tumor-suppressor function for desmosomes, but in some circumstances desmosomal proteins have been linked to oncogenic effects (see review in ref. 42). miR-1 is downregulated in various cancer types and regulates genes involved in cancer and metastasis (see review in ref. 43). SFRP1 is a Wnt antagonist. Silencing of SFRP1 leads to Wnt-pathway activation, whereas overexpression can confer metastatic properties (see review in ref. 44). COL9A3 is one of the three α-chains of type IX collagen, the major collagen component of hyaline cartilage that has roles in locomotion, cell adhesion, and ossification. It has been reported to interact with EGFR (45) although whether this may be linked to EGFR overexpression in HNSCC has not been investigated. miR-345 has been found to be upregulated in a meta-analysis of cancers (46), and its expression increases with progression of leukoplakia to oral carcinoma (47). The IHH and GPR55 methylation sites have been shown to have opposite and significant levels of methylation in a study of HNSCC, which is commensurate with our study (48). In conclusion, the directions of gene expression and methylation changes in general align with what would be expected from previous work.

An intriguing finding was that there is little difference between the DNA copy number landscapes of R1 and non-R1 HNSCC samples. A possible interpretation is that the landscapes of alterations somehow reflect the underlying organization of chromosomes (reviewed in ref. 49) rather than tumorigenic potential. However, there are regional chromosome gains and losses that occur more frequently in the R1 samples. One explanation is that the increase in chromosomal aberrations in the R1 tumor specimens could be related to less efficient DNA repair. Alternatively, they may represent accumulations of DNA changes in longer-lived more aggressive cells that have persisted in the R1 tumors. One of the major changes in R1 tumors is loss in the 13q14.2-3 region, which contains the RB1 tumor-suppressor gene.

In conclusion, integrated analysis of multi-omic datasets on an important patient cohort has allowed us to uncover and identify a robust category of head and neck tumors that are at high risk for future distant metastasis and which are significantly deregulated in molecular pathways that play pivotal role in the acquisition of metastatic features (cell–cell adhesion, ECM, EMT, immune response, and apoptosis). These results set the foundations for the identification of HNSCC biomarkers with a functional basis that cross the frontiers between different omic analyses and select important functional changes with clinical relevance.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: A.C. Jung, J. Abecasis, A. de Reyniès, B. Wasylyk

Development of methodology: S. Job, A. de Reyniès, B. Wasylyk

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S. Ledrappier, C. Macabre, B. Wasylyk

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.C. Jung, S. Job, A. de Reyniès, B. Wasylyk

Writing, review, and/or revision of the manuscript: A.C. Jung, S. Job, J. Abecasis, A. de Reyniès, B. Wasylyk

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S. Ledrappier, C. Macabre

Study supervision: A.C. Jung, J. Abecasis, A. de Reyniès, B. Wasylyk

Acknowledgments

This work is part of the Cartes d’Identité des Tumeurs (CIT) program of the Ligue Nationale Contre le Cancer. The authors thank Pr. Jacqueline Godet and Dr. Jaqueline Metral for their invaluable contribution to this project, Dr. M Weber for help with the methylation assays, R Million for participation in the initial phase of the work, and Wasylyk team members for discussions. The authors also thank the participation of the CIT Affymetrix DNA microarray (Philipppe Kastner, Christelle Thibault, and Doulaye Dembele) and the St. Louis Hospital qRT–PCR (Jessica Zurn-Rossi) platforms. Tumor samples were provided by Dr. Guy Brunner.

Grant Support

This study was supported by Ligue Nationale Contre le Cancer CIT Programme, Ligue Regionale CCGE (CCIR-GE), CNRS, INSERM, UDS, and INCA.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 3, 2012; revised May 18, 2013; accepted June 3, 2013; published OnlineFirst June 24, 2013.

References


3. Fojo T, Parkinson DR. Biologically targeted cancer therapy and marginal benefits: are we making too much of too little or are we achieving too little by giving too much? Clin Cancer Res 2010;16:5972–80.


8. Nohata N, Hanazawa T, Kinoshita T, Okamoto Y, Seki N. MicroRNAs function as tumor suppressors or oncogenes: aberrant expression of

www.aacajournals.org

Clin Cancer Res; 19(19) August 1, 2013 4183

Downloaded from clinincancerres.aacajournals.org on April 5, 2017. © 2013 American Association for Cancer Research.