



Article

MicroRNA-Target Network Inference and Local Network Enrichment Analysis Identify Two microRNA Clusters with Distinct Functions in Head and Neck Squamous Cell Carcinoma

Steffen Sass^{1,†}, Adriana Pitea^{1,2,†}, Kristian Unger^{2,3}, Julia Hess^{2,3}, Nikola S. Mueller^{1,*} and Fabian J. Theis^{1,4,*}

Received: 12 November 2015; Accepted: 9 December 2015; Published: 18 December 2015

Academic Editor: Martin Pichler

¹ Institute of Computational Biology, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany; steffen.sass@helmholtz-muenchen.de (S.S.); adriana.pitea@helmholtz-muenchen.de (A.P.)

² Research Unit Radiation Cytogenetics, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany; unger@helmholtz-muenchen.de (K.U.); julia.hess@helmholtz-muenchen.de (J.H.)

³ Clinical Cooperation Group “Personalized Radiotherapy in Head and Neck Cancer”, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

⁴ Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany

* Correspondence: nikola.mueller@helmholtz-muenchen.de (N.S.M.); fabian.theis@helmholtz-muenchen.de (F.J.T.); Tel.: +49-89-3187-1174 (N.S.M.); +49-89-3187-4030 (F.J.T.); Fax: +49-89-3187-3369 (N.S.M. & F.J.T.)

† These authors contributed equally to this work.

Abstract: MicroRNAs represent ~22 nt long endogenous small RNA molecules that have been experimentally shown to regulate gene expression post-transcriptionally. One main interest in miRNA research is the investigation of their functional roles, which can typically be accomplished by identification of mi-/mRNA interactions and functional annotation of target gene sets. We here present a novel method “miRlastic”, which infers miRNA-target interactions using transcriptomic data as well as prior knowledge and performs functional annotation of target genes by exploiting the local structure of the inferred network. For the network inference, we applied linear regression modeling with elastic net regularization on matched microRNA and messenger RNA expression profiling data to perform feature selection on prior knowledge from sequence-based target prediction resources. The novelty of miRlastic inference originates in predicting data-driven intra-transcriptome regulatory relationships through feature selection. With synthetic data, we showed that miRlastic outperformed commonly used methods and was suitable even for low sample sizes. To gain insight into the functional role of miRNAs and to determine joint functional properties of miRNA clusters, we introduced a local enrichment analysis procedure. The principle of this procedure lies in identifying regions of high functional similarity by evaluating the shortest paths between genes in the network. We can finally assign functional roles to the miRNAs by taking their regulatory relationships into account. We thoroughly evaluated miRlastic on a cohort of head and neck cancer (HNSCC) patients provided by The Cancer Genome Atlas. We inferred an mi-/mRNA regulatory network for human papilloma virus (HPV)-associated miRNAs in HNSCC. The resulting network best enriched for experimentally validated miRNA-target interaction, when compared to common methods. Finally, the local enrichment step identified two functional clusters of miRNAs that were predicted to mediate HPV-associated dysregulation in HNSCC. Our novel approach was able to characterize distinct pathway regulations from matched miRNA and mRNA data. An R package of miRlastic was made available through: <http://icb.helmholtz-muenchen.de/mirlastic>.

Keywords: miRNA expression; mRNA expression; mi-/mRNA regulatory network; elastic net regression; local enrichment analysis; head and neck squamous cell carcinoma

1. Introduction

MicroRNAs (miRNAs) represent small-single stranded RNA molecules that own the ability to bind to the complementary 3'-untranslated region (3' UTR) of messenger RNA (mRNA) sequences, to post-transcriptionally fine-tune target mRNA gene expression [1,2]. MiRNA biogenesis was proven to be under tight spatio-temporal control and targeting relationships were shown to be largely restricted to specific cell types or tissues [3–5]. MiRNAs are thought to act in a combinatorial manner and were also shown to modulate oncogene expression [6,7]. Dysregulation of miRNAs has been associated with human cancer and proven to be sufficient for driving oncogenesis in mouse models, while changes on the genetic and epigenetic levels of the miRNA biogenesis have been associated with cancer initiation [8].

miRNAs may act as oncogenes or tumor-suppressor genes, as demonstrated in various studies across many cancer types [9–11], including our own work on head and neck squamous cell carcinoma (HNSCC) [12,13]. HNSCC represents one of the most commonly diagnosed carcinomas worldwide with an incidence of ~600,000 patients per year [14] and is characterized by phenotypic, etiological, biological and clinical heterogeneity [15–18]. Heavy smoking, alcohol abuse and infection with high-risk types of human papilloma virus (HPV), mostly HPV-16, represent the major risk factors associated with HNSCC. Combined treatment procedures include surgery, radiotherapy alone or in conjunction with chemotherapy or immunotherapy. However, the survival rate for patients with advanced HNSCC remains limited to ~50% [19–21]. An oncogenic HPV infection was associated with expression of the viral oncogenes E6 and E7, leading to cell cycle deregulation through E6-induced degradation of p53 and E7-mediated inactivation of the Retinoblastoma (Rb) protein [22,23]. The subsequently caused promotion of cell cycle progression and proliferation were considered to be the onset of HPV-mediated carcinogenesis. HPV+ tumors represent a distinct group within HNSCCs, differing from HPV-tumors in pathogenesis, histopathology, clinical outcome—prognosis was more favorable for HPV+ patients—and molecular biology [15,17,24]. Specific molecular characteristics related to HPV status include gene mutations, genomic copy number aberrations, changes in DNA methylation, mRNA and miRNA expression patterns [17,25]. Lajer *et al.* identified a set of core miRNAs implicated in HPV pathogenesis [25,26]. Strikingly, those HPV core miRNAs are related to the E6/p53 and E7/Rb pathways of HPV induced malignant pathogenesis. The reported involvement of miRNAs in HPV related pathogenesis motivated us to further explore miRNAs associated with HPV infection in HNSCC and to reveal how those miRNAs regulate underlying cellular mechanisms in HNSCCs with respect to HPV status.

Regulation of cellular processes by miRNAs was thought to be mediated through their target genes. It is thus worthwhile to infer target relationships of miRNAs in order to reveal their functional roles. The idea builds on two subsequent steps: (1) identification of mi-/mRNA gene interactions and (2) functional annotation of target genes. As for (1) obtaining mi-/mRNA gene interactions, both computational and experimental methods were used. Several bioinformatics algorithms have been proposed to predict miRNA targets: TargetScan uses a statistical model based on stringent seed pairing, site features and likelihood of preferential conservation [27], miRanda uses a moderately stringent seed pairing algorithm that also considers site number, conservation and free energy [28], while mirSVR represents a machine learning method based on a down-regulation score [29]. The disadvantages of *in silico* target prediction methods reside in the high false positive prediction rates and unspecificity for a given setup [30]. Other approaches inferred miRNA targets solely from expression values [31]. Experimental identification of mi-/mRNA interactions can be achieved through genetic screening, quantification of gene expression changes caused by

miRNA transfection or methods based on crosslinking of mRNA and miRNA-containing argonaute (AGO) complexes followed by immunoprecipitation like Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) [32] or High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) [33]. The StarBase database stores mi-/mRNA target relationships that were identified using PAR-CLIP or HITS-CLIP [34]. Other methods such as the micro-multivariate Markov modeling inference engine (microMUMMIE) [32], PAR-CLIP miRNA assignment (PARma) [35] or MIRZA [36] used CLIP data to computationally infer the specific miRNA that guides the interaction of AGO with a gene target.

While simultaneous measurements of entire transcriptomes including mRNAs and miRNAs have become relatively straightforward with high-throughput techniques, their integration is not trivial. Several methods have been proposed to integrate miRNA and mRNA data for the identification of mi-/mRNA networks. An intuitive approach for finding associations between miRNAs and mRNAs on expression level is correlation analysis [37,38]. Generally, correlation analyses cannot adequately model the demonstrated joint effects of several miRNA on a shared target [39]. To account for the miRNA joint effects, multiple linear regression models have been proposed. An appropriate model, which has been used for target inference on expression level [40], was given by the least absolute shrinkage and selection operator (lasso) regression [41]. It imposes sparsity by an L1 penalty on the regression coefficients shrinking them towards zero, thus allowing for feature selection on the variables in the model. However, lasso selects one representative from each correlated group of miRNAs, so it does not account for co-expression. In contrast to lasso, the ridge regression model maintains all predictors in the model by using an L2 penalty but does, in turn, not perform feature selection. To overcome these drawbacks, an elastic net penalty [42] was used for the integration of miRNA and gene expression data [43], which combines the L1 and L2 penalties in order to account for co-expression among miRNAs and, at the same time, performs feature selection for potential regulatory relationships with target genes. We also previously showed that the usage of elastic net regression in combination with a negativity constraint on the coefficients provides reasonable results both on mRNA and proteome level [44,45].

To reveal functional properties of target gene sets, common approaches were based on statistical enrichment testing to identify over-represented functional categories [46–48]. For example, miRGator [47] was proposed to infer miRNA functions by performing a statistical enrichment test of target genes in each term for gene ontology, pathway and disease annotations available on the PhenoMir platform [49]. These methods typically generate a large amount of significantly enriched functional categories, however, many of which are unrelated to the diseases and conditions. A recent study showed that the most commonly used functional enrichment test returns significant *p*-values for targets of randomly selected miRNAs, revealing the intrinsic bias in the target prediction [50]. A combined approach that identifies functional roles of miRNAs from experimental data, together with adequate statistical modeling and miRNA-tailored functional annotation analysis, remains to be done.

We have set out to develop a method, called miRlastic, to infer miRNA functions in a data-driven manner using statistical inference of mi-/mRNA interactions, as well as functional characterization of the inferred network (Figure 1A).

To obtain condition-specific mi-/mRNA interactions, we developed a statistical inference method taking into account matched miRNA and mRNA expression data of the underlying conditions together with prior knowledge of sequence-based predictions.

We used a multiple linear regression model with elastic net penalty, which is a trade-off between lasso and ridge regression that accounts for both joint effects of several miRNAs on a common target (Figure 1B) and co-expression between miRNAs [42]. We ensured selecting down-regulation effects with a negativity-constraint on the regression coefficients.

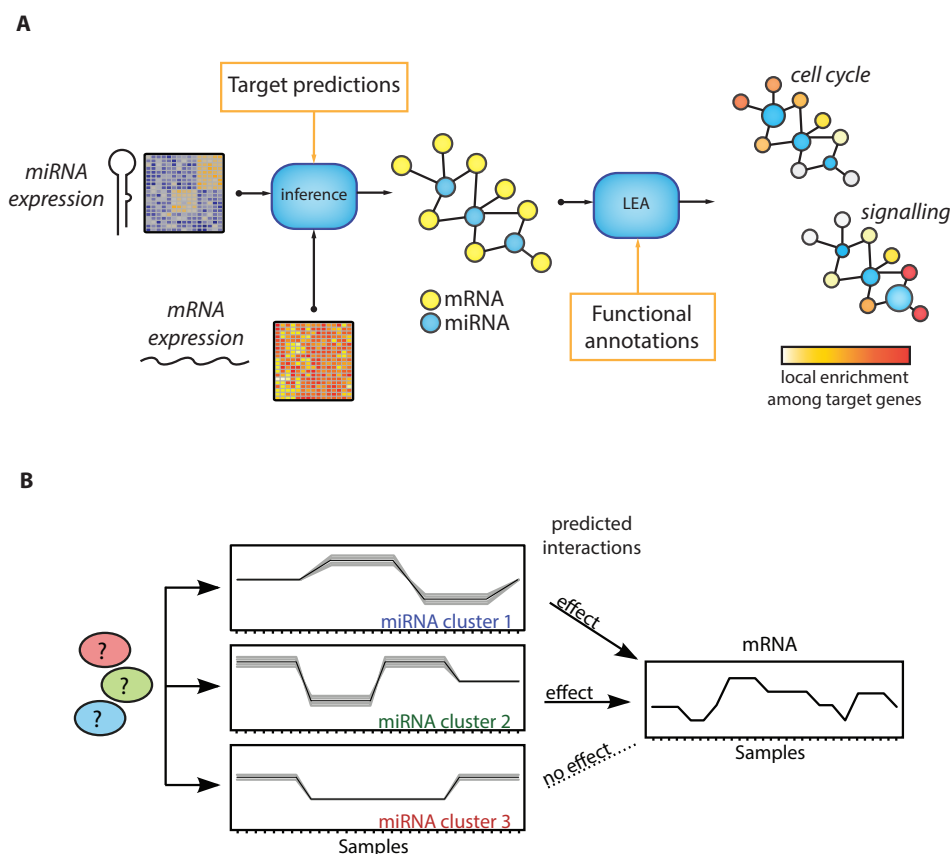


Figure 1. Functional characterization of miRNAs in a condition-specific manner by data-driven inference of mi-/mRNA networks and subsequent functional annotation in the networks. (A) MiRlastic uses a two-step approach integrating prior knowledge and mi-/mRNA expressions to infer mi-/mRNA network and miRNA functional annotation; (B) Schematic drawing of miRNAs co-expressed in clusters induced by an yet unknown regulatory layer. Only several of the putative miRNA regulators are collectively regulating the mRNA expression.

On synthetic data, miRlastic outperformed commonly used methods and was especially suitable for low sample sizes. To functionally annotate miRNAs based on the inferred mi-/mRNA network, we introduced a local enrichment analysis that scores miRNAs given the underlying network structure with respect to functional annotations of target genes.

We thoroughly evaluated miRlastic on a cohort of HNSCC patients provided by The Cancer Genome Atlas (TCGA). We inferred an mi-/mRNA regulatory network for human papilloma virus (HPV)-associated miRNAs in HNSCC. The resulting network best enriched for experimentally validated miRNA-target interactions, when compared to commonly used methods. Finally, the local enrichment analysis (LEA) procedure identified two functionally distinct clusters of miRNAs that were predicted to mediate HPV-associated dysregulation in HNSCC. An R package of miRlastic was made available through miRlastic [51].

2. Methods and Materials

2.1. HNSCC Data

We downloaded Level 3 mRNA and miRNA RNA-seq data from the TCGA data portal. The mRNA Level 3 data consisted of gene expression measurements which were generated following the protocol previously described by the TCGA consortium [52]. For analysis on mRNA expression level, we selected only the mRNAs with non-zero count values in more than 80% of the

patients, non-zero standard deviation and applied a \log_2 transformation. For miRNA analysis, we obtained normalized expression levels for miRNA precursors after selecting only those miRNAs that accomplished the same criteria as the mRNAs. In the first step, we overlaid the precursor entries with associated entries in TargetScan. By doing so, we considered only those miRNAs that are supposed to be incorporated into the RNA-induced silencing complex (RISC) complex and thus not subject to degradation [53]. For each miRNA, for which TargetScan provides predictions on both -3p and -5p strands, we unify the two sets of putative targets and assign them to the corresponding precursors. After running the inference step, we intersect the resulting mi-/mRNA interactions from the individual precursors with the predictions of the mature miRNAs provided by TargetScan. We performed differential miRNA expression analysis on a subcohort of 244 patients for which the human papillomavirus (HPV) status clinical parameter was provided [17] in order to identify deregulated miRNAs between HPV+ and HPV- patients. When using the edgeR package [54], we also included the age and gender as confounder variables. We controlled for a 5% false discovery rate (FDR) using Benjamini and Hochberg algorithm [55].

2.2. Preliminaries

We defined \mathbf{X} and \mathbf{Y} as two matrices that contain miRNA (x_{jk}) and mRNA (y_{ik}) expression data, both simultaneously measured in s samples, with $k \in \{1, \dots, s\}$, $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m\}$, such that n and m represent the number of measured miRNAs and mRNAs, respectively. We denoted the regulatory interaction of miRNAs and their putative mRNA targets as a bipartite graph G . The bipartite graph G captures all putative mi-/mRNA interactions in $G = (V^{miR}, V^{mR}, E)$ with disjoint sets of two node types as V^{miR} and V^{mR} . The set of all mRNAs represents all the nodes listed in $V^{mR} = \{v_1^{mR}, \dots, v_m^{mR}\}$, and likewise the miRNAs represent the nodes in $V^{miR} = \{v_1^{miR}, \dots, v_n^{miR}\}$. The edge set $E = \{e_1, \dots, e_z\}$ connects the nodes from V^{miR} with the ones from V^{mR} as $e_l = (v_u^{miR}, v_w^{mR})$ with $l \in \{1, \dots, z\}$. The edges e_l were extracted from the TargetScan database (Version 6.2) [27] and used as input graph G for the miRlastic inference method yielding an optimized graph, $G' = (V^{miR}, V^{mR}, E')$ with $E' \subseteq E$.

2.3. Group-Wise Correlation of miRNAs with a Shared Target Gene

We exemplarily calculated pairwise Pearson correlation coefficients among all miRNA expression profiles from the HNSCC dataset, which were predicted by TargetScan with the common target C9orf85 (Figure 2A). Note that this data was representative for the entire dataset. We observed subgroups of high correlation which corresponds to the expectation that functionally related miRNAs or miRNAs in high proximity on the chromosome tend to be co-expressed [7,56].

To systematically analyze whether miRNA expression profiles were typically correlated when sharing a putative target, we next defined a measure of correlation strength $c(\mathbf{X})$ for a matrix of miRNA expression measurements $\mathbf{X} = (x_{ik})$ as

$$c(\mathbf{X}) = \frac{\|\mathcal{R}(\mathbf{X})\|_F}{\sqrt{(n^2 - n)/2}}$$

with the Pearson correlation matrix $\mathcal{R}(\mathbf{X}) = (\rho_{i_1 i_2}) = (r_{\mathbf{X}_{i_1}, \mathbf{X}_{i_2}})$ for $i_1, i_2 \in i$. The Frobenius norm of the correlation matrix $\mathcal{R}(\mathbf{X})$ was calculated as

$$\|\mathcal{R}(\mathbf{X})\|_F = \sqrt{\sum_{i_1 < i_2} \rho_{i_1 i_2}^2}$$

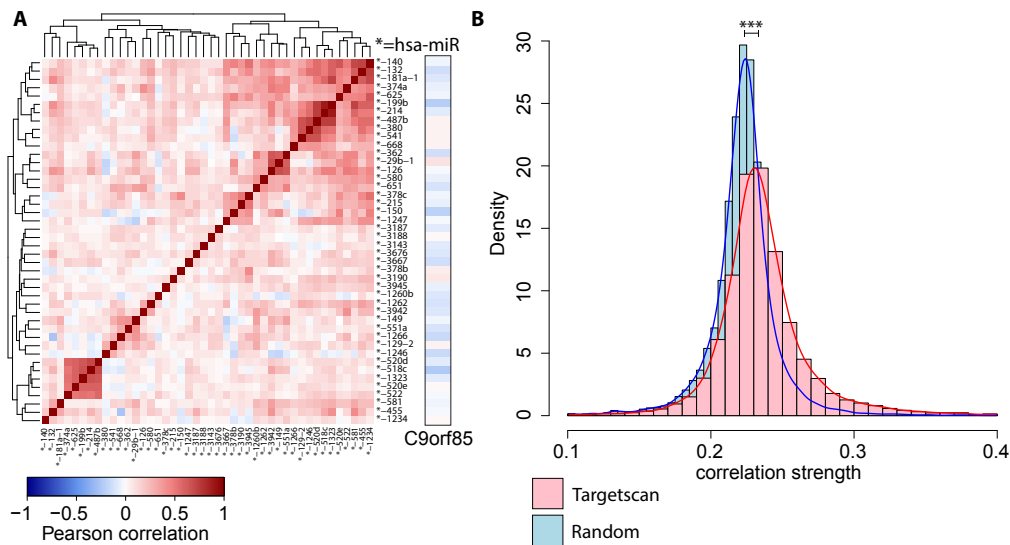


Figure 2. Collective effects of co-expressed miRNAs. (A) Pairwise correlation of putative, expressed miRNA regulators together with their target gene C9orf85 (obtained from the HNSCC dataset). The miRNAs are themselves clustered into several co-expressed groups; (B) The distribution of correlation strengths $c(X)$ of miRNA sets, which are predicted to target a common gene, (red curve and histogram) is higher than for randomly re-sampled mi-/mRNA associations (blue, Wilcoxon rank sum test has $p < 1 \times 10^{-80}$) in the HNSCC miRNA expression dataset.

Note that only the upper triangular matrix with $(n^2 - n)/2$ elements was considered for the calculation of the Frobenius norm. As all elements of $\mathcal{R}(X)$ range between $[-1, 1]$, all values of $c(X)$ range between $[0, 1]$. The extreme values $c(X) = 0$ and $c(X) = 1$ indicate an entirely uncorrelated and perfectly (anti-)correlated set of miRNAs, respectively.

Using the correlation strength measure $c(X)$, we evaluated the properties of collective miRNA regulation by assessing the (anti-)correlation strength across all miRNAs, which are predicted to target a common mRNA. We found that these sets of miRNAs were more correlated among each other than randomly sampled sets of miRNAs ($p < 1 \times 10^{-80}$ Wilcoxon rank sum test, Figure 2B).

2.4. mi-/mRNA Network Inference

To model the regulatory behavior of several miRNAs jointly modulating one common mRNA target v_j^{mR} , we extracted from G all connected n^* miRNAs v_i^{miR} with $i^* = \{i | \exists (v_i^{miR}, v_j^{mR}) \in E\}$. We refer to these observations of one mRNA j as y_j and to its associated miRNA observations as $X(j) = X_{i^*}^T$, where the one-dimensional vector y_j and the $s \times n^*$ -dimensional matrix $X(j)$ represent the response and the predictors of a regression model.

The mi-/mRNA interaction can be modeled without feature selection by a linear regression model, separately for each mRNA as:

$$y_j \sim \beta_{j0} + X(j)\beta_j + \epsilon$$

with the normally distributed error $\epsilon \sim N(0, \sigma)$, parameters $\beta_j = (\beta_{j1}, \dots, \beta_{jn^*})$ and intercept β_{j0} .

As shown above, high co-expression among miRNAs is typical for miRNAs with functional similarity and thus, needs to be taken into account when performing feature selection. From correlation between miRNAs and their target gene, the effects of joint actions of miRNAs were detectable.

For network inference, we next performed feature selection using penalized regression (elastic net [42]). We therefore introduced a negativity constraint on the coefficients β_j to account for the fact

that miRNAs mainly down-regulate mRNA target expression levels [57]. The β_j values were obtained by solving the following optimization problem:

$$\hat{\beta}_j = \underset{\beta_j}{\operatorname{argmin}} |y_j - X(j)\beta_j + \lambda \mathcal{P}_\alpha(\beta_j)|,$$

where

$$\mathcal{P}_\alpha(\beta_j) = (1 - \alpha) \frac{1}{2} \|\beta_j\|_2^2 + \alpha \|\beta_j\|_1.$$

The parameter α_j with $0 \leq \alpha_j \leq 1$ denotes the elastic net mixing parameter. The second parameter λ was identified by 10-fold cross-validation using the glmnet package. For estimating the regression coefficients $\hat{\beta}_j$, we used a coordinate descent approach [58], which is also implemented in the glmnet package.

To tune the elastic net penalty, we adjusted α with respect to the potentially expected fraction of correlated predictor groups. The parameter α_j of the elastic net regression model of y_j , given $X(j)$, is then defined as $\alpha_j = 10^{-c(X(j))}$.

The choice of the parameter α allowed for an unbiased parameter tuning whereas lower values of α_j were slightly preferred. However, we wanted α_j to remain non-zero in any case such that feature selection was performed in all models. Therefore, gene-specific choice of α_j parameter was a good trade-off.

2.5. Functional Characterization of miRNA-Target Networks

To identify local, closely-connected functions within the inferred network, we designed a local enrichment analysis (LEA) procedure. We evaluated whether node arrangements were assigned to a certain term, describing e.g., a molecular function or biological process, occurred by chance or not. Proximity of two target gene nodes in the network were measured by their shortest path. The bipartite graph $G' = (V^{miR}, V^{mR}, E')$ was the input for LEA. The edges and edge weights were represented as a matrix $W = (w_{ij})$ with z non-zero entries. In this case, edge weights were given by scaled negative regression coefficients. We transformed weights by e^w to obtain positive edge weights.

A path between two gene nodes was defined as $P(a, b) = (v_1^{mR}, \dots, v_p^{mR})$ with $v_1^{mR} = v_a^{mR}$ and $v_p^{mR} = v_b^{mR}$ such that there exists an miRNA node connected to both nodes v_k^{mR} and v_{k+1}^{mR} in $P(a, b)$ for all $1 \leq k < p$. The distances of a path between two nodes v_a^{mR} and v_b^{mR} were computed as:

$$d(a, b) = \sum_{k=1}^{p-1} \min_{i^*} (w_{i^*k} + w_{i^*k+1})$$

where i^* denotes the miRNAs that target the mRNAs k and $k + 1$. A path $P(a, b)$ between the mRNAs a and b is then called shortest path $P_{min}(a, b)$ if it minimizes the distance $d(a, b)$.

Next, we scored local neighborhoods in the network. Let $M_k = g_{1k}, \dots, g_{mk}$ denote the gene set consisting of m annotated genes for a specific term k with $k \in 1, \dots, l$, which may be, for example, retrieved from an online repository like the Kyoto Encyclopedia of Genes and Genomes (KEGG) [59], and let $M_{k,G'} := M_k \cap V^{mR}$ be the set of genes in M_k that overlap with the genes in G' . In order to determine the enrichment of $M_{k,G'}$ around a certain mRNA j in G' , we compared the distribution of the shortest path distances $D_{min}(j, M_{k,G'})$ to the background distribution $D_{min}(j, G' \setminus M_{k,G'})$. Note that $D_{min}(j, M_{k,G'})$ included the shortest path distance to a node itself, which is defined as zero, if $v_j \in M_{k,G'}$. We applied a left-tailed Wilcoxon rank-sum test, which resulted in a p -value indicating whether the values of $D_{min}(j, M_{k,G'})$ were significantly shifted towards lower values when compared to values of $D_{min}(j, G' \setminus M_{k,G'})$. We used p -values to assess a score $S(v_j)$ as $S(v_j) = -\log_{10}(p_j)$ that described the enrichment of the term for the given functional group M_k around gene j .

To finally characterize the importance of miRNAs in G' , we calculated the miR score $S_{miR}(v_i)$ for every node $v_i \in V^{miR}$ by considering the set $\mathcal{V} := \{v_j | (v_j \in V_{mR}) \wedge (\exists (v_i, v_j) \in E)\}$ of associated mRNA nodes. We also introduced a weight that accounts for the number of corresponding targets $|\mathcal{V}_i|$ for each miRNA. The score was then defined as:

$$S_{miR}(v_i) = \left(\frac{1}{|\mathcal{V}_i|} \sum_{v_j \in \mathcal{V}_i} S(v_j) \right) \cdot \sqrt{|\mathcal{V}_i|}$$

Figure 3 illustrates the LEA scoring. The inferred mi-/mRNA interactions were subjected to LEA with the nodes A, I, J, P being part of the gene set assigned to the term (term nodes, Figure 3A, diamond shaped). Exemplary shortest distances between node M (blue) or B (purple) and respective term nodes were calculated.

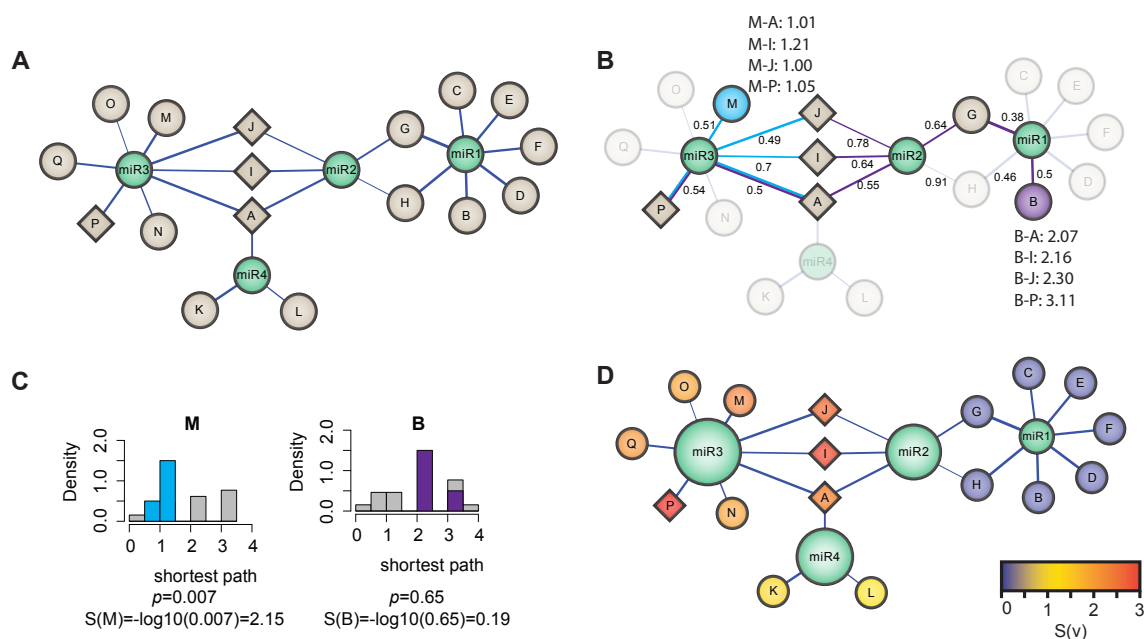


Figure 3. Node scoring strategy of local enrichment analysis. (A) Example of a network in which the nodes are represented by miRNAs (green) and genes (grey) and the edges are represented by the mi-/mRNA interactions. The network is given as input for the functional annotation step. We assume that four genes are assigned to a certain functional group—A, I, J, P (diamond shape); (B) The transformed network after computing the shortest path distances between these four nodes and the two nodes M (blue) and B (purple). The edge labels denote the weights after the transformation. The edge weight indicates the strength of the mi-/mRNA negative regulation; (C) The distribution of shortest paths from node M to the nodes A, I, J and P is significantly shifted to lower values. No shift can be observed for node B. The p -values determined by left-tailed Wilcoxon rank-sum test are converted to the node scores; (D) Network visualization after annotating functional groups. The node scores are indicated by the color. The size of the miRNA nodes corresponds to the miR score.

We next compared the distributions of shortest distances to the background distribution that represents the shortest distances to genes, which did not overlap with the gene set assigned to the term (Figure 3C). We observed that, in the case of node M, the distribution of shortest distances to the gene set assigned to the term tends to be shifted to lower values as compared to the background distribution, which was not the case for node B. In order to statistically test for this shift to lower values, we apply a left-tailed Wilcoxon rank-sum test and obtain a significant p -value that is ($p = 7 \times 10^{-3}$) in the case of node M, while the p -value for node B was not significant $p = 6.5 \times 10^{-1}$. We can thus conclude that node M was actually located in close proximity of the nodes associated

with the term k , whereas node B is not. Given the p -values from the left-tailed Wilcoxon rank-sum test, we can calculate the scores of the two nodes M and B as $S(M) = -\log_{10}(7 \times 10^{-3}) = 2.15$ and $S(B) = -\log_{10}(6.5 \times 10^{-1}) = 0.19$, respectively. This score indicated the proximity of a gene to the genes in the functional group.

Finally, we were interested in finding local enrichment of a set of terms given their associated gene sets. With G' being the mi-/mRNA network and $M_{k,G'}$ the set of overlapping genes between the gene set M_k of term k and the genes in G' , we selected all shortest distances $D_{min}(M_{k,G'})$ between the nodes $v_j \in M_{k,G'}$. We compared the distribution of this set to the distribution of all other shortest distances $D_{min}(V_{mR} \setminus M_{k,G'})$. We applied a left-tailed Wilcoxon rank-sum test to assess the enrichment of associated terms in a local area of the network. Terms were then considered to be locally enriched if their adjusted p -values (Bonferroni corrected) were below a significance threshold of $p < 0.05$.

2.6. Implementation and Availability

The whole miRlastic pipeline was implemented within the R environment for statistical computing [60]. For elastic net regression we used the *glmnet* package [58]. To calculate the shortest paths in a given mi-/mRNA network, we used the implementation of Dijkstra's algorithm [61] in the *igraph* [62] package.

MiRlastic can be downloaded as an R package from <http://icb.helmholtz-muenchen.de/mirlastic> [51].

2.7. Synthetic Data

We set up a test environment: in each of the test runs, we generated a set of synthetic miRNA and mRNA expression values adapting to biological features. We modeled a set of miRNAs with expression values $x_i \sim \mathcal{N}(0,1)$ targeting a common mRNA. Furthermore, expression levels were modeled for a set of unknown factors, $h_j \sim \mathcal{N}(0,1)$ which were assumed to target a distinct subset of the predicted miRNAs (Figure 4a). We generated synthetic miRNA expression values by assuming a randomly coordinated regulation as it is the case, for example, for clustered miRNAs. Assuming a repressive effect of the targeting miRNAs, the mRNA expression profile was then generated as

$$y = \sigma\epsilon + \sum_i -\hat{x}_i$$

where $\epsilon \in \mathcal{N}(0,1)$ corresponds to the noise arising from biological reasons or experimental artifacts. We performed tests with different magnitudes of the error weight σ . In addition to the miRlastic inference, we also applied correlation analysis and lasso on the generated profiles. For the correlation analysis, a synthetic miRNA was considered to be a true regulator of a mRNA if the adjusted p -value (Bonferroni corrected) of the negative Pearson correlation coefficient was below 0.05. For lasso, we used the miRlastic inference method with a fixed $\alpha = 1$. The entire procedure was repeated three times with 10, 30 and 50 samples each across 500 runs and varying noise levels. In each parameter setting, the corresponding F_1 measure was calculated.

We counted the number of falsely chosen miRNAs (false positives, FP) and the number of correctly missed miRNAs (false negatives, FN) obtained after several runs. Based on precision and recall rate, we computed $F_1 = 2 \cdot (\text{precision} \cdot \text{recall}) / (\text{precision} + \text{recall})$, where $\text{precision} = TP / (TP + FP)$ and $\text{recall} = TP / (TP + FN)$, for any related confusion matrix comparing actually classes (true and false) to any classification results (positive and negative).

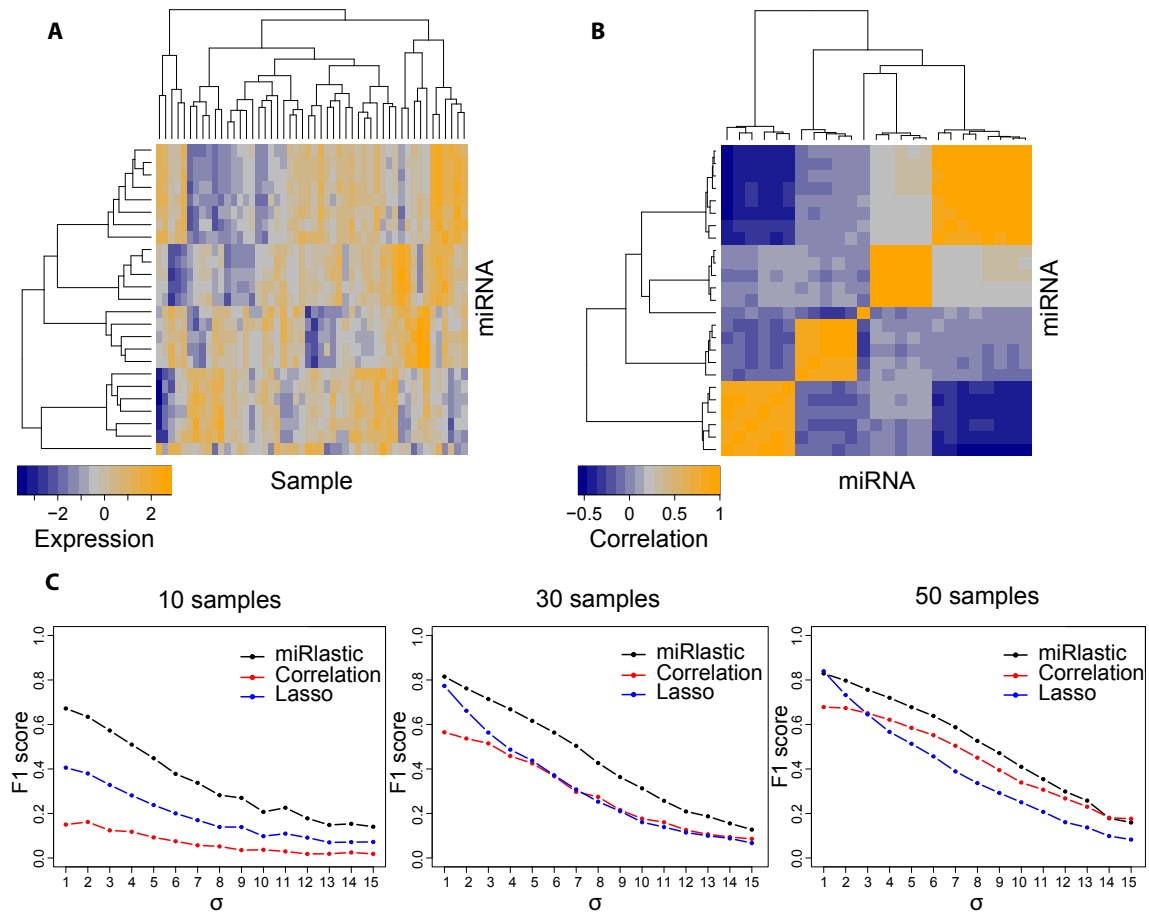


Figure 4. Benchmark with synthetic data. (A) Heatmap illustrating a set of randomly generated synthetic miRNAs with 30 samples; (B) Heatmap of pairwise correlations between the generated miRNAs; (C) Success-rate (F_1) of all algorithms across varying sample numbers and noise levels to recover the true synthetic mi-/mRNA associations.

2.8. Validation of Network Inference on Experimental HNSCC Data

We compared the network inference results, by accounting for how many mi-/mRNA target predictions were experimentally validated. Experimentally validated interactions were downloaded from StarBase [34]. As background set, we considered only those mi-/mRNA interactions with medium stringency from the starBase database intersected with the set of miRNAs and mRNAs also predicted by TargetScan. Note that TargetScan was used as prior network for miRlastic and all related methods, such that we compared all inference methods to the fraction of experimentally validated interactions in TargetScan itself. For validation, we performed Fisher’s exact test to test whether inferred mi-/mRNA pairs by one method (e.g., inferred using miRlastic) were enriched for experimentally validated pairs from StarBase.

3. Results and Discussion

We extensively evaluated the performance of the miRlastic network inference using synthetic data and further applied the miRlastic pipeline on HNSCC patients available on the TCGA data portal.

3.1. Robust mi-/mRNA Network Inference for Small Sample Sizes

In order to assess the performance of our inference approach, we built a test environment generating synthetic miRNA and mRNA expression profiles with characteristic biological features.

We generated expression profiles of sets of miRNAs targeting the same mRNA (Figure 4A) such that miRNAs were set-wise highly correlated among each other while the correlation to miRNAs in different groups was rather low (Figure 4B). The mRNA expression was then calculated as being regulated by only a subset of miRNAs and additional noise. The set-wise miRNA correlations for the synthetic expression data was in accordance with the HNSCC expression data as shown in Figure 2A.

We evaluated the performance of miRlastic in comparison with the related methods, namely pairwise Pearson correlation analysis and lasso. Correlation analysis showed low performance for low sample numbers, whereas the results improved for high sample numbers. Lasso performed good for medium and high sample numbers, yet only for low noise levels, indicating reduced robustness against noisy observations. Throughout all settings, the miRlastic inference method outperformed correlation analysis and Lasso, especially for low sample numbers (Figure 4C).

The benchmark on synthetic data indicated that the miRlastic inference method was able to provide reasonable results even when applied on low-dimensional datasets. This comes as another advantage, since low sample datasets are ubiquitous in biological research due to the costs required for large-scale techniques. The number of matched samples tend to be lower especially for combined expression data (miRNA and mRNA), since the measurements need to be performed twice. Here, we could show that the miRlastic inference was able to reliably identify true regulators with high specificity and sensitivity in a biologically reasonable synthetic test environment.

3.2. mi-/mRNA Regulatory Network of Head and Neck Squamous Cell Carcinoma

To understand HPV-associated miRNA-mediated gene regulation of HNSCC, we applied the miRlastic inference to the HNSCC TCGA data set. As HPV was shown to disrupt cellular differentiation in HNSCC [17,63,64], we used only those 244 patient samples with reported HPV status. Subsequent to preprocessing, target predictions were available for 16,617 mRNAs and 600 miRNAs. We performed a differential analysis between HPV+ and HPV- patients yielding 44 significantly differentially expressed miRNAs (Figure 5A). Among the differentially expressed miRNAs, we observed the miR-9 family, miR-363, miR-20b which have been associated with the HPV status in several independent studies [26,65–69]. In agreement with the results of the Wald *et al.* study [70], miR-363 showed up-regulation in the HPV infected patients. We performed the miRlastic inference using 135,391 targets predicted by TargetScan in combination with the respective miRNA and mRNA expression values. MiRlastic inferred 766 mi-/mRNA interactions (Figure 5B).

We next asked whether the network inferred by miRlastic was specifically enriched for experimentally validated mi-/mRNA interactions. Therefore, we collected experimentally validated interactions from StarBase and compared performance to three other existing methods, namely lasso and Pearson as well as Spearman correlation analysis. A Fisher's exact test was conducted in order to determine whether the fraction of inferred and validated interactions was higher than expected from the prior target network (TargetScan) with respect to the number of inferred interactions (Figure 5C). For miRlastic, the test yielded a highly significant p -value of $p = 8.736821 \times 10^{-4}$. We obtained a p -value of $p = 1.059271 \times 10^{-2}$ for lasso, $p = 1.228527 \times 10^{-1}$ for Pearson correlation and $p = 9.978787 \times 10^{-1}$ for Spearman. These results indicated that the miRlastic approach was able to identify a higher fraction of validated target predictions as compared to the other methods. Lasso also performed well with a significantly fraction of experimentally identified interactions but when compared to miRlastic was clearly lower (Figure 5C). Pearson and Spearman correlation did not show any significant difference. Taken together, miRlastic inference outperformed other commonly used methods with respect to the over-representation of experimentally validated mi-/mRNA interactions.

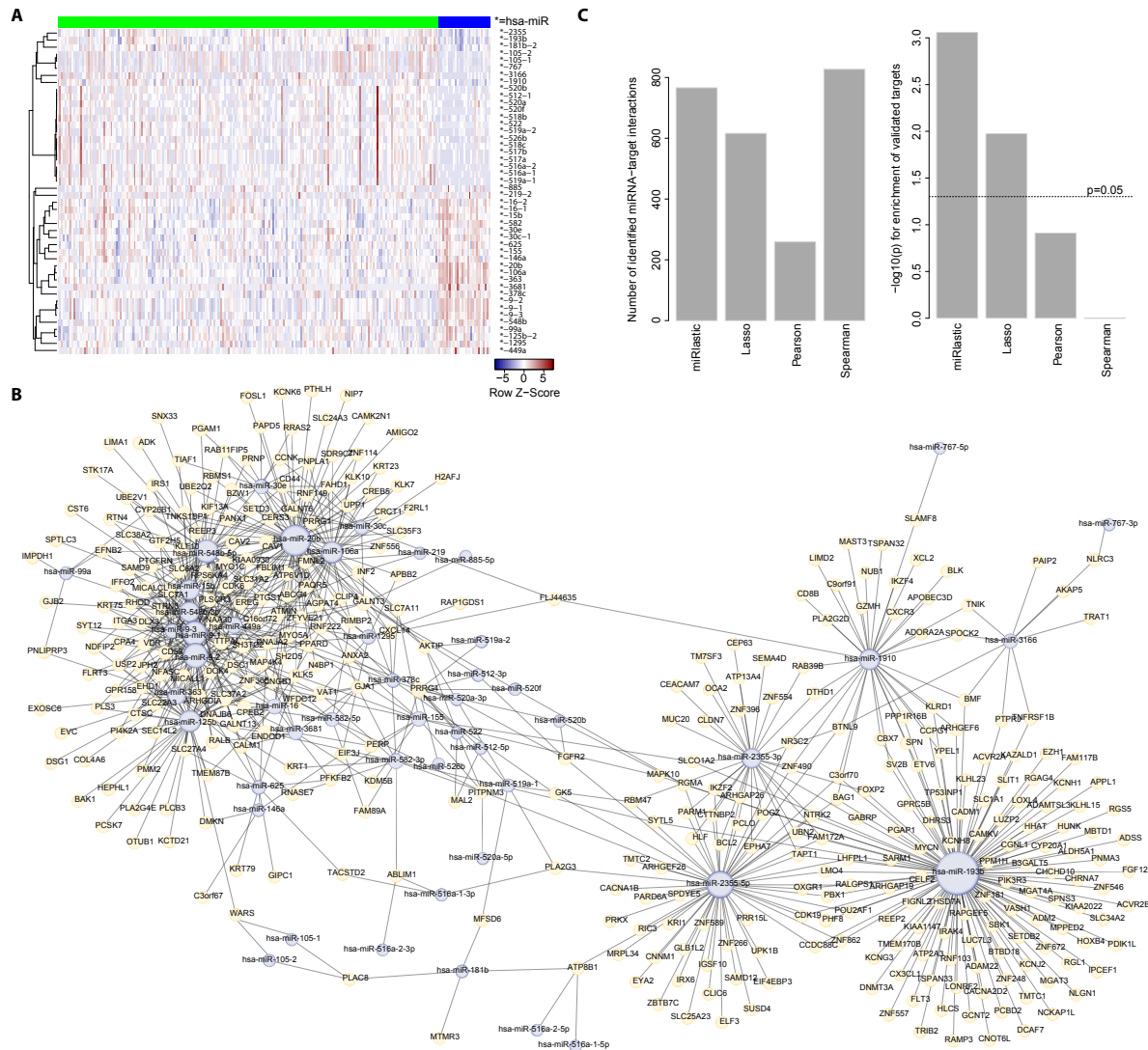


Figure 5. mi-/mRNA regulatory network inference in HNSCC samples. (A) 44 miRNA expression profiles show differential expression between HPV+ (blue) and HPV- (green) patients; (B) mi-/mRNA regulatory network generated by our inference algorithm. The network consists of 766 regulations between 44 miRNAs (light blue) and 16,617 genes (light yellow). The edges represent miRNA-target relationships in the context of the TCGA HNSCC cohort; (C) Performance evaluation of the mi-/mRNA associations inference module. The left bar plot indicates the number of mi-/mRNA interactions detected by our approach, lasso, Pearson and Spearman. The right bar plot shows the $-\log_{10}$ -transformed p -values when testing for enrichment of experimentally validated targets within the results of each method. We provide an interactive representation of this network at <http://icb.helmholtz-muenchen.de/mirlastic/hnsc> [71].

3.3. Two Functional Clusters in miRNA-Mediated HPV-Associated Dysregulation

In the previous sections, we described the network that resulted from the analysis of mi-/mRNA data from tumors of patients from the TCGA HNSCC cohort with known HPV status. To finally gain insights into the functional role of the identified mi-/mRNA pairs, we analyzed miRNA-target genes for local enrichment given the HNSCC network. For the application of LEA on the HNSCC mi-/mRNA interaction network, we downloaded 108 pathways from KEGG for gene annotations [59].

In total, we obtained nine significantly locally enriched pathways (Figure 6). Overall, distinct frequencies of genetic alterations affecting the key signaling pathways of G protein Ras (RAS) and

mitogen-activated protein kinases (MAPK), as well as apoptosis, were reported before for HPV+ and HPV–HNSCC on gene level: for example, TP53, HRAS, MYC, BIRC2 and CASP8 were most often altered in HPV– tumors, whereas HPV+ cases were characterized by PIK3CA mutations, inactivation of TRAF3, the viral genes E6, E7 and amplification of E2F1 [17].

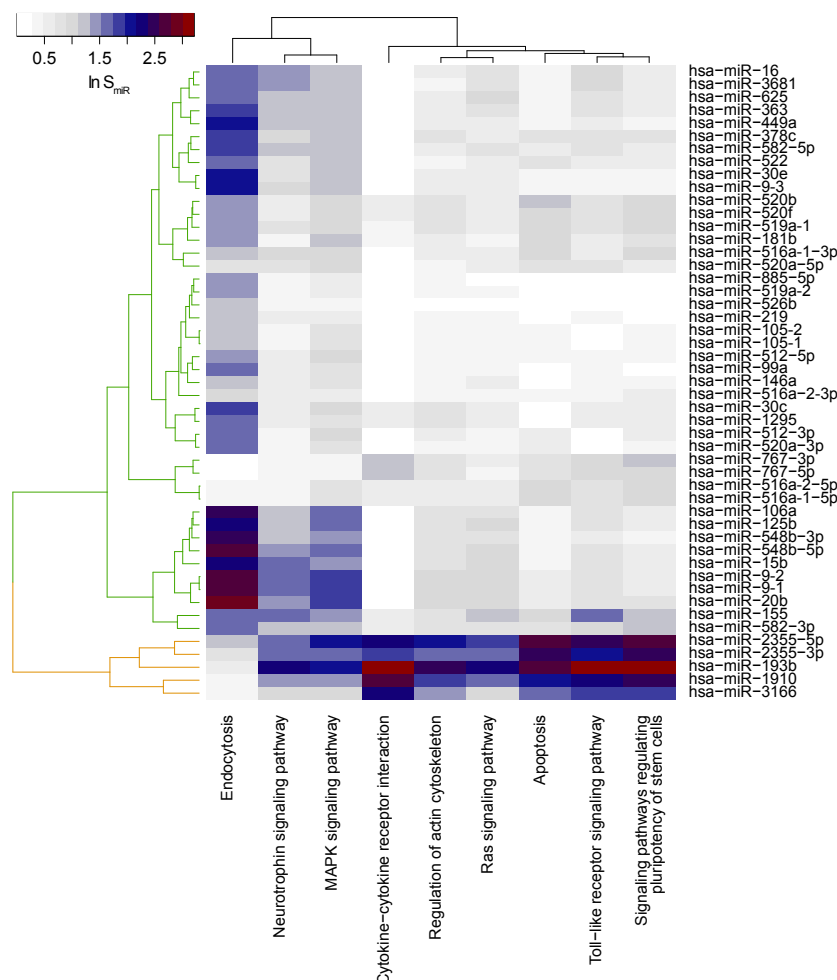


Figure 6. Functional characterization of the HNSCC mi-/mRNA network. Heatmap of miR scores $S_{miR}(v)$ for each miRNA v in the network indicating the functional role in the significantly locally enriched KEGG pathways.

We next clustered miRNAs according to their functional miR score that reflected the strength of interactions of miRNAs with a particular pathway identified two clusters with similar pattern of miR scores across pathways (Figure 6).

The two pathway MAPK- and neurotrophin-signaling pathway associated with both clusters suggesting that they were orchestrated by most miRNAs in HPV-associated dysregulation. Notably, MAPK-, RAS-, neurotrophin and toll-like receptor (TLR) signaling share common elements such as a series of MAPKs, protein kinase B (AKT) and most importantly the downstream located transcription factor NF- κ B, which is involved in regulating chief processes governing tumorigenesis such as immune response, apoptosis, proliferation and angiogenesis [72,73]. Interestingly, hsa-miR-193b showed a strong association with these pathways which was so far shown to play important roles in MAPK signaling [74] and independently shown for HNSCC [75]. This finding validated the miRlastic approach analyzing miRNA regulation in a data-driven manner with subsequent scoring by functional annotation in a network-context.

The first miRNA cluster was characterized by associations of many miRNAs (hsa-miR-106a, -125b, -548b-3p/5p, -15b, -9-2, -9-1, -20b, -155, and -582-3p) mostly with the endocytosis pathway. Endocytosis was shown to be deregulated in cancer cells in such a way that cell surface molecules with growth advantage effects were recycled at higher rates while molecules with reduced growth effects or those making cells recognized by the immune system were removed from the cell surface [76].

The second miRNA cluster was functionally broader and characterized by associations between hsa-miR-2355-5p/3p, -193b, -1910 and 3166 with cytokine-cytokine receptor (CCR) interaction, regulation of actin cytoskeleton, RAS signaling, apoptosis, toll-like receptor (TLR) signaling and signaling involved in regulating pluripotency of stem cells. Disruption of the apoptosis pathway is one of the hallmarks of cancer and is the result of bypassing cell cycle control in cancer cells, which, in turn, leads to uncontrolled proliferation of cells [77]. Disturbance of the regulation of the actin cytoskeleton was shown to play a role in the motility and invasion potential of HNSCC cells, which were key features of metastasizing cells [78]. Cancer stem cells (CSC) were described to be involved in therapy resistance of HNSCC, while HPV+ HNSCCs have been shown to harbor smaller proportions of CSC compared to HPV-HNSCCs [79]. Consequently, these results provided a possible explanation for the favorable prognosis of HPV+ HNSCC patients and link the pathway “signaling pathways regulating pluripotency of stem cells” also to HPV-infection. The expression patterns of TLR, which not only play a role in the immunological defense against pathogens but also in cancer, were demonstrated to be related to the HPV-status [80]. The HPV oncogenes E6 and E7 have been shown to reduce expression of TLRs, thus linking the TLR signaling pathway to HPV+ HNSCCs [81].

The oncoproteins E6 and E7 are known to bind several cellular targets, which have been described to regulate, among others, cytokine signaling, thereby providing a link between HPV-status and the “cytokine-cytokine receptor interaction” pathway [82]. Cytokine-cytokine receptor interaction in general is a common mechanism of cell-cell communication between tumor cells and surrounding non-tumor cells and plays an important role in regulation of the tumor driving mechanisms immune suppression and angiogenesis [83]. The miRlastic analysis suggested that few miRNAs, predominantly hsa-miR-193b and -2355-5p, mediated gene dysregulation largely through modulation of TLR-CCR-interaction signaling and signaling regulating pluripotency of stem cells.

In all, the pathways identified applying miRlastic were in context with signaling disturbance known to be associated with HNSCC tumorigenesis. Moreover, the identified pathways were also known to be related to HPV-status of HNSCC, strengthening the validity of the generated mi-/mRNA regulatory network. Our finding suggested that the two clusters of miRNA-mediated target gene regulation were characterized by distinct signaling pathways deregulated in HPV+ *versus* HPV-HNSCC and, therefore, support plausibility of the results in the context of the biology of HPV+ HNSCC tumors.

4. Conclusions

We have demonstrated the power of coupling an miRNA-characteristic mi-/mRNA target inference method to a local enrichment analysis on the inferred network, in order to identify functional roles of miRNAs in a given experimental setting. The proposed target inference of miRlastic employed a linear regression model with elastic net penalization and negativity-constrained coefficient estimation. The choice of balancing between L1 and L2 was achieved by capturing the co-expression structure of miRNAs independently for each common target gene. The miRlastic inference was able to best enrich for experimentally validated interactions. The local enrichment analysis and subsequent clustering of functional miR scores across pathways, finally allowed us to assign functional roles to sets of miRNAs as given by short distances in the inferred networks between target genes of a selected pathway. The method was made publicly available as an R package. We applied miRlastic to HPV-associated mi-/mRNA regulation in HNSCC data and were able to identify two miRNA clusters of each distinct functional roles. The HPV-associated analysis in HNSCC patients provided, via the identified mRNA targets, descriptive and mechanistic insights

into the molecular phenotype of HPV-driven HNSCC and further validated the meaningfulness of the miRlastic method at a biological and clinical level. We used precursor miRNA expression instead of mature miRNA expression, which was used in previous applications of our approach [44,45], and thus proved that miRlastic is independent from the transcriptional status. In summary, we designed a method for which we showed that it provided a valuable basis for identifying functional roles of miRNAs in a disease context. On the one hand, we selected miRNAs that have a promising role as potential biomarkers due to the significant association with a disease-specific cellular process. On the other hand, this knowledge can further be used to identify potential therapeutic targets, e.g., for cancer treatment, since it has been shown that tumor growth may be prevented either by systemic administration of miRNAs [84] or miRNA silencing [85].

Acknowledgments: We thank Ivan Kondofersky for statistical help in developing the inference method and Gökçen Eraslan for further statistical help. This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grant # 01ZX1313C) and the European Research Council (Latent Causes: 259294); the Deutsche Forschungsgemeinschaft (InKoMBio: SPP 1395 (TH 900/3-2)) and (Psyncourse).

Author Contributions: Steffen Sass, Nikola S. Mueller and Fabian J. Theis developed and refined the method. Adriana Pitea analyzed HNSCC data and evaluated the method. Nikola S. Mueller and Fabian J. Theis planned and supervised the study. Kristian Unger and Julia Hess analyzed biological results. Steffen Sass and Adriana Pitea drafted the manuscript and Steffen Sass, Adriana Pitea, Nikola S. Mueller, Kristian Unger, Julia Hess and Fabian J. Theis wrote the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lewis, B.; Shih, I.; Jones-Rhoades, M.; Bartel, D.; Burge, C. Prediction of mammalian microRNA targets. *Cell* **2003**, *115*, 787–798.
2. Schmiedel, J.; Klemm, S.; Zheng, Y.; Sahay, A.; Bluthgen, N.; Marks, D.; van Oudenaarden, A. MicroRNA control of protein expression noise. *Science* **2015**, *348*, 128–132.
3. Guo, Z.; Maki, M.; Ding, R.; Yang, Y.; Zhang, B.; Xiong, L. Genome-wide survey of tissue-specific microRNA and transcription factor regulatory networks in 12 tissues. *Sci. Rep.* **2014**, *4*, doi:10.1038/srep05150.
4. Barad, O.; Mann, M.; Chapnik, E.; Shenoy, A.; Blelloch, R.; Barkai, N.; Hornstein, E. Efficiency and specificity in microRNA biogenesis. *Nat. Struct. Mol. Biol.* **2012**, *19*, 650–652.
5. Cohen, S. Use of microRNA sponges to explore tissue-specific microRNA functions *in vivo*. *Nat. Methods* **2009**, *6*, 873–874.
6. Pasquinelli, A. MicroRNAs and their targets: Recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.* **2012**, *13*, 271–282.
7. Sass, S.; Dietmann, S.; Burk, U.; Brabletz, S.; Lutter, D.; Kowarsch, A.; Mayer, K.; Brabletz, T.; Ruepp, A.; Theis, F.; *et al.* MicroRNAs coordinately regulate protein complexes. *BMC Syst. Biol.* **2011**, *5*, 136.
8. Lin, S.; Gregory, R. MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer* **2015**, *15*, 321–333.
9. Ma, X.; Conklin, D.; Li, F.; Dai, Z.; Hua, X.; Li, Y.; Xu-Monette, Z.; Young, K.; Xiong, W.; Wysoczynski, M.; *et al.* The oncogenic microRNA miR-21 promotes regulated necrosis in mice. *Nat. Commun.* **2015**, *6*, 7151.
10. Medina, P.; Nolde, M.; Slack, F. OncomiR addiction in an *in vivo* model of microRNA-21-induced pre-B-cell lymphoma. *Nature* **2010**, *467*, 86–90.
11. Esquela-Kerscher, A.; Slack, F. Oncomirs—MicroRNAs with a role in cancer. *Nat. Rev. Cancer* **2006**, *6*, 259–269.
12. Summerer, I.; Hess, J.; Pitea, A.; Unger, K.; Hieber, L.; Selmsberger, M.; Lauber, K.; Zitzelsberger, H. Integrative analysis of the microRNA-mRNA response to radiochemotherapy in primary head and neck squamous cell carcinoma cells. *BMC Genom.* **2015**, *16*, 654.
13. Summerer, I.; Niyazi, M.; Unger, K.; Pitea, A.; Zangen, V.; Hess, J.; Atkinson, M.; Belka, C.; Moertl, S.; Zitzelsberger, H. Changes in circulating microRNAs after radiochemotherapy in head and neck cancer patients. *Radiat. Oncol.* **2013**, *8*, 1–9.
14. Du, Y.; Peyser, N.D.; Grandis, J.T. Integration of Molecular Targeted Therapy with Radiation in Head and Neck Cancer. *Pharmacol. Ther.* **2014**, *142*, 88–98.

15. Leemans, C.; Braakhuis, B.; Brakenhoff, R. The molecular biology of head and neck cancer. *Nat. Rev. Cancer* **2011**, *11*, 9–22.
16. Stransky, N.; Egloff, A.; Tward, A.; Kostic, A.; Cibulskis, K.; Sivachenko, A.; Kryukov, G.; Lawrence, M.; Sougnez, C.; McKenna, A.; *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **2011**, *333*, 576–582.
17. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **2015**, *517*, 576–582.
18. Pai, S.; Westra, W. Molecular pathology of head and neck cancer: Implications for diagnosis, prognosis and treatment. *Annu. Rev. Pathol.* **2009**, *4*, 49–70.
19. Bar-Ad, V.; Palmer, J.; Yang, H.; Cagnetti, D.; Curry, J.; Luginbuhl, A.; Tuluc, M.; Campling, B.; Axelrod, R. Current management of locally advanced head and neck cancer: The combination of chemotherapy with locoregional treatments. *Semin. Oncol.* **2014**, *41*, 798–806.
20. Orth, M.; Lauber, K.; Niyazi, M.; Friedl, A.; Li, M.; Maihoefer, C.; Schuettrumpf, L.; Ernst, A.; Niemoeller, O.; Belka, C. Current concepts in clinical radiation oncology. *Radiat. Environ. Biophys.* **2014**, *53*, 1–29.
21. Schmitz, S.; Ang, K.; Vermorken, J.; Haddad, R.; Suarez, C.; Wolf, G.; Hamoir, M.; Machiels, J. Targeted therapies for squamous cell carcinoma of the head and neck: Current knowledge and future directions. *Cancer Treat. Rev.* **2014**, *40*, 390–404.
22. Psyrris, A.; DiMaio, D. Human papillomavirus in cervical and head-and-neck cancer. *Nat. Clin. Pract. Oncol.* **2008**, *5*, 24–31.
23. Moody, C.; Laimins, L. Human papillomavirus in cervical and head-and-neck cancer. *Nat. Rev. Cancer* **2010**, *10*, 550–560.
24. Pannone, G.; Santoro, A.; Papagerakis, S.; Muzio, L.L.; Rosa, G.D.; Bufo, P. The role of human papillomavirus in the pathogenesis of head and neck squamous cell carcinoma: An overview. *Infect. Agents Cancer* **2011**, *6*, 4.
25. John, K.; Wu, J.; Lee, B.; Farah, C. MicroRNAs in Head and Neck Cancer. *Int. J. Dent.* **2013**, *2013*, 650218.
26. Lajer, C.; Garnaes, E.; Friis-Hansen, L.; Norrild, B.; Therkildsen, M.; Glud, M.; Rossing, M.; Lajer, H.; Svane, D.; Skotte, L.; *et al.* The role of miRNAs in human papilloma virus (HPV)-associated cancers: Bridging between HPV-related head and neck cancer and cervical cancer. *Br. J. Cancer* **2012**, *106*, 1526–1534.
27. Lewis, B.; Burge, C.; Bartel, D. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **2005**, *120*, 15–20.
28. John, B.; Enright, A.; Aravin, A.; Tuschl, T.; Sander, C.; Marks, D. Human MicroRNA Targets. *PLoS Biol.* **2004**, *2*, e363.
29. Betel, D.; Koppal, A.; Agius, P.; Sander, C.; Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* **2010**, *11*, R90.
30. Bassett, A.R.; Azzam, G.; Wheatley, L.; Tibbit, C.; Rajakumar, T.; McGowan, S.; Stanger, N.; Ewels, P.A.; Taylor, S.; Ponting, C.P.; *et al.* Understanding functional miRNA-target interactions *in vivo* by site-specific genome engineering. *Nat. Commun.* **2014**, doi:10.1038/ncomms5640.
31. Li, X.; Gill, R.; Cooper, N.; Yoo, J.; Datta, S. Modeling microRNA-mRNA interactions using PLS regression in human colon cancer. *BMC Med. Genom.* **2011**, *19*, 44.
32. Hausser, J.; Zavolan, M. Identification and consequences of miRNA-target interactions—beyond repression of gene expression. *Nat. Rev. Genet.* **2014**, *15*, 599–612.
33. Chi, S.; Zang, J.; Mele, A.; Darnell, R. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **2009**, *460*, 479–486.
34. Yang, J.H.; Li, J.H.; Shao, P.; Zhou, H.; Chen, Y.Q.; Qu, L.H. starBase: A database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acid Res.* **2011**, *39*, D202–D209.
35. Majoros, W.; Lekprasert, P.; Mukherjee, N.; Skalsky, R.; Corcoran, D.; Cullen, B.R.; Ohler, U. MicroRNA target site identification by integrating sequence and binding information. *Nat. Methods* **2013**, *10*, 630–633.
36. Khorshid, M.; Hausser, J.; Zavolan, M.; van Nimwegen, E. A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. *Nat. Methods* **2013**, *10*, 253–255.
37. Li, W.; Chen, L.; Li, W.; Qu, X.; He, W.; He, Y.; Feng, C.; Jia, X.; Zhou, Y.; Lv, J.; *et al.* Unraveling the characteristics of microRNA regulation in the developmental and aging process of the human brain. *BMC Med. Genom.* **2013**, *6*, 55.

38. Sales, G.; Coppe, A.; Bisognin, A.; Biasiolo, M.; Bortoluzzi, S.; Romualdi, C. MAGIA, a web-based tool for miRNA and Genes Integrated Analysis. *Nucleic Acids Res.* **2010**, *38*, 352–359.
39. Rinck, A.; Preusse, M.; Laggenbauer, B.; Lickert, H.; Engelhardt, S.; Theis, F. The human transcriptome is enriched for miRNA-binding sites located in cooperativity-permitting distance. *RNA Biol.* **2013**, *10*, 1125–1135.
40. Lu, Y.; Zhou, Y.; Qu, W.; Deng, M.; Zhang, C. A Lasso regression model for the construction of microRNA-target regulatory networks. *Bioinformatics* **2011**, *27*, 2406–2413.
41. Muniategui, A.; Nogales-Cadenas, R.; Vasquez, M.; Arangueren, X.; Luttun, A.; Prosper, F.; Pacual-Montano, A.; Rubio, A. Quantification of miRNA-mRNA Interactions. *PLoS ONE* **2012**, *7*, e30766.
42. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* **2005**, *67*, 301–320.
43. Beck, D.; Ayers, S.; Wen, J.; Brandl, M.; Pham, T.; Webb, P.; Chang, C.; Zhou, X. Integrative analysis of next generation sequencing for small non-coding RNAs and transcriptional regulation in Myelodysplastic Syndromes. *BMC Med. Genom.* **2011**, *4*, 19.
44. Meyer, S.; Sass, S.; Mueller, N.; Krebs, S.; Bauersachs, S.; Kaiser, S.; Blum, H.; Thirion, C.; Krause, S.; Theis, F.J.; *et al.* Integrative Analysis of MicroRNA and mRNA Data Reveals an Orchestrated Function of MicroRNAs in Skeletal Myocyte Differentiation in Response to TNF- α or IGF1. *PLoS ONE* **2015**, *10*, e0135284.
45. Waldera-Lupa, D.; Kalfalah, F.; Florea, A.; Sass, S.; Kruse, F.; Rieder, V.; Tigges, J.; Fritsche, E.; Krutmann, J.; Busch, H.; *et al.* Proteome-wide analysis reveals an age-associated cellular phenotype of in situ aged human fibroblasts. *Aging* **2014**, *6*, 856–878.
46. Kowarsch, A.; Preusse, M.; Marr, C.; Theis, F. miTALOS: Analyzing the tissue-specific regulation of signaling pathways by human and mouse microRNAs. *RNA* **2011**, *17*, 809–819.
47. Nam, S.; Kim, B.; Shin, S.; Lee, S. miRGator: An integrated system for functional annotation of microRNAs. *Nucleic Acids Res.* **2008**, *36*, 159–164.
48. Wong, N.; Wang, X. miRDB: An online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* **2015**, *43*, D146–D152.
49. Ruepp, A.; Kowarsch, A.; Schmidl, D.; Buggenthin, F.; Brauner, B.; Dunger, I.; Fobo, G.; Frishman, G.; Montrone, C.; Theis, F. PhenomiR: A knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.* **2010**, *11*, R6.
50. Bleazard, T.; Lamb, J.; Griffiths-Jones, S. Bias in microRNA functional enrichment analysis. *Bioinformatics* **2015**, *31*, 1592–1598.
51. Sass, S.; Pitea, A.; Mueller, N.S.; Theis, F.J. The miRlastic R package. Available online: <http://icb.helmholtz-muenchen.de/mirlastic> (accessed on 14 December 2015).
52. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **2012**, *489*, 519–525.
53. Kawamata, T.; Tomari, Y. Making RISC. *Trends Biochem. Sci.* **2010**, *35*, 368–376.
54. Robinson, M.; McCarthy, D.; Smyth, G. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2009**, *26*, 139–140.
55. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. B* **1995**, *57*, 289–300.
56. Baskerville, S.; Bartel, D. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **2005**, *11*, 241–247.
57. Guo, H.; Ingolia, N.; Weissman, J.; Bartel, D. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **2010**, *43*, 835–840.
58. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22.
59. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2012**, *40*, D109–D114.
60. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2008.
61. Dijkstra, E.W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1*, 269–271.
62. Csardi, G.; Nepusz, T. The igraph software package for complex network research. *Int. J. Complex Syst.* **2006**, 1–9.

63. Gross, A.; Orosco, R.; Shen, H.; Egloff, A.; Carter, H.; Hofree, M.; Choueiri, M.; Coffey, C.; Lippman, S.; Hayes, D.; *et al.* Multi-tiered genomic analysis of head and neck cancer ties TP53 mutation to 3p loss. *Nat. Genet.* **2014**, *46*, 1–7.
64. D'Souza, G.; Kreimer, A.; Viscidi, R.; Pawlita, M.; Fakhry, C.; Koch, W.; Westra, W.; Gillison, M. Case-control study of human papillomavirus and oropharyngeal cancer. *N. Engl. J. Med.* **2007**, *356*, 1944–1956.
65. Hui, A.; Lin, A.; Xu, W.; Waldron, L.; Perez-Ordóñez, B.; Weinreb, I.; Shi, W.; Bruce, J.; Huang, S.; O'Sullivan, B.; *et al.* Potentially prognostic miRNAs in HPV-associated oropharyngeal carcinoma. *Clin. Cancer Res.* **2013**, *19*, 2154–2162.
66. Gao, G.; Gay, H.; Chernock, R.; Zhang, T.; Luo, J.; Thorstad, W.; Lewis, J.J.; Wang, X. A microRNA expression signature for the prognosis of oropharyngeal squamous cell carcinoma. *Cancer* **2013**, *119*, 72–80.
67. Hui, A.; Lenarduzzi, M.; Krushel, T.; Waldron, L.; Pintilie, M.; Shi, W.; Perez-Ordóñez, B.; Jurisica, I.; O'Sullivan, B.; Waldron, J.; *et al.* Comprehensive MicroRNA profiling for head and neck squamous cell carcinomas. *Clin. Cancer Res.* **2010**, *16*, 1129–1140.
68. Minor, J.; Wang, X.; Zhang, F.; Song, J.; Jimeno, A.; Wang, X.; Lu, X.; Gross, N.; Kulesz-Martin, M.; Wang, D.; *et al.* Methylation of microRNA-9 is a specific and sensitive biomarker for oral and oropharyngeal squamous cell carcinomas. *Oral Oncol.* **2012**, *48*, 73–78.
69. Salazar, C.; Nagadia, R.; Pandit, P.; Cooper-White, J.; Banerjee, N.; Dimitrova, N.; Coman, W.; Punyadeera, C. A novel saliva-based microRNA biomarker panel to detect head and neck cancers. *Cell Oncol.* **2014**, *37*, 331–338.
70. Wald, A.; Hoskins, E.; Wells, S.; Ferris, R.; Khan, S. Alteration of microRNA profiles in squamous cell carcinoma of the head and neck cell lines by human papillomavirus. *Head Neck* **2011**, *33*, 504–512.
71. Sass, S.; Pitea, A.; Mueller, N.S.; Theis, F.J. MiRlastic on HNSCC dataset. Available online: <http://icb.helmholtz-muenchen.de/mirlastic/hnsc> (accessed on 14 December 2015).
72. Gilmore, T.D. Introduction to NF- κ B: Players, pathways, perspectives. *Oncogene* **2006**, *25*, 6680–6684.
73. Bancroft, C.; Chen, Z.; Yeh, J.; Sunwoo, J.; Yeh, N.; Jackson, S.; Jackson, C.; van Waes, C. Effects of pharmacologic antagonists of epidermal growth factor receptor, PI3K and MEK signal kinases on NF- κ B and AP-1 activation and IL-8 and VEGF expression in human head and neck squamous cell carcinoma lines. *Int. J. Cancer* **2002**, *99*, 538–548.
74. Ikeda, Y.; Tanji, E.; Makino, N.; Kawata, S.; Furukawa, T. MicroRNAs associated with mitogen-activated protein kinase in human pancreatic cancer. *Mol. Cancer Res.* **2012**, *10*, 259–269.
75. Lenarduzzi, M.; Hui, A.; Alajez, N.; Shi, W.; Williams, J.; Yue, S.; O'Sullivan, B.; Liu, F. MicroRNA-193b enhances tumor progression via down regulation of neurofibromin 1. *PLoS ONE* **2013**, *8*, e53765.
76. Mellman, I.; Yarden, Y. Endocytosis and cancer. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, doi:10.1101/cshperspect.a016949.
77. Hanahan, D.; Weinberg, R. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674.
78. Kelley, L.; Shahab, S.; Weed, S. Actin cytoskeletal mediators of motility and invasion amplified and overexpressed in head and neck cancer. *Clin. Exp. Metastasis* **2008**, *25*, 289–304.
79. Rietbergen, M.; de Kemp, S.M.; Bloemena, E.; Witte, B.; Brink, A.; de Jong, R.B.; Leemans, C.; Braakhuis, B.; Brakenhoff, R. Cancer stem cell enrichment marker CD98: A prognostic factor for survival in patients with human papillomavirus-positive oropharyngeal cancer. *Eur. J. Cancer* **2014**, *50*, 765–773.
80. Jouhi, L.; Datta, N.; Renkonen, S.; Atula, T.; Maekitie, A.; Haglund, C.; Ahmed, A.; Syrjaenen, S.; Grenman, R.; Auvinen, E.; *et al.* Expression of toll-like receptors in HPV-positive and HPV-negative oropharyngeal squamous cell carcinoma—an *in vivo* and *in vitro* study. *Tumour Biol.* **2015**, *36*, 7755–7764.
81. Amador-Molina, A.; Hernandez-Valencia, J.; Lamoyi, E.; Contreras-Paredes, A.; Lizano, M. Role of innate immunity against human papillomavirus (HPV) infections and effect of adjuvants in promoting specific immune response. *Viruses* **2013**, *5*, 2624–2642.
82. Rautava, J.; Syrjaenen, S. Biology of human papillomavirus infections in head and neck carcinogenesis. *Head Neck Pathol.* **2012**, *6*, 3–15.
83. Pries, R.; Wollenberg, B. Cytokines in head and neck cancer. *Cytokine Growth Factor Rev.* **2006**, *17*, 141–146.

84. Kota, J.; Chivukula, R.R.; O'Donnell, K.A.; Wentzel, E.A.; Montgomery, C.L.; Hwang, H.W.; Chang, T.C.; Vivekanandan, P.; Torbenson, M.; Clark, K.R.; *et al.* Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell* **2009**, *137*, 1005–1017.
85. Cheng, C.J.; Bahal, R.; Babar, I.A.; Pincus, Z.; Barrera, F.; Liu, C.; Svoronos, A.; Braddock, D.T.; Glazer, P.M.; Engelman, D.M.; *et al.* MicroRNA silencing for cancer therapy targeted to the tumour microenvironment. *Nature* **2015**, *518*, 107–110.



© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).