Human against Machine? Machine Learning Identifies MicroRNA Ratios as Biomarkers for Melanoma

Melody H. Shellman¹ and Yiqun G. Shellman²,³

Identification of quantitative molecular biomarkers to distinguish melanoma from nevi is highly desirable. Expressions of microRNAs (miRNAs) are promising candidates but lack consensus in many studies. Torres et al. (2020) utilized a machine learning pipeline to identify miRNA ratios as strong biomarkers. Results indicate that machine learning, although powerful, requires human input to identify high quality biomarker signatures.

Although microRNAs (miRNAs) are promising candidates for biomarkers to distinguish melanoma from nevi, many challenges exist in the field, especially the lack of reproducibility among the studies (Elmore et al., 2017; Jayawardana et al., 2016; Witwer and Halushka, 2016). Torres et al. (2020) attempted to address this issue with machine learning. The significance of the authors’ work is two-fold. First, they identified two genetic and clinical confounding factors contributing to the variations for detecting miRNAs in formalin-fixed paraffin-embedded (FFPE) biopsies. Second, they identified a promising set of high sensitivity and specificity miRNA ratios to classify melanoma versus nevi, independent of the confounding factors.

Tumor cellularity and patient age are confounding factors for miRNA detection in FFPE samples

miRNAs are promising candidates as molecular biomarkers for diseases, because they are relatively small and easy to detect at a low cost, even in small samples such as skin biopsies. In addition, miRNAs are more stable than mRNAs in FFPE samples. Consequently, many studies have explored miRNAs as potential biomarkers to distinguish melanoma from nevi. However, there is a high level of discrepancy among these studies (Elmore et al., 2017; Jayawardana et al., 2016; Witwer and Halushka, 2016). In a meta-analysis, including seven publicly available miRNA datasets, Torres et al. (2020) found that no single miRNA was differentially expressed across all datasets, and only four miRNAs were differentially expressed when limiting the analysis to four of the seven datasets. Thus, there is a substantial discrepancy in miRNA signatures for melanoma versus nevi, which needs to be addressed before any miRNA signature will be valuable for clinical use.

The issue of high discordance is known among studies of miRNAs used as biomarkers (Elmore et al., 2017; Jayawardana et al., 2016; Witwer and Halushka, 2016); however, the cause is not well-defined. Discrepancies have been attributed to multiple confounding variables, including sample-to-sample variation of FFPE biopsies, with differing amounts of contamination from nontumor cells, as well as platform-to-platform variation of quantification for miRNA expression (Witwer and Halushka, 2016). By utilizing machine learning, Torres et al. (2020) first aimed to find genetic or clinical confounding variables that affect the quantification of miRNA expression in FFPE samples of melanoma and nevi. Torres et al. (2020) established a well-curated dataset by meticulously assembling and annotating an initial training cohort of microdissected primary melanomas with their adjacent nevi. Each sample was annotated with miRNA expression data, diagnosis, genetic information (tumor cellularity, mutation burden, and copy number variation), and clinical features (patient age, sex, and anatomical location of lesion). This dataset provided well-controlled materials for the study, and the authors identified two confounding variables: tumor cellularity and patient age (to a lesser extent). Tumor cellularity is the relative proportion of tumor cell, normal cell, and other cell types in a sample, and was estimated by allele frequencies and magnitudes of copy number changes. Results thus indicated that any high quality miRNA biomarkers should be independent of these variables in the samples.

The ratios, not expressions, of miRNAs are better biomarkers to distinguish melanoma from nevi

Tumor cellularity is influenced by contaminating nontumor cells in the samples of skin biopsies, which can vary dramatically depending on sample size, histologic type (predominantly junctional vs intradermal), and preparation (precision of microdissection). The most common miRNA biomarkers are disease-associated miRNA expression. However, Torres et al. (2020) found that quantification of miRNA expression in FFPE skin samples was significantly affected by tumor cellularity, undermining the use of total miRNA expression alone as biomarkers for melanoma versus nevi.

miRNA ratios have been used successfully in other cancers to improve prediction accuracy (Avissar et al., 2009; Reddy et al., 2015), thus Torres et al. (2020) explored using miRNA ratios to distinguish melanoma from nevi. The ratios were derived from...
expressions of miRNAs that were either enriched (ME-miRNAs) or depleted (MD-miRNAs) in melanoma compared with nevi. These miRNAs were identified by comparing the matched pair of microdissected primary melanomas with their adjacent nevi. Consequently, MD-miRNAs were more likely to be derived from melanocytes, the cell origin of melanoma, and less likely to be from infiltrating or adjacent non-melanocytic cells. Therefore, the ratios of ME to MD may normalize the amount of input RNA of melanocytic origin and be less influenced by changes in biopsy composition. Although similar, this normalization technique is different from standard normalization to total RNA or a housekeeping gene; ME to MD ratios normalize the fraction of input RNA that originated from melanocytic cells, thus highlighting the malignant content in each sample. Indeed, the model of ME to MD-miRNA ratios has a far better predictive value than the model of miRNA expression alone, in both the training dataset and the external validation cohort. This model also classified all available published cohorts with an area under the receiver operating characteristic (ROC) curve of 0.98 and was independent of tumor cellularity or patient age (Torres et al., 2020). These results suggest that strategic usage of the ratios, instead of total expression as a normalization, may provide a way to control for malignant content of samples. This strategy may be used for other expression biomarkers as internal control standardization, including mRNAs or proteins.

### Clinical Implications
- Tumor cellularity and patient age affect microRNA quantification in formalin-fixed paraffin-embedded samples.
- MicroRNA expression ratios more accurately classify melanoma and nevi than expression alone.
- Machine learning can be used to successfully identify biomarkers for melanomas.

**Human against machine? Using machine learning to identify biomarkers**

Machine learning has a growing presence in biology, particularly in the diagnosis of various diseases. It is a branch of artificial intelligence (AI) that enables a system to learn from data without explicitly programmed commands; pattern recognition and inference provide the basis for such algorithms. Compared with humans, machine learning has the potential to more accurately identify a diseased sample.

---

**Figure 1. Scheme of machine learning to identify miRNA signature to distinguish melanoma vs nevi.**

(a) Candidate Identification (feature selection). This is an important step to identify variables for training the machine learning algorithm. (b) MiRNA Expression Trained Model. Expression levels from selected miRNAs (a) were used to train a machine learning melanoma versus nevus classification model, which then was tested on an external validation dataset. The model yielded an ROC of 0.67 and sensitivity and specificity of 1.00 and 0.34. (c) MiRTM. ME to MD ratios of the selected miRNAs (a) were used to train a machine learning melanoma vs nevus classification model, which was tested on the same external validation dataset as (b). The model yielded an ROC of 0.90 and sensitivity and specificity of 0.81 and 0.88. The ROC graphs in (b) and (c) are from (Torres et al., 2020). MD, melanoma depleted; ME, melanoma enriched; miRNA, microRNA; MiRTM, MiRNA Ratio Trained Model; ROC, receiver operating characteristic.
state, but the algorithm can only do so if the training data is well-curated and cleaned of trivial variance. In other words, “garbage in, garbage out.” Unfortunately, biological samples are often imperfect, and it is challenging to obtain a large dataset. Therefore, it is important to have a high quality and well-curated training set for biological studies, which was a strength of Torres and colleagues’ study (2020). Their initial feature selection dataset contained 15 matched pairs of microdissected primary melanomas with their adjacent nevi, and each sample was annotated with genetic and clinical information, thus allowing the authors to extract useful information from a small dataset.

Machine learning has been explored previously to screen for disease-associated miRNA biomarkers, although with miRNA expression only (Adam et al., 2013; Chen et al., 2018). In the field of melanoma diagnosis, algorithms have been developed to facilitate melanoma detection from a dataset of 300 dermoscopic images (Haenssle et al., 2018), a significantly larger sample size than most datasets of biological samples. Yet, no studies have utilized machine learning to screen for quantitative molecular markers to classify melanoma from nevi.

In the manuscript from Torres et al. (2020), machine learning is applied in two steps (Figure 1): feature selection (candidate identification, Figure 1a) and melanoma versus nevi classification models (Figure 1b and c). Importantly, both steps require human input to make the machine learning feasible within the dataset available. Feature selection is an important step to identify variables for training the machine learning algorithm. A dataset of expression across all measured miRNA represents far too many variables for practical model building. Most of these variables are irrelevant to classification, and many may not have explainable relevance in advance. Additionally, many machine learning algorithms exhibit a decrease of accuracy with a large number of variables because of overtraining of the model (Kursa and Rudnicki, 2010). Torres et al. (2020) used the Boruta R package algorithm to reduce expression across all miRNAs to a smaller subset. In order to reduce false positive identification, an artificial feature (shadow miRNAs) generated through randomized read counts across samples represented a negative control for each unique miRNA. The Boruta algorithm was then applied to the miRNAs paired with the shadow miRNAs. This process yielded six differentially expressed miRNAs for the focus of the machine learning classification models, two ME-miRNAs and four MD-miRNAs.

For the melanoma versus nevi classification models, two different machine learning classifiers were trained (Figure 1b and c) (Torres et al., 2020). The MiRNA Expression Trained Model was trained with individual expression values of the six miRNAs (Figure 1b), whereas the MiRNA Ratio Trained Model (MiRTM) was trained with the eight expression ratios of the same miRNAs (Figure 1c). From the two ME-miRNAs and four MD-miRNAs, eight miRNA ratios were created from all possible combinations of ME-miRNA to MD-miRNA. Each classifier was cross-validated with the aggregated public datasets. The expression classifier resulted in an area under the ROC curve of 0.95. The MiRTM resulted in an area under the ROC curve of 0.98. However, when validated with an external dataset with various degrees of tumor cellularity, the miRNA expression trained model dropped drastically in classification accuracy, resulting in an area under the ROC curve of 0.67 (Figure 1b). The sensitivity and specificity of the test were 1.00 and 0.34, respectively. The MiRTM on the other hand, maintained similar results as the aggregated datasets, resulting in an area under the ROC curve of 0.90 and sensitivity and specificity of 0.81 and 0.88 (Figure 1c).

The success of the MiRTM over the expression trained model demonstrates that human design of the training strategy can help improve the prediction accuracy of machine learning algorithms. Human decisions based off deep understanding of biological samples drive these intentional data manipulations. Regarding the presence of machine learning in biology, the power of utilizing these algorithms has already been demonstrated. For the present, however, human selection of training data is critical to realize the power of AI. It is not human against machine, but rather human guides machine for the foreseeable future.

CONFLICT OF INTEREST
The authors state no conflict of interest.

REFERENCES