Combined Amplification and Molecular Classification for Gene Expression Diagnostics

Gokul Gowri[1], Randolph Lopez[2], and Georg Seelig[2(✉)]

[1] California Institute of Technology, Pasadena, CA, USA
[2] University of Washington, Seattle, WA, USA
gseelig@uw.edu

Abstract. RNA expression profiles contain information about the state of a cell and specific gene expression changes are often associated with disease. Classification of blood or similar samples based on RNA expression can thus be a powerful method for disease diagnosis. However, basing diagnostic decisions on RNA expression remains impractical for most clinical applications because it requires costly and slow gene expression profiling based on microarrays or next generation sequencing followed by often complex *in silico* analysis. DNA-based molecular classifiers that perform a computation over RNA inputs and summarize a diagnostic result *in situ* have been developed to address this issue, but lack the sensitivity required for use with actual biological samples. To address this limitation, we here propose a DNA-based classification system that takes advantage of PCR-based amplification for increased sensitivity. In our initial scheme, the importance of a transcript for a diagnostic decision is proportional to the number of molecular probes bound to that transcript. Although probe concentration is similar to that of the RNA input, subsequent amplification of the probes with PCR can dramatically increase the sensitivity of the assay. However, even slight biases in PCR efficiency can distort weight information encoded by the original probe set. To address this concern, we developed and mathematically analyzed multiple strategies for mitigating the bias associated with PCR-based amplification. We evaluate these amplified molecular classification strategies through simulation using two distinct gene expression data sets and associated disease categories as inputs. Through this analysis, we arrive at a novel molecular classifier framework that naturally accommodates PCR bias and also uses a smaller number of molecular probes than required in the initial, naive implementation.

1 Introduction

Detection and quantification of RNA molecules in blood or tissue can be a powerful tool for disease diagnosis. Although detection of a single, differentially expressed molecular marker might be ideal, such a distinctive marker may not exist for a disease state of interest. Instead, a diagnostic decision may have to be

C. Thachuk and Y. Liu (Eds.): DNA 25, LNCS 11648, pp. 159–173, 2019.
https://doi.org/10.1007/978-3-030-26807-7_9

based on panels of differentially expressed genes and may require careful weighing of the contributions of each gene in the panel. The conventional workflow for building such multi-gene classifiers consists of experimentally measuring gene expression in different "states" (e.g. healthy and disease) using microarrays or high-throughput sequencing, followed by training of a computational classifier that learns to assign labels to the samples based on differences in gene expression between those states. Once the classifier is trained it can be used to label previously unseen samples based on their gene expression profiles, thus providing a powerful tool for aiding diagnostics decisions. For example, an *in silico* whole blood gene expression classifier has been developed to distinguish bacterial infections, viral infections, and non-infectious disease [1]. Similarly, an *in silico* blood platelet gene expression classifier has been developed to distinguish six different types of cancer [2]. Both of these classifiers were trained using support vector machine (SVM) methods, and each involved more than 100 features (i.e. differentially expressed genes) [1,2]. These examples provide proof-of-principle for the power of a diagnostic approach based on analysing multi-gene panels but the complexity and cost of gene expression classification has limited its implementation to a research setting and to a small subset of clinical problems.

Using molecular computation to implement a disease classifier could overcome some of the limitations of existing technologies. In such an approach, the molecular diagnostic "computer" could be mixed with the patient sample, perform an analysis of the sample through a sequence of molecular interactions and then summarize the result into an easy-to-interpret signal that represents the diagnosis. If realized, such an approach could accelerate the diagnosis by simplifying the measurement process and eliminating the need for separate computational analysis.

DNA-based circuits have long been proposed as a platform for performing gene expression diagnostics [3]. In our own previous work, we demonstrated a platform to translate a linear classifier into a DNA strand displacement circuit that can operate on a set of RNA inputs and classify a sample by calculating a weighted sum over those inputs [4]. However, our work was limited to a proof-of-principle implementation with *in vitro* transcribed RNA since the concentration of RNA molecules in biological samples is significantly below the detection limit of that assay.

In this work, we propose a method for combining targeted RNA amplification and molecular classification for gene expression diagnostics. We begin by proposing a classifier design that combines computation with enzymatic amplification (Sect. 2). In our approach, molecular probes are used to assign weights to each RNA feature in the classifier and PCR-based amplification is used to amplify the signal resulting from each probe. The workflow for such an approach is shown in Fig. 1. We then demonstrate a computational approach, starting from gene expression data, for learning a sparsely-featured classifier that is well-suited for molecular implementation (Sect. 3). However, this formal framework assumes that molecular classifier components behave ideally, while in practice nucleic acid amplification mechanisms are biased, i.e. different probe

sequences get amplified to varying degrees which distorts the classification result (Sect. 4). We then explore the effects of amplification bias and developed strategies to mitigate amplification bias and maintain classification accuracy. First, we explore a strategy for averaging over multiple probes, each with an unknown bias (Sect. 5). Although successful to an extent, we will argue that this naive strategy is impractical because of the large number of probes required. Moreover, because in practice the amplification bias is consistent across samples, it is possible to actually measure the probe bias and incorporate that information into the classifier design. Second, we thus ask whether accurate classification can be achieved if individual probes are biased but well-characterized (Sect. 6). We will show that this approach results in a more compact molecular implementation of the classifier but that classification accuracy is still limited compared to an ideal classifier. Finally, we develop an approach wherein each weight associated with a transcript is implemented by just two competing probes, one associated with a negative and the other with a positive weight (Sect. 7). We show that even with biased probes any target value of the weight can be closely approximated as long as we have precise control over probe concentrations. We validate each of our designs by simulating molecular classification of gene expression profiles in the context of cancer and infectious disease diagnostics. Our results indicate that probe characterization and subsequent selection enable the construction of sensitive, accurate and cost-effective molecular classifiers for gene expression diagnostics.

2 A Molecular Classifier with Built-In Amplification

Here, we lay out a strategy for creating a molecular classifier with a built-in amplification mechanism. We first review our earlier classifier design that provides that basis for our work and then discuss how an enzymatic amplification step can be used to dramatically increase the specificity of the earlier approach.

In the molecular classifiers reported in [4], the magnitude of a weight associated with a transcript is implemented by the number of probes designed to target that transcript; an identity domain appended to the hybridization domain indicates the sign of a each probe and thus of each weight. The total concentration of all positive (negative) identity domains thus represents the sum of all positive (negative) weights in the system. Identity domains interact with a downstream DNA circuit that performs subtraction through pairwise annihilation of positive and negative signal [4,5]. The remaining signal is converted to fluorescent output. Alternatively, if reporter constructs labeled with distinct fluorophores are used to separately read out the concentrations of positive and negative identity domains, the subtraction can be performed in silico. In either case, this approach requires weights to be small (<10) integers, because each unit weight requires a unique binding site on a transcript. This design is demonstrated in Fig. 2. Moreover, in this framework, the output signal has similar magnitude to the input signal, so sensitivity is low, and not sufficient for clinical disease diagnostics.

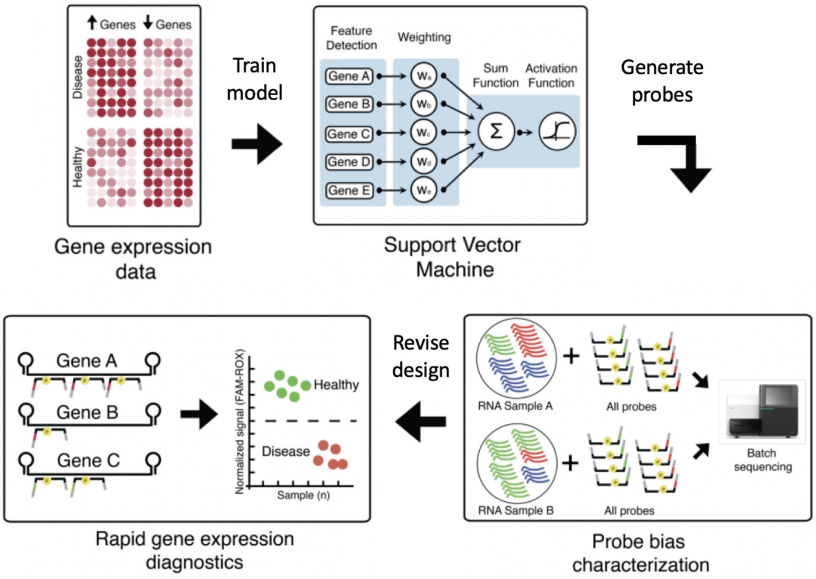


Fig. 1. Workflow for combined amplification and classification of RNA samples. An in-silico classifier is trained and validated on publicly available gene expression data. Then, multiple probes are designed and synthesized to target the set of genes in the classifier. This set of probes is amplified using an RNA sample with known gene expression data. Using next-generation sequencing, the number of amplified probes can be counted in batch in order to determine probe specific amplification bias. Subsequently, this data informs a design for molecular classifier that will implement the desired classification model.

Increasing Sensitivity. Amplification reactions based on strand displacement cascades [6–9] provide one interesting approach to signal amplification and such systems have even been used to create amplifiers with controllable gain that can be combined into linear classifier circuits [10,11]. Still, the gain that can be achieved with a single strand displacement-based catalytic amplifier is typically limited and cascades of multiple amplifiers that might have higher gain are often leaky or simply not robust enough for practical applications.

In order to address the lack of sensitivity intrinsic to our initial classification scheme we thus compared several enzyme-based approaches for targeted RNA amplification including NASBA [14], rolling-circle amplification [15,17,18], RASL [12] and multiplex PCR [19]. Upon evaluating and characterizing several of these methods, we decided to implement a molecular classification workflow using RASL probes. RASL (RNA-mediated oligonucleotide annealing, selection, and ligation) is commonly used for targeted RNA sequencing for the purpose of gene expression quantification [12].

Probe Amplification with RASL. In the RASL protocol, shown in Fig. 3a, custom probe pairs are designed for each RNA of interest. The two probes of

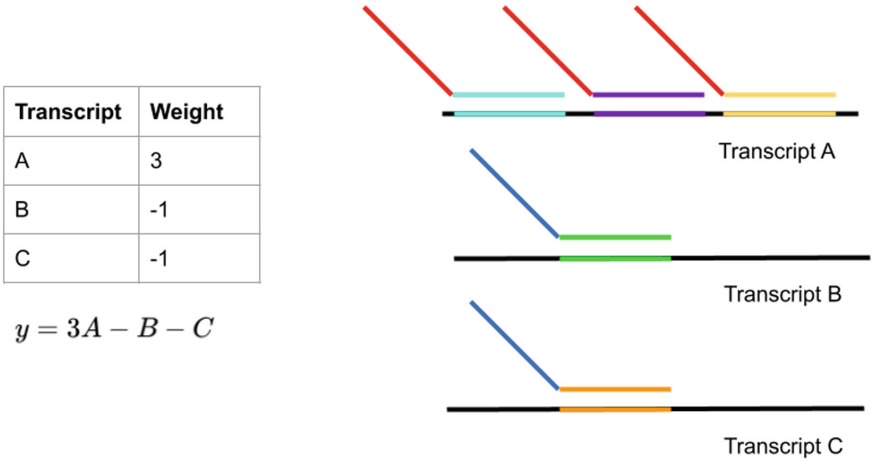


Fig. 2. An example molecular implementation of a linear classifier with the approach previously demonstrated in [4]. The red and blue domains represent positive and negative values respectively for downstream subtraction and reporting mechanisms. (Color figure online)

a pair bind neighboring sites on an mRNA target such that they create a contiguous sequence with a nick at the center. Adjacent probes are then ligated together. Importantly, ligation can only occur if probes are bound to the RNA. Each probe also contains universal amplification overhangs. PCR amplification of the ligated probes enables accurate amplification of low total RNA amounts (10 nanograms) for gene expression quantification using next-generation sequencing. Because only ligated probes can be efficiently amplified and because ligation requires the RNA template, RASL is specific and quantitative method for RNA detection.

A molecular classifier using RASL probes is shown in Fig. 3b. Different integer weights can be implemented by varying the number of probes per transcript. To implement positive and negative weights, each probe also contains a sequence barcode that triggers either a green or red fluorescent reporter during amplification. The difference between the aggregate fluorescent signals from the two channels is the classification output.

Amplification by RASL is consistent across samples, but is inconsistent within samples, such that certain probes are amplified with greater propensity than others [12]. As a result, n probes will not necessarily yield n times as large a signal as a single probe representing unit weight, so molecular classifiers designed with the approach of [4] will not behave similarly to the *in silico* classifier upon amplification. Below we discuss how such bias can be accounted for in the classifier design.

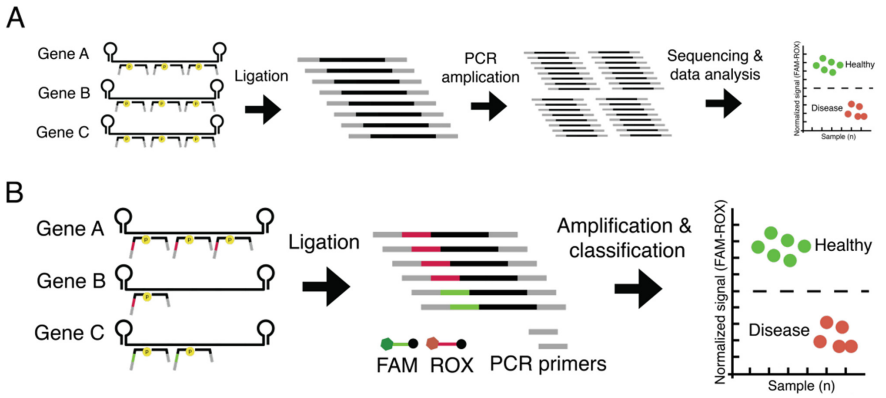


Fig. 3. Overview of existing RASL-seq and our modified method for molecular classification. In panel A, the standard RASL-seq protocol for targeted gene expression counting using next-generation sequencing is shown. Multiple probe pairs are hybridized to their RNA targets followed by ligation of gaps. The yellow circled P's indicate the 5' phosphates required for ligation. The ligated products can be amplified using common primers and then used as input for next generation sequencing. Sequencing data is then used for gene expression analysis. In panel B, the protocol for molecular classification using RASL is shown. Our modified RASL probes contain a positive or negative barcode (red or green domain) associated with each probe. The number of probes that bind to each transcript can be varied to tune the effective amplification weight on each transcript. During amplification, two fluorescent reporters (FAM and ROX) are triggered based on the presence of each barcode. This fluorescent signal can then determine the classification outcome for a given sample. (Color figure online)

3 Computational Design of Sparsely Featured Diagnostic Classifiers

Before building a molecular classifier for gene expression diagnostics, we first train an *in silico* disease classifier on available gene expression data. This computational classifier is then mapped onto a molecular classifier. In a molecular classifier, each feature corresponds to a target transcript, and probes are implemented as oligonucleotides that can bind to a unique region within a transcript. Although feasible in principle, molecular classifiers with large numbers of features are currently not practical because of the large numbers of probes required to implement them. As s first step in the classifier design workflow we thus aim to create sparsely featured yet clinically valuable classifiers.

Using the publicly available data sets associated with [1] and [2], we identified reliable classifiers that have feasible molecular implementations by limiting the scope of the classification problems and training support vector machines with high feature selection penalties. Regularization parameters were determined through iterative search, and a hard sparsity constraint of a maximum of 50 features was used. In particular, we identified a sparsely-featured classifier that

differentiates between bacterial infections and viral infections from whole blood RNA samples, similar to that previously implemented as a molecular classifier for *in vitro* use [4]. We also created a sparsely-featured classifier that discriminates between cancer and healthy patients based on platelet RNA samples. Both classifiers were tested on validation sets that were disjoint from the training sets, and consisted of approximately $\frac{1}{6}$ the number of samples of the training sets, corresponding to validation sets of 35 and 56 samples respectively. The accuracy of the classifiers on the validation set will henceforth be referred to as model accuracy, and are shown in Table 1.

4 Simulating the Accuracy of a Molecular Classifier

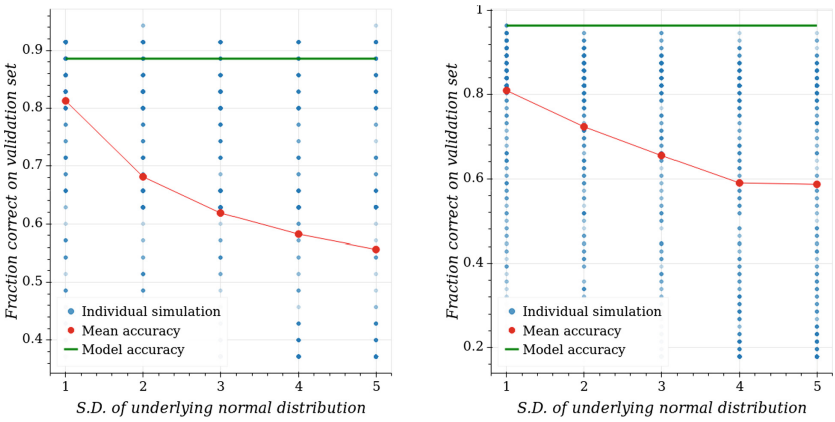
The sequence dependent amplification bias in RASL is not well characterized, but based on the evaluation of probe bias in [12], we will roughly approximate it as a $\log_2$-normal distribution. The largest difference in amplification factor of the 25 probes tested in [12] is approximately 2^6, so we approximate the underlying normal distribution of probe bias to have $\mu = 0, \sigma = 3$.

We estimate the accuracy of a molecular implementation by generating random amplification biases according to the previously described distribution for each probe, then summing the biases of the probes for each target transcript in order to determine an effective weight for each feature. The resulting set of weights is then evaluated on the validation set to obtain an accuracy that can be compared to the model accuracy. Several such simulations are executed to determine an expected accuracy of a molecular implementation. The simulated accuracy of molecular classifiers implemented with the approach of [4] are shown in Table 1.

As mentioned before, the parameters of the distribution of probe bias are rough estimates. In Fig. 4, we see the effect of the standard deviation of the underlying normal distribution of probe bias on the simulated accuracy. This indicates that unless we have $\sigma < 1$, the accuracy of the molecular implementation of a classifier deviates highly from the model accuracy. Based on the data presented in [12], it seems unreasonable to expect that variation in amplification bias is so small. As such, a method of mitigating the effect of amplification bias is necessary.

Table 1. Properties of the sparsely-featured disease diagnostic classifiers. Basic molecular classifier accuracy is the estimated accuracy of the molecular classifiers implemented with amplification and without bias mitigation, averaged over 500 simulations.

Task	# of features	*In silico* model accuracy	Simulated basic molecular classifier accuracy
Viral vs. bacterial	5	0.89	0.59
Cancer diagnosis	24	0.96	0.60



(a) Bacterial vs. viral classifier (b) Cancer diagnosis classifier

Fig. 4. Effect of variation in amplification bias of DNA probes on simulated accuracy of molecular classifiers implemented with no amplification bias mitigation. σ is the standard deviation of the normal distribution underlying the $\log_2$-normal distribution of probe bias, with a mean $\mu = 0$. For each σ, 500 independent simulations were done, and the shade of blue indicates the frequency of an outcome. (Color figure online)

5 Scaling up Weights to Reduce Bias Variation

The ideal method of mitigating amplification bias would involve no experimental procedures beyond those previously characterized in molecular gene expression classifiers. One such approach is to simply increase the number of probes used to implement a unit weight, such that a weight of magnitude k would be implemented by $c * k$ probes, where c is constant across all weights, as depicted in Fig. 5. This is equivalent to scaling up all weights by a constant factor. Intuitively, this will reduce the effect of probe bias because the average of c amplification factors should be less variable than the amplification factor of a single probe. The desired effect of this approach can be seen in the following mathematical characterization:

For a molecular classifier with a set of target transcripts $\mathcal{T}$, let

$$\mathcal{P}_t = \{p_1^t, p_2^t, ...p_{c*|t|}^t\}$$

be the set of amplification factors of the probes used to implement desired weight $|t|$ of a transcript $t \in \mathcal{T}$, where each p_i^t represents the amplification factor of the i-th probe targeting transcript t. We will define the effective weight of a target transcript t as

$$\mathcal{W}_t = \sum_{i=1}^{c*|t|} p_i^t$$

The effective weight is the weight implemented by the molecular classifier after amplification. Using the approximation that amplification factors are distributed

Transcript	Weight
A	3
B	-1
C	-1

$$y = 3A - B - C$$

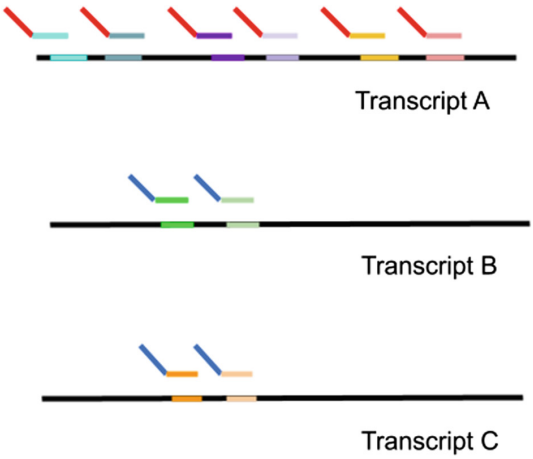


Fig. 5. Molecular implementation of bias mitigation through scaling up weights, with scaling factor $c = 2$ for a simple toy classifier. Probes are not drawn to scale, and represent the RASL-seq probes shown in Fig. 3.

as $p_i^t \sim 2^{\mathcal{N}(\mu,\sigma)}$, we see that for large c,

$$\mathcal{W}_t \approx c * |t| * 2^{\mu}$$

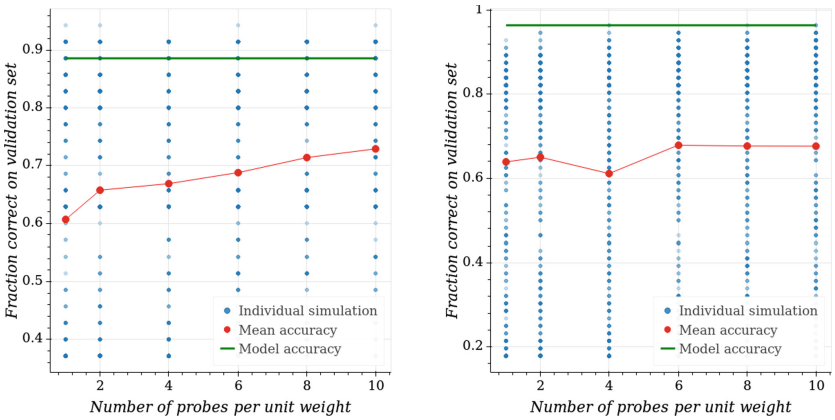
In order to implement the desired classifier, the effective weights do not need to be the same as the desired weights – constantly scaled weights will yield identical classification results. As a result, if the effective weights have the same relative values as the desired weights, the implementation can be considered successful. We see that for any pair of transcripts $u, v \in \mathcal{T}$,

$$\frac{\mathcal{W}_u}{\mathcal{W}_v} \approx \frac{c * |u| * 2^n}{c * |v| * 2^n} = \frac{|u|}{|v|}$$

so the classifier will implement the desired weights, regardless of individual amplification biases.

However, c cannot be arbitrarily large, as it is limited by the number of unique binding sites per transcript and the cost of oligonucleotide synthesis. Instead, we must determine if an experimentally feasible c will yield a sufficient approximation for gene expression classification.

In Fig. 6, we see the effect of c on the simulated accuracy of a molecular classifier. The plots indicate that for even $c = 10$, corresponding to up to 50 distinct probes per transcript, there is still a large disparity between the predicted implementation accuracy and the model accuracy. Based on these results, it seems that knowledge of the amplification biases of the probes is necessary in order to construct a molecular classifier.



(a) Bacterial vs. viral classifier (b) Cancer diagnostic classifier

Fig. 6. Simulated accuracy of molecular classifiers using c probes to implement a unit weight. The shade of blue indicates the frequency of an outcome. 500 independent simulations were done for each c. (Color figure online)

6 Classifier Calibration by Screening Probes

A simple extension of the approach to implementing molecular classifiers presented in [4] is to measure the amplification biases of a set of probes for each target transcript, then select a subset of these probes that will yield effective weights that best approximate the desired weights, as shown in Fig. 7. This approach can be mathematically characterized similarly to the approach described in Sect. 5:

For a molecular classifier with a set of target transcripts $\mathcal{T}$, let

$$\mathcal{P}_t = \{p_1^t, p_2^t, ...p_n^t\}$$

be the set of amplification factors of n probes targeting a transcript $t \in \mathcal{T}$ with desired weight $|t|$, where p_i^t represents the amplification factor of the i'th probe targeting transcript t. Let $\wp(\mathcal{P}_t)$ be the power set of $\mathcal{P}_t$, representing the set of all weights that could be implemented using the given set of probes. The effective weight of a subset of probes $\mathcal{S}_t \in \wp(\mathcal{P}_t)$ is

$$\mathcal{W}_t = \sum_{p_i^t \in \mathcal{S}_t} p_i^t$$

Our goal is to select subsets of probes $\mathcal{S}_t$ for all transcripts $t \in \mathcal{T}$ that yield

$$min\left(\sum_{u,v \in \mathcal{T}} \left(\frac{\mathcal{W}_u}{\mathcal{W}_v} - \frac{|u|}{|v|} \right)^2 \right)$$

This minimization problem requires us to search $|\wp(\mathcal{P}_t)|^{|\mathcal{T}|} = 2^{n*|\mathcal{T}|}$ possible solutions. For the bacterial vs. viral classifier, with an experimentally reasonable 10 probes per target transcript, there would be a massive 2^{50} possible sets

Transcript	Weight
A	3
B	-1
C	-1

$$y = 3A - B - C$$

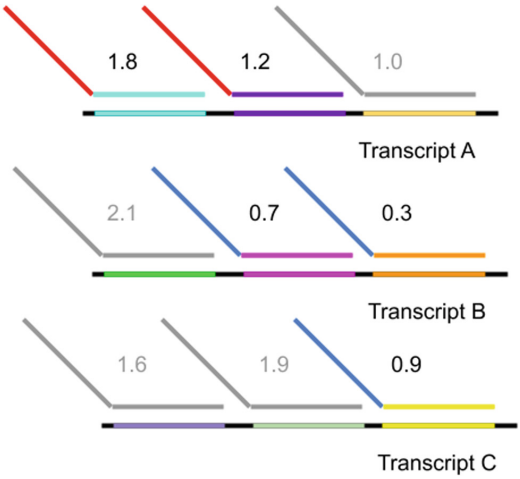
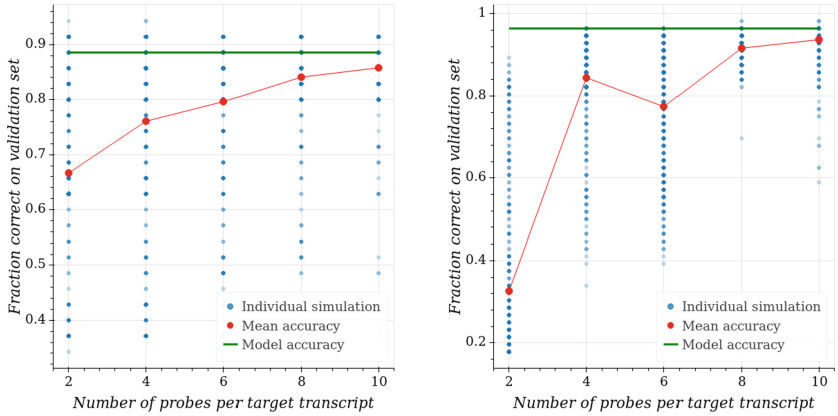


Fig. 7. Molecular implementation of classifier built by selecting optimal subset of probes, with $n = 3$ probes per transcript, for a simple toy classifier, with probe biases (denoted by the number next to the probe) selected for demonstration purposes. The grey probes are omitted in the classifier implementation, and the colored probes are the subset that best approximates the desired weight. Probes are not drawn to scale, and represent the RASL-seq probes shown in Fig. 3. (Color figure online)

of weights, so it seems likely that the optimal selection of probes would yield desired classifier behavior. However, the large solution space also makes identifying the solution challenging. In order to estimate the accuracy of this approach analogously to Fig. 6, with 500 simulations of randomly biased probes, it is not feasible to search the entire solution space. Instead, for each transcript, we simply find the subset of probes whose sum is closest to the weight normalized to $n * \mu$, where μ is the mean of the $\log_2$-normal distribution. In other words, we select subsets of probes $\mathcal{S}_t$ for all transcripts $t \in \mathcal{T}$ that yield

$$min\left(\sum_{t \in \mathcal{T}} (\mathcal{W}_t - |t| * \mu)^2\right)$$

This minimization problem requires us to search $|\wp(\mathcal{P}_t)| * |\mathcal{T}| = |\mathcal{T}| * 2^n$ possible solutions, which makes the search much more feasible for simulation purposes. The accuracy of the classifier built with this selection of probes gives us a reasonable lower bound on the accuracy of the optimal subset, which could feasibly be identified when implementing a disease diagnostic classifier. These estimated lower bounds of accuracy are shown for different n in Fig. 8. The lower bound of expected molecular classifier accuracy begins to approach the model accuracy at $n = 10$, which is experimentally feasible.



(a) Bacterial vs. viral classifier (b) Cancer diagnostic classifier

Fig. 8. Simulated accuracies of molecular classifiers using subsets of n probes per transcript. The shade of blue indicates the frequency of an outcome. 500 independent simulations were done for each n. (Color figure online)

7 Encoding Weights in Probe Concentration

In the previously described methods of implementing molecular classifiers, in order to change the effective weights of the classifier, one must use a different set of probes to construct the classifier. It may be possible to instead encode classifier weights in probe concentration, similar to previous DNA-based neural network constructions [13]. Because disease diagnostic molecular classifiers must accept analog input, and we wish to minimize the number of strand displacement reactions in the system, the mechanism of weight multiplication should be different from previous DNA-based neural network constructions. An implementation of analog multiplication by DNA strand displacement has been proposed [16], but the design involves a complex reaction network that would likely be difficult to experimentally construct for the purposes of *in situ* molecular classification.

We have designed a method of implementing arbitrary classifier weights with a constant number of probes per transcript, by encoding classifier weights in the concentration of the probes. For each transcript, we select exactly one binding domain. For each binding domain, we design one positive probe and one negative probe. When constructing the classifier, we add all probes in excess, while controlling the ratios of each pair of positive and negative probes. The difference of the relative concentrations of the probes will encode the weight of the transcript. If a positive probe has the same concentration as its negative counterpart, then the weight of their target transcript should be 0. Similarly, a larger concentration of positive probe will encode a positive weight, and a larger concentration of negative probe will encode a negative weight. This weight implementation strategy is depicted in Fig. 9.

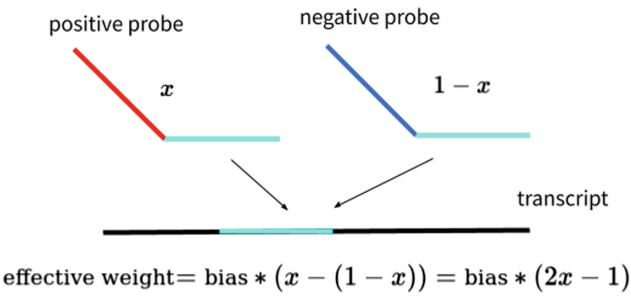
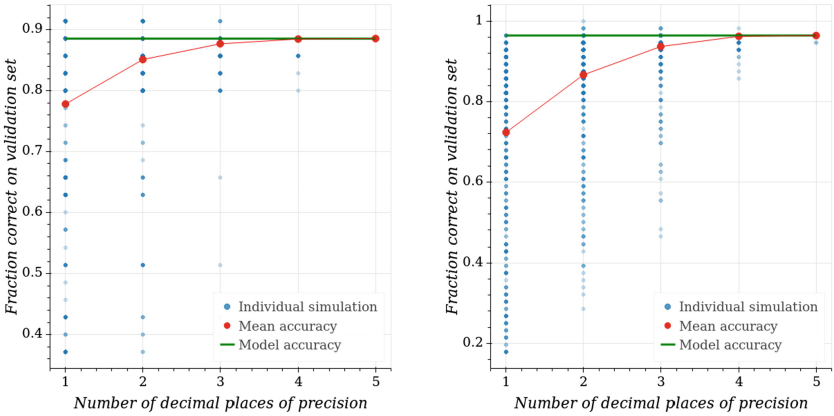


Fig. 9. Cartoon depiction of competitive hybridization based weight implementation strategy. The arrows indicate complementary domains. Probes are not drawn to scale, and represent the RASL-seq probes shown in Fig. 3. $x = \frac{|\text{positive probe}|}{|\text{positive probe}|+|\text{negative probe}|}$, the concentration of the positive probe normalized to the sum of the concentrations of the positive and negative probe.



(a) Bacterial vs. viral classifier (b) Cancer diagnostic classifier

Fig. 10. Simulated accuracies for molecular classifiers built with weights encoded in probe concentration. Decimals of precision indicates the control of the ratios. For example, 1 decimal of precision means that only ratios with increments of ten percent can be implemented.

Implementing weights in this fashion would allow for manipulation of classifier weights without changing the probes used. One could measure probe biases, then adjust the concentrations such that the effective weights are identical to the weights of the *in silico* model. Since the effective weight is simply a linear function of the relative concentrations, the ability to implement arbitrary weights is dependent on the ability to arbitrarily control the concentrations of probes. The simulated accuracy of molecular classifiers built with this weight implementation strategy after bias mitigation are shown in Fig. 10 for different degrees of concentration precision. We see that if ratios can be controlled to precision

10^{-3} (i.e. being able to implement concentrations in ratio 0.501:0.499), then we expect the molecular classifiers to approach the model accuracy. This seems experimentally feasible, but in the case that such precision cannot be attained, a larger number of probes per transcript can be characterized, and the probes with the most similar biases can be selected to decrease the minimum difference in concentration.

8 Discussion

Based on these analyses, both selecting subsets of probes and encoding classifier weights in probe concentration appear to be viable approaches to mitigating amplification bias in molecular classifiers. When mitigating bias by selecting subsets of probes, one can use almost exactly the experimental construction presented in [4]. The only additional steps involve characterizing probe bias. For this reason, it is nearly certain that such a strategy could be used to build molecular classifiers. However, it is possible that many probes will need to be screened in order to identify a subset of probes that implements the desired classifier.

While encoding the classifier weights in probe concentration requires a slightly different, currently experimentally unverified construction, it would likely use a much smaller set of probes in order to implement the desired classifier, and thus be more cost effective. Even with very limited implementation precision, one could imagine having to screen far fewer probes than in the subset selection method. Furthermore, it allows one to implement weights with more specificity than the subset selection method, so provided that there is sufficient implementation precision, it should allow for more accurate implementation of an *in silico* classifier.

Gene expression profiling is increasingly an important clinical metric for diagnosis a wide number of human diseases. The approach presented here would enable amplification and classification of RNA samples containing multiple biomarkers in a single reaction with a two-channel fluorescent readout. This could drastically reduce the complexity of gene expression classification and potentially enable a point-of-care solution to this type of diagnostics. Even though this approach would result in a more complex development stage, once a final construction is found, the implementation is drastically simpler than existing alternatives.

Acknowledgment. G. G. was supported by Caltech's Summer Undergraduate Research Fellowship program. R. L. and G. S. were supported by NSF grant CCF-1714497.

References

1. Tsalik, E.L., et al.: Host gene expression classifiers diagnose acute respiratory illness etiology. Sci. Transl. Med. **8**, 322ra11 (2016)
2. Best, M.G., et al.: RNA-Seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. Cancer cell **28**, 666–676 (2015)
3. Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., Shapiro, E.: An autonomous molecular computer for logical control of gene expression. Nature **429**, 423–429 (2004)
4. Lopez, R., Wang, R., Seelig, G.: A molecular multi-gene classifier for disease diagnostics. Nat. Chem. **10**, 746–754 (2018)
5. Zhang, D.Y.: Cooperative hybridization of oligonucleotides. J. Am. Chem. Soc. **133**, 1077–1086 (2010)
6. Turberfield, A.J., Mitchell, J., Yurke, B., Mills Jr., A.P., Blakey, M., Simmel, F.C.: DNA fuel for free-running nanomachines. Phys. Rev. Lett. **90**, 118102 (2003)
7. Dirks, R.M., Pierce, N.A.: Triggered amplification by hybridization chain reaction. Proc. Nat. Acad. Sci. USA **101**, 15275–15278 (2004)
8. Seelig, G., Yurke, B., Winfree, E.: Catalyzed relaxation of a metastable DNA fuel. J. Am. Chem. Soc. **128**, 12211–12220 (2006)
9. Zhang, D.Y., Turberfield, A.J., Yurke, B., Winfree, E.: Engineering entropy-driven reactions and networks catalyzed by DNA. Science **318**, 1121–1125 (2007)
10. Zhang, D.Y., Seelig, G.: DNA-based fixed gain amplifiers and linear classifier circuits. In: Sakakibara, Y., Mi, Y. (eds.) DNA 2010. LNCS, vol. 6518, pp. 176–186. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-18305-8_16
11. Chen, S.X., Seelig, G.: A DNA neural network constructed from molecular variable gain amplifiers. In: Brijder, R., Qian, L. (eds.) DNA 2017. LNCS, vol. 10467, pp. 110–121. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66799-7_8
12. Li, H., Qiu, J., Fu, X.D.: RASL-seq for massively parallel and quantitative analysis of gene expression. Curr. Protoc. Mol. Biol. **98**, 4–13 (2012)
13. Cherry, K.M., Qian, L.: Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. Nature **559**, 370 (2018)
14. Burchill, S.A., Perebolte, L., Johnston, C., Top, B., Selby, P.: Comparison of the RNA-amplification based methods RT-PCR and NASBA for the detection of circulating tumour cells. Br. J. Cancer **86**, 102 (2002)
15. Deng, R., Zhang, K., Sun, Y., Ren, X., Li, J.: Highly specific imaging of mRNA in single cells by target RNA-initiated rolling circle amplification. Chem. Sci. **8**, 3668–3675 (2017)
16. Song, T., Garg, S., Mokhtar, R., Bui, H., Reif, J.: Analog computation by DNA strand displacement circuits. ACS Synth. Biol. **5**, 898–912 (2016)
17. Stougaard, M., Juul, S., Andersen, F.F., Knudsen, B.R.: Strategies for highly sensitive biomarker detection by Rolling Circle Amplification of signals from nucleic acid composed sensors. Integr. Biol. **3**, 982–992 (2011)
18. Takahashi, H., Matsumoto, A., Sugiyama, S., Kobori, T.: Direct detection of green fluorescent protein messenger RNA expressed in Escherichia coli by rolling circle amplification. Anal. Biochem. **401**, 242–249 (2010)
19. Stahlberg, A., Krzyzanowski, P.M., Jackson, J.B., Egyud, M., Stein, L., Godfrey, T.E.: Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing. Nucleic Acids Res. **44**, e105–e105 (2016)