chemprobstats: A pipeline for differential analyses of RNA chemical probing data

other possible names: diffProbeR, DARC-Probe (Differential Analysis of RNA Chemical PROBing data)

previous title: Modeling overdispersion in RNA chemical probing data and application to secondary structure prediction

Abstract

Chemical probing can uncover biological properties of RNA nucleotides, including structural context, chemical modification, and protein binding \cite{26575240, 28216634, 26544910}. Technological advances are continually expanding the applications of these techniques, as probes for new RNA properties are developed and probing assays have been adapted to *in vivo* conditions, with results read out by high throughput sequencing. Despite the great utility of chemical probing experiments for RNA biology, comparatively little is known about the statistical properties of chemical probing experiments. This is particularly important, because inference of nucleotide properties from chemical probing relies upon differential analysis of RT stops or mutations between treated samples and untreated controls. Moreover, properties like RNA structure and chemical modification are thought to differ tissue specifically and in other biological contexts, and differential analysis of probing experiments between conditions would be greatly aided by more detailed statistical modeling. Here we show that chemical probing data–like count data from many other biological assays, but in contrast to previous assumptions in the probing field–are overdispersed relative to the Poisson distribution. We devise a method to model probing data using the negative binomial distribution, implemented through an R package that we call *chemprobstats*. Our model of probing data provides an improved fit to a variety of RNA probing data, and resulting p-values track better with expected biochemical biases of probing reagents than previous models. We further incorporate our models into a method for RNA secondary structure prediction, which we benchmark against RNAs of known structure and apply to novel data for the Fendrr lncRNA. Finally, we apply our model to test for differential pseudouridine modification of RNA bases in different genetic contexts in yeast, laying the analytical groundwork for studies of tissue or condition specific RNA nucleotide properties.

Introduction

Comment [MR1]: Is there an appropriate review to cite?

Comment [MR2]: If calibration experiments from Simon lab succeed, mention them!

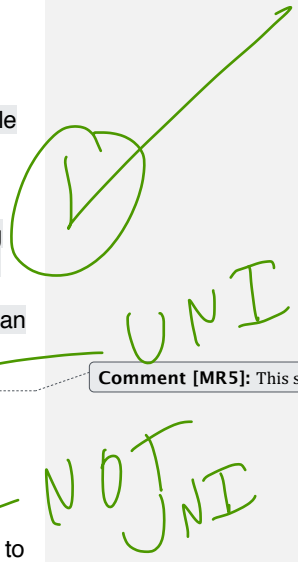
Comment [MR3]: Working on implementation for these things

Chemical probing techniques can reveal biologically important properties of RNA molecules at single nucleotide resolution. The applications of this versatile set of tools include studies of RNA structure, chemical modification, and interactions with proteins, all of which aid mechanistic investigations of RNA function and regulation. The utility of chemical probing experiments to study RNA biology has expanded in scope, as probes have been developed that work *in vivo* and techniques have been adapted to a sequencing platform for both transcriptome wide \cite{ 25192136, 28504680, structure: 24270811, 27819661, 24336214, 25951283} and targeted \cite{26646615, 26544910, 27578869} analyses. Technical improvements are also enabling experimentation under different biological conditions, which will enable a greater understand of how properties like RNA structure and chemical modification are regulated. The most widely used probes–e.g. dimethyl sulfate (DMS) and selective 2' hydroxyl acylating (SHAPE) reagents–aid RNA secondary structure determination by selectively modifying single-stranded and flexible nucleotides. Nucleotides modified by chemical probes are then read out by reverse transcriptase (RT), which terminates cDNA synthesis or inserts incorrect bases at chemical adducts (we refer to RT stops and mutations more generally as RT events). Comparing results of probing experiments to controls with no chemical treatment enables calculation of nucleotide reactivities, which are then converted into probabilistic constraints for RNA secondary structure prediction, or parallel inferences about other nucleotide properties \cite{19109441}.

Though chemical probing experiments can uncover diverse biochemical information about RNA nucleotides, the extent to which these experimental signals can be confounded by noise is not well understood. Ideally, if the conditions of experiments were exactly the same–such that every RNA molecule had the same probability of generating an RT event at a given nucleotide, both within and between experimental replicates–then observational noise could be modeled accurately using distributions that make this simplifying assumption. Indeed, the Poisson distribution, which is often used to model chemical probing data, assumes that the underlying statistical process being modeled is uniform. Moreover, most analysis methods for chemical probing analysis address the results of a single replicate, or the pooled results of multiple replicates rather than exploring the variability between experimental samples \cite{25028896, 28501650}. There is reason to be concerned that the Poisson model may underestimate the variability observed in probing data. Indeed, many types of biological data – ranging from gene expression (RNA-Seq)\cite{17728317, 19910308 } to mutation rates in cancer genomes \cite{ 26304545} – are often overdispersed relative to Poisson statistics (have greater variability than expected), due to heterogeneity in biological conditions within or between experimental samples. If chemical probing data are also overdispersed relative to Poisson statistics, explicit modeling of replicate data could be critical to assessing confidence in nucleotide reactivities and for downstream applications such as structure prediction. Statstical modeling would be all the more important

Comment [MR4]: Too similar to first abstract sentence

Comment [MR5]: This section needs to be tightened



to identification of nucleotides with different properties under different biological conditions. Supporting the opportunities to improve the analysis of chemical probing experiments, one recent method, BUM-HMM (Beta-Uniform Mixture Hidden Markov Model) used the empirical variability in control chemical probing experiments to assess the significance of differences between treated and control experiments\cite{ 27819660}, which showed improvements in robustness of observations for RNAs with low sequencing coverage. However, to our knowledge, no existing method provides a straightforward statistical treatment of independent probing experiments that explicitly models variability in both treated and control experiments.

Here we conduct a systematic investigation of statistical overdispersion in data from a variety of chemical probing techniques. We demonstrate that across multiple probes, types of experimental conditions (*in vitro* vs. *in vivo*), and in both targeted and genome-wide studies, counts observed from probing experiments are often overdispersed. To address this problem, we develop *chemprobstats*, a new tool that uses replicate observations to model overdispersion in chemical probing data. Our tool more accurately models the count data that are produced in chemical probing experiments, enabling statistical identification of significantly modified bases. We show an application of our approach to RNA secondary structure prediction, one of the traditional focuses of chemical probing analysis, developing a model-based approach to consider observational noise in converting chemical probing reactivities into constraints for structure prediction. We demonstrate the utility of our package for prediction of a variety of RNAs of known structure, as well as novel probing data for the 5' region of the Fendrr RNA. We further apply our model to perform differential analysis of pseudouridine modification in between wild type yeast and those lacking enzymes that catalyze the pseudouridine modification.

Comment [MR6]: These two analyses are underway! See text and plans below.

Results

RNA chemical probing data are overdispersed

To motivate development of a method to consider observational noise in the analysis of chemical probing data, we began by examining the replicate observations of a sample dataset: *in vivo* DMS probing of the mouse 18S rRNA. We focused on replicate RT stop counts from chemically treated RNA samples. Models commonly used to simulate or analyze probing data employ the Poisson distribution, which assumes the probability of producing and RT stop at a given nucleotide is the same for each RNA molecule, both within and between experimental samples. We hypothesized that probabilities of producing an RT stop at a given nucleotide may in fact be heterogenous between RNA molecules, which would the Poisson model to underestimate experimental variability. This heterogeneity could arise through differences in experimental conditions (particularly for *in vivo* samples) between experiments, or from differences in how

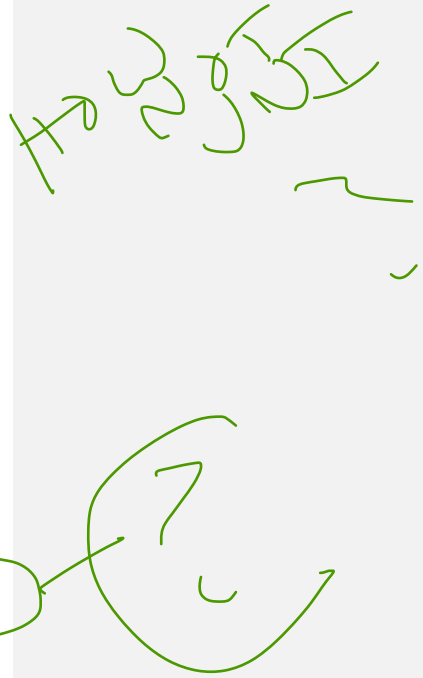
different RNA molecules (and cDNAs) behave through the entire process of probing and library preparation.

To gain an intuitive feel for whether our sample probing dataset follows Poisson statistics, we plotted the RT stop counts for two samples for a sample 50 nt region of the mouse 18S rRNA (Figure 1a), after normalizing for the number of reads where the RT reached each nucleotide of interest (see methods). Treating one replicate as a reference, we drew 95% confidence intervals assuming that data fit the Poisson distribution, and observed that for 19 of the 50 nucleotides, counts for the second replicate were outside the confidence interval of the reference replicate (2.5 outliers would be expected). To investigate the level of variability in probing data across a larger region of the 18S rRNA, we estimated the mean counts at each nucleotide, then simulated replicate datasets according to the Poisson distribution (Figure 1b), and finally compared the agreement between these simulated replicate observations to observed replicates (Figure 1c). Consistent with our initial observations, observed variability was much greater than that assumed by the Poisson model, implying that the data are overdispersed.

To investigate the overdispersion of chemical probing data more formally, we considered the p-values from the Poisson exact test between coverage corrected counts of replicates of the DMS-treated 18S rRNA data. Since these are replicate experiments, we expect that observations from both replicates at each nucleotide come from the same distribution. If the model accurately describes the variability of the data, this would lead p-values comparing replicate observations to follow the uniform distribution. To test whether this is the case, we plot the ordered Poisson exact p-values between replicates against the quantiles of the uniform distribution (quantile-quantile plot, Figure 1d) and observe that the Poisson exact p-values are almost all more extreme than any p-value. This observation is borne out by using the Kolmogorov-Smirnov test, a standard test for whether the goodness of fit of two distributions, which shows that Poisson p-values differ greatly from the uniform distribution ($p < 2.2*10\text{-}^{16}$). Together, these observations show that the Poisson distribution greatly underestimates the variability in *in vivo* DMS probing data of the Xist RNA, and that these data are overdispersed.

We next considered whether the overdispersion we observed in *in vivo* mouse 18S rRNA DMS probing data was particular to this dataset, or to some of its properties. It would be logical, for example to suggest that overdispersion comes primarily from biological variation, and that noise in *in vitro* probing experiments would be modeled more accurately by the Poisson distribution. Further, overdispersion could plausibly be observed specifically in a particular dataset, or only with certain chemical probing reagents. To address this question, we processed a variety of chemical probing data, both *in vitro* and *in vivo*, and for probing reagents including DMS, SHAPE reagents, and CMC which irreversibly modifies pseudouridine residues. We find that across these different conditions, all datasets tested display overdispersion as measured by the Kolmogorov-

Comment [MR7]: This section is clunky. Will refine later. I think that the language of hypothesis is good, but we want this section not to be too redundant with the similar section in the introduction.



Comment [MR8]: Improve descriptions of reagents, give final list when all data are processed.

Smirnov test (Table S1 ***not made yet). Interestingly, we find that RT event counts in most control experiments are also overdispersed, implying that heterogeneity in experimental steps conducted after probing is still sufficient to cause overdispersion (Table S1 ***not made yet).

Modeling overdispersion in chemical probing data by adapting methodology used for RNA-Seq analysis

Having established that chemical probing data are overdispersed, we next sought to develop a more accurate way to model count data produced by chemical probing experiments. This posed a challenge as relatively few replicates are typically conducted because of cost constraints, making it hard to make accurate variance estimates for each data point (nucleotide) individually. As we considered this problem, we noted that chemical probing techniques can be viewed largely as an extension of RNA-Seq experiments, where instead of counting reads at genes, RT events are counted at nucleotides. With the exception of chemical treatment, the key steps of the two techniques–reverse transcription, library preparation, and sequencing–are highly similar. Moreover, cost also limits the number of replicates produced for RNA-Seq experiments, and RNA-Seq data are well known to be overdispersed \cite{17881408, 20979621, 25516281}. We therefore considered whether we could adapt methods used for RNA-Seq analysis to model the overdispersion of chemical probing data.

To model overdispersion in chemical probing data, we chose to adapt the RNA-Seq analysis tool, DESeq2, which takes advantage of common information among many measurements (of gene expression) made in parallel to aid inference of count distributions \cite{25516281}. DESeq2 employs the negative binomial distribution, which is closely relative to the Poisson distribution but contains a dispersion parameter, α, which is zero when there is no overdispersion (Poisson) but takes higher values when data are overdispersed (see Methods). DESeq2 estimates the dispersion parameter by first making estimates for each gene (or nucleotide for chemical probing), then observing a trend between mean counts and dispersion values, and finally adjusting dispersion values toward the trend. Though DESeq2 can analyze normalized counts of any type, standard normalization for RNA-Seq is based upon the total number of reads in the experiment. In contrast, as above, we normalize input counts to the number of reads that reach the nucleotide of interest (for RT stops) or that cover the nucleotide of interest (for mutations).

We used DESeq2 to model normalized counts for our sample dataset: RT stop counts for *in vivo* DMS probing of the mouse 18S rRNA. We observe that as for RNA-Seq data, the dispersion parameters fit for each nucleotide trend with the mean counts (Fig 2a). To gain the same intuitive feel for the fit of the negative binomial models to the data, we compared simulated negative binomial replicates (Fig 1d) to observed replicates (Fig 1c) and Poisson replicates (Fig 1b), finding that the negative binomial replicates are much more similar to the real replciates

Real,

Comment [MR9]: Should this point be included somewhere in the introduction?

than the Poisson replicates. To evaluate the negative binomial models produced by DESeq2 more formally, we fit a model using 5 replicates of the Xist DMS data and computed negative binomial p-values given the model for the observations of a sixth replicate. As in our analysis of fit to the Poisson distribution, if the negative binomial model matches the variability of the data, then the p-values for the sixth replicate relative to the model should follow the uniform distribution. We observe that our negative binomial p-values are much closer to following the uniform distribution than Poisson p-values (Figure 1e), with less significant evidence that the negative-binomial p-values differ from expected uniform quantiles than we do for the Poisson distribution (Kolmogorov-Smirnov test, Poisson: $p < 2.2*10^{-16}$; Negative Binomial: $p = 2*10^{-5}$).

We next modeled the variety of chemical probing datasets of different types that we had found to be overdispersed relative to the Poisson distribution. For each dataset, we fit either Poisson or negative-binomial (using DESeq2) models to all but one replicate and then compared p-values of a test replicate to the uniform distribution. The Kolmogorov-Smirnov statistics for the negative binomial p-values are consistently lower (better fit) than for Poisson p-values (Figure 1f, Table S1), implying that our negative binomial fits improve modeling of a wide variety of chemical probing datasets.

Comment [MR10]: The negative binomial model actually has a non-significant KS-test p-value for some datasets. This is one of the worst fits and we should probably use something else (especially because some of the replicates were done on different days).

The new experiments will be ideal for this.

Using p-values from *chemprobstats* for biochemical inference

A more accurate statistical model of chemical probing data should aid biochemical inference from these datasets. The structure probing reagents SHAPE and DMS are expected to modify single-stranded nucleotides preferentially over double-stranded bases, while DMS is also selective for A and C bases. Meanwhile carbodiimide probing followed by reversal is selective for the modified base, pseudouridine \cite{25192136}. We performed statistical tests comparing treated and control RT event counts to identify bases that have significantly more RT events due to treatment. We then compared the distinctive ability of p-values from our negative binomial model to that of Poisson p-values, as well as the empirical p-values produced by recently published BUM-HMM method that uses empirical differences in untreated samples as the null distribution against which to test treatment-control comparisons. Precision-recall curves, which measure precision as increasing numbers of positive identifications are made, show that negative binomial p-values have greater distinctive ability than either of the other two statistical tests for distinction of AC vs. GU nucleotides for *in* vivo, targeted DMS probing of 3 RNAs: Xist, U2 snRNA, and 7SK RNA (Fig 2a-e). The negative binomial p-values outperform other tests for distinction of pseudouridine bases from unmodified uridines in carbodiimide probing data of the human 18S rRNA, and perform comparably to the other tests for distinction of single stranded vs. double stranded nucleotides from *in vitro* SHAPE probing of the *E coli* 5S rRNA \cite{27064082}. We report performance for an expanded list of RNAs, using the area under both precision-recall and

receiver operator curves as metrics, in Table 1 (***Table 1 not made yet***). We note that we would not expect perfect performance from a statistical metric in any of these comparisons, because the in no case is the chemical probing reagent perfectly selective for the type of base.

Because chemical probes are often not perfectly selective for nucleotides with the property of interest, with the potential for nonzero reactivity toward all bases, many analysis methods focus on the degree of reactivity of each nucleotide, as opposed to the statistical significance of its difference from the background control. Our negative binomial fits enable us to consider both the magnitude and confidence in our reactivity estimates simultaneously. To illustrate this, we made violin plots showing the inferred distribution of a common measurement of reactivity–the increase in probability of stopping or mutation due to treatment (denoted ΔP_{stop}, ΔP_{mut}, or more generally, ΔP; see Methods)– by simulating from the treated and control count distributions for each nucleotide \cite{ 26646615} (Fig 2e-h). We show these plots for segments of each RNA used in the p-value vs. biochemical property comparisons above. The use of inferred reactivity distributions has the potential to be useful for inference of the many different nucleotide properties that can be measured by chemical probing.

chemprobstats enables incorporates modeling of overdispersion into RNA secondary structure prediction

As an application of our statistical modeling of chemical probing counts, we sought to develop a method to incorporate inferred reactivity distributions into RNA secondary structure prediction. RNA structure prediction methods typically rely on levels of nucleotide reactivity, based on the observation that the degree of reactivity is more structurally meaningful than simply whether the reactivity is above zero \cite{19109441}. This may be because chemical probing reagents are selective for certain bases, but often have positive, nonzero reactivity toward all bases \cite{27803152}. RNA structure prediction methods typically perform one of a variety of normalization methods on ΔP or similar metrics of reactivity, and then define functions that convert these normalized reactivities into probabilities that each nucleotide is paired. These probabilities are converted to pseudoenergy terms that can be used to constrain free energy-based RNA secondary structure prediction algorithms \cite{19109441, 24895857}.

To create a replicate-aware method to incorporate chemical probing data into RNA secondary structure prediction, we represent counts for treated and control experiments by the negative binomial distributions we fit using DESeq2. Similar to the violin plots for ΔP above, we simulated many replicates of our experimental data. We then calculated ΔP, used the so-called boxplot method (see Methods, \cite{ 19109441 }) to create normalized reactivities. These reactivities were converted to probabilities using the method proposed by Deigan et al.:

Comment [MR11]: This is a bit redundant with the paragraph above.

Comment [MR12]: What is the best reference for this?

Comment [MR13]: These were previously mentioned earlier, and I think they probably should be returned to an earlier place in the results.

$$\Delta G = a*\log(R+1) + b$$
$$p(\text{paired})/p(\text{single-stranded}) = \exp(-\Delta G/R_{gas}T)$$
$$p(\text{single-stranded}) = 1/(1+\exp(-\Delta G/R_{gas}T))$$

Here, R represents normalized reactivity, R_{gas} is the gas constant, T is temperature, which is taken to be 37C (310.15 K). The final probability estimate was taken as the mean of the probabilities computed from sampling from treated and control count distributions. Of note is that there are two free parameters in the Deigan method that control the degree to which high reactivity indicates low pairing probability (a, above) and that no reactivity indicates higher pairing probability (b, above) (Figure 3b). The best values for a and b are typically determined based on optimization of prediction performance for RNAs of known structure.

We implemented our replicate-sensitive method for converting raw chemical probing counts to structure prediction constraints using *in vitro* SHAPE data for the 5S ribosomal RNA, an RNA of known structure \cite{ 25303992, 22976082, maybe others}. Pairing probabilities from our method were input into the RNAstructure software package for structure prediction. To evaluate our predictions, we use two metrics: sensitivity, the proportion of correct base pairs that are predicted; and positive predictive value (PPV), the proportion of predicted base pairs that are correct. Our probabilistic method enables correct prediction of the 5S rRNA structure (sensitivity = 100%, PPV = 100%, Fig 3a), in contrast to an inaccurate prediction from an unconstrained structure (sensitivity = 27.0%, PPV = 24.3%, Fig 3b). We can also see that prediction accuracy is robust to parameters of the Deigan pseudoenergy function that control the degree to which high reactivity indicates low pairing probability (a) and that no reactivity indicates higher pairing probability (b) (Fig 3c). This demonstrates the ability of our replicate-sensitive method to aid accurate prediction of RNA structure on a model RNA.

Comment [MR14]: I am working on adding other RNAs!

Comment [MR15]: Using a single replicate to predict the structure using standard methods is also accurate.

Application of noise aware RNA structure prediction to probing of the Fendrr lncRNA

The Fendrr lncRNA is essential to development in mouse models \cite{23369715, 24381249} and expression of its human transcript is associated with prognosis in gastric cancer \cite{25167886}. We sought to investigate this RNA's structure by combining chemical probing and our *chemprobstats* pipeline. We *in vitro* transcribed and the 5' region of Fendrr, and probed this RNA with the SHAPE reagent 1M7.

Plan:

- Fold 5' region of Fendrr lncRNA (maybe other regions as well)
- Plot structure with coloring for SHAPE constraints

- Assess confidence in predicted base pairs (pairing probability and Shannon entropy calculations)

chemprobstats **enables differential analysis of nucleotide properties between conditions**

The analyses conducted thus far consider probing experiments in a single condition, but nucleotide properties like RNA structure and chemical modifications can change substantially under different biological conditions. For example, RNA chemical modifications are thought to vary tissue specifically in humans, based upon the expression of enzymes that catalyze or regulate the deposition of modifications. We thus sought to extend our approach to identification of bases with different reactivity toward chemical probes under different conditions. The generalized linear model framework used by DESeq2 to fit our negative binomial models also enables testing of contrasts more complicated than simple comparison of two conditions. Here we ask whether the treated vs. control difference in one condition is greater than that in another. As a first test dataset, we used Pseudo-Seq data in yeast cells with and without deletion of the Pus7 gene, which is known to catalyze deposition of pseudouridine sequence specifically \cite{25192136}.

Plan:

- Quantify RT stops/coverage for PseudoSeq data in yeast
- Identify differentially modified bases
- Show motif enrichment in top hits, similar to that seen by Wendy Gilbert and Aviv Regev groups.

Discussion

In this study, we investigate the noise characteristics of chemical probing experiments on RNA. This field has expanding applications, ranging from a traditional focus on RNA secondary structure to increasing interest in identification of chemically modified nucleotides and sites of RNA-protein interaction. We establish that data from these experiments are often overdispersed, and often do not fit well to previous noise models proposed by other groups that make simplifying assumptions. We develop a pipeline – *chemprobstats* – that fits negative binomial distributions to the counts of reverse transcription events – either mutations or stops – that result from chemical treatment. *chemprobstats* further includes an application of statistical modeling of chemical probing data to RNA structure prediction. We show that the p-values from the *chemprobstats* more clearly distinguish nucleotides according to the expected selectivity of chemical probes in a variety of ways. Use of inferred from treated and control count distributions from chemical probing experiments in

structure prediction yields similar results to previous methods. We also develop a novel method to search for nucleotides with differential properties between biological states.

Several avenues exist for improvement of statistical modeling of chemical probing data. First, the biases leading to overdispersion of probing data are not fully understood. While some heterogeneity undoubtedly comes from biological variability for *in vivo* probing data, technical factors are also likely at play, especially given our finding that *in vitro* data are also overdispersed. It may be possible to correct or better model biases such as level of probe treatment, primer bias, and read mappability. Though this study has focused primarily on targeted experiments, progress is already being made in correcting for priming biases, read mappability, and transcript abundances in genome-wide experiments \cite{28501650}. We also recently showed that nucleotide adducts can induce RT stops or mutations preferentially at different nucleotides within the same experiment, and both can be meaningful for interpretation of RNA structure.

In addition to investigation of experimental biases, it may also be possible to improve statistical modeling of count data from chemical probing. Our method, based upon the RNA-Seq tool DESeq2, employs the negative binomial distribution and uses mean counts as a covariate to help infer which nucleotides have similar distribution parameters. Other statistical distributions, such as the ß-binomial distribution, may also be appropriate for modeling, and adding other covariates, such as nucleotide context may better inform which nucleotides have similar distributional properties.

Comment [MR16]: This paragraph is weak

Beyond modeling of the statistical count distributions of chemical probing data, we provide an application, incorporating reactivity distributions we fit into RNA structure prediction. It is notable that because chemical probes can likely react to some extent with all nucleotides, that simple detection of positive reactivity is not enough to yield information about a biochemical property of interest. Gold standard data are essential to the development of methods to infer biochemical properties from probing data. In the case of RNA structure prediction, this is further complicated by the fact that predictions involve probabilistic constraints from probing data and nearest-neighbor free energy models. The method we develop here should be viewed as a starting point, but a wide variety of methods have been considered to incorporate probing data into RNA secondary structure predictions, and merging these methods with statistical models of count data, or development of new methods, may improve structure prediction. Similarly, the statistical modeling techniques that we develop provide a foundation for considering observational noise when inferring biochemical properties from any chemical probing method.

We apply our chemprobstats pipeline to analysis of the secondary structure of the 5' region of the Fendrr lncRNA. We find …… This illustrates the value of the *chemprobstats* package, and suggests …… about Fendrr biology.

Comment [MR17]: Also to be written

Add something about results from differential pseudouridine modification analysis!

Figure legends

Figure 1. Overdispersion of chemical probing data and modeling using the negative-binomial distribtution and DESeq2. (Panels A-F go from left to right, then top to bottom)

A. Plot of mouse 18S DMS stop counts from a reference replicate with Poisson confidence intervals (black). Counts for a second replicate are plotted as well, with points marked in red if they exceed the Poisson confidence interval and otherwise in gray.
B. Scatterplot of Poisson simulated replicates for mouse 18S DMS stop counts.
C. Scatterplot of two biological replicates of mouse 18S DMS treated with DMS *in vivo* stop counts.
D. Scatterplot of negative-binomial simulated replicates for mouse 18S stop counts. Distribution was fit using DESeq2.
E. Relationship between mean counts and the negative binomial dispersion parameter (which controls overdispersion) while fitting mouse 18S stop counts using DESeq2. Dispersion parameters are initially fit separately to each nucleotide (black dots), then a trend is fit to these dispersion estimates (red line), and final estimates (blue dots) are made combining information from the individual estimates and the trend.
F. Quantile-quantile plot for stop counts from mouse 18S treated with DMS *in vivo*. P-values of observed data against assumed distributions are compared to the uniform distribution on log10 scale. From six replicates, one was chosen to fit the Poisson model and the other five were tested against this replicate.

Figure 2. Evaluation of biochemical inferences made with negative binomial p-values (from DESeq2) and comparison to other methods
(Panels A-I go from top to bottom, starting with the left column)

A. Precision recall plot for A and C bases for *in vivo* targeted DMS probing of the mouse 18S rRNA. Negative binomial p-values fit with DESeq2 are compared to those from BUM-HMM and the Poisson exact test.
B. Precision recall plot for A and C bases for *in vivo* targeted DMS probing of the mouse U2 snRNA.

C. Precision recall plot for A and C bases for *in vivo* targeted DMS probing of the mouse Xist snRNA.
D. Precision recall plot for pseudouridine bases for *in vivo* carbodiimide probing (with reversal) of the human 18S rRNA. Positives are pseudouridine, and negatives are uridines; other bases are excluded.
E. Precision recall plot for A and C bases for *in vitro* SHAPE probing of the E coli 5S rRNA.
F. (F-J)Violin plots showing distributions of ΔP_{stop}, a measure of nucleotide reactivity. Plots show data for (F) Targeted *in vivo* DMS probing mouse of the mouse 18S rRNA RNA, (G) *in vivo* DMS probing mouse of the mouse U2 snRNA, (H) *in vivo* DMS probing mouse of the mouse Xist RNA, (I) *in vivo* carbodiimide probing (with reversal) of the human 18S rRNA, and (J) *in vitro* SHAPE probing of the E coli 5S rRNA. Violins represent ΔP values calculated from 10,000 sets of treated and control counts, simulated from negative binomial distributions fit to real data. Violins for DMS treatment colored by RNA base, while for SHAPE, violins are colored by when nucleotides are single-stranded (green) or paired (gray). Nucleotides with significant levels of reactivity (FDR < 0.01) are indicated by bold letters with stars below along the x-axis. Bars in the centers of violins indicate mean estimated reactivity. Dots indicate estimated reactivities from individual replicates.

Figure 3. Structure prediction of 5S rRNA with SHAPE data
A. 5S rRNA structure predicted using *in vitro* SHAPE constraints with the *chemprobstats* approach. Nucleotides are colored by the probabilities of pairing produced with probing data and our negative binomial models. This predicted structure is correct (100% sensitivity, 100% positive predictive value).
B. 5S rRNA structure predicted without constraints. Nucleotides are colored by the probabilities of pairing produced with probing data and our negative binomial models. This predicted structure has sensitivity = 27.05%, positive predictive value = 24.3%,
C. Sensitivity of predicted 5S rRNA structures with SHAPE constraints, using different parameters of the Deigan pseudoenergy function.
D. Positive predictive value of predicted 5S rRNA structures with SHAPE constraints, using different parameters of the Deigan pseudoenergy function.