Rubric for peer review

Intro:

Unfortunately, there is not standard way to conduct peer review, but below is a system I tend to use when reviewing papers. Also, as an appendix, I have provided an example of a mock review I did for a bioRxiv paper as an exercise training one of my grad students. I was not actually a reviewer of this paper, and in this exercise I is focused on just Quality of Science rather than Impact for this paper (since it is in bioRxiv and a bit outside of my expertise for knowing where the eRNA field is at the moment, without additional reading to figure out impact).

When I review a paper for publication I think about two major things.

1) Quality of the Science (this is where I put 80-90% of my effort)
2) The Impact of the Findings (how novel/different is this? Does it belong in this journal).

To evaluate #1, the big picture focus is on establishing the claims that the make (what are their big picture conclusions) and determining if the data support those conclusions. If not, then they either need to alter their conclusions (for a lower impact journal) or do additional data. So mostly it is about evaluating the quality of the data – are the experiments the right experiments? Or are there more definitive methods or additional methods that need to be brought to bear? Are they well performed? How is the N? Is the statistical approach appropriate? *A key one: Are additional controls needed?* I would say 25% of my comments are requests for additional controls.

The other part of this is to make sure that they are not ignoring key findings in the field, and appropriately putting their work in the context. This influences both #1 and #2. For #1, if there are major discrepancies between their observations and others, those need to be discussed. For #2, if this work is repeating what is previously done, or findings are a logical extension of current models, it is often of somewhat lessor impact. However, #2 is fundamentally very subjective and honestly how things end up in different journals is a bit random in a way that is probably unfortunate for the functioning of science. But I digress. Going back to #1, this is step by step how I review a paper.

Step by Step:

1) Read title and abstract, and extract the major claims of the paper. I will often immediately paraphrase these into my own words (as the intro to my review) and then as I read the rest of the paper I am going to be looking at how well they support these claims.
2) Take a glance at the authors, and think about their prior work and how this builds on this, and how much expertise they have in this area (sometimes I leave thinking about this for last, to stay more unbiased in my reading. I actually think one could make a strong argument for review being more fair if it were always double blind)
3) Read the introduction. How well do they set up the question? With regard to impact, is it an important question. The main thing – are they ignoring key findings from other labs that they should be introducing in the background? Some labs are known for really hacking the literature to make their work more novel.

4) Now results and figures. The most important part.
    a. Are the experiments the right experiments? Do they show what the authors say they do? Or are there more definitive methods or additional methods that need to be brought to bear?
    b. Is the data presented in a convincing way?
    c. Are they well performed? How is the N? Is the statistical approach appropriate? *Are additional controls needed?*
    d. I also make notes as to things I want to specifically check for in methods (controls they didn't mention. N. Read depth, ect.).
5) Now methods
    a. Did they mitigate any concerns I noted after reading results?
    b. Are the methods well described enough I could do these experiments in my lab?
    c. Are they sharing any data that they are supposed to be sharing?
    d. I also note to the editor (in private comments) which parts of the paper I simply felt I was not qualified to evaluate (e.g. electrophysiology. Human brain imaging). And if I have critiques that I think I see in these areas, I will state them to the editor but also make it clear that for this area I am not as confident. (I do this because I think it sucks having a reviewer ask for stupid experiments because they don't understand the technique, so I try not to be that guy).
6) Now discussion. I am a bit more permissive with discussion. I think authors should have the right to discuss data in any way that they like with three caveats. 1) They should not over-reach the results. They are, however, allowed to speculate as long as they write in a way that is speculative. ("Our mouse results indicate that X could be a promising treatment option for human autism" vs. "We have discovered the cure for autism") 2) They should indicate limitations of their study a bit (Note though in high profile journals, authors usually don't do this because that provides fodder for lazy reviewers. Less nefariously, when you are limited on space, it is often the first thing to go). 3) They should not ignore important literature that would change either the interpretation of impact of their findings. This is more insidious and harder to detect as it requires either a good familiarity with the literature, or at least a cursory google scholar /pubmed search. As in many things, omission in this area can either be sneaky/deliberate, or merely an oversight. The literature is broad and deep.
7) Throughout my reading I am glancing at how they are referencing things to make sure it makes sense. I usually don't go read more than the titles of those though until I finish the paper I am currently reading (I want to get a sense of the flow). Sometimes I then go read key refs to understand background, methods, or context.
8) Because I spend so much time copy editing for my students I tend to also copy edit as I go and add a list of typos at the end of my review. This is a compulsion of mine. Most reviewers don't bother, aside from a blanket statement the effect of "There were too many typos, is your science as sloppy as your writing?"
9) Finally, I look at all my comments, re-read the sections that made me think of them. I double check that they are not addressed somewhere else in paper (i.e. methods), and cross them out if they have been resolved (though if they were raised just because they were not clearly written, I'll keep a note of that and mention it as a minor concerns needing revision). Then I break my comments into 4 categories:

a. Major, likely unaddressable paper-killers. There is some flaw so big and deep this paper needs to be rejected. The idea is either clearly completely wrong, or it would take > 3-4 months to do all the experiments needed to fix it, and thus it is better off the authors get rejected and start fresh in 4 months.
b. Major, but could be addressable with the right experiments: Usually this is to the effect of key controls being missing that are needed to support their major conclusions. However these controls are feasible based on demonstrated methods (esp. demonstrated by authors). Concerns about insufficient N or wrong statistic might fall here too. Basically, they need to go do more exps, and if those experiments pan out, then they have a paper. If not, they wont.
c. Minor: This is where most of my comments often go. If I think data should be displayed differently, or writing is unclear, or certain refs or method details are missing, they go here. These need to be addressed, but don't have the potential to kill the paper if the go awry. Many can be addressed with changes in writing but without additional experiments. This might also be where I talk about potentially weak support for conclusions they draw, but conclusions that are not key to their paper. Often I advise authors to simply remove or reduce these conclusions, if they are tangential to the main point of the paper anyway. Or if an experiment is premature and poorly executed, but not really informative anyway, just take it out. Or discuss it very differently
d. Things I would like to see, but aren't needed: I am unusual in my approach to this too. All readers/reviewers but think of exps whose outcomes would be interesting and informative if they were to do them. If these fill a key hole, then I propose them for b or c. But if it is just something I think would be interesting, I just say so. The way I am unusual is that a lot of reviewers would force you to do them anyway. I guess you could lump under here also experiments to increase 'impact.' At high profile journals you will often get this kind of comment as a major concern. "All of these experiments are well executed and support the conclusions as stated, but for this to be suitable for the Nature Underpants audience, the authors also need to show that their new underpants provide additional key support in humans as well as in their mouse model. Or that they treat disease X, or whatever." Just basically, "For this journal, I want more." Also you see a fair amount of "I care about hormone receptors, so please do experiments tying your data to hormone receptors." Clearly the author side of me kind of dislikes some of these comments.
e. Finally, I list typos.

Appendix – example review:

Mock review of: https://www.biorxiv.org/content/biorxiv/early/2018/07/17/270967.full.pdf

In this paper, Carullo et al attempt to address the question of whether eRNAs are necessary and sufficient for activity dependent gene expression in the context of neurons. They start by analyzing existing data to look for eRNA induction in cultured neurons in response to activity. They show they can

detect putative eRNAs, that their expression and induction can correlate with nearby gene expression. The focus on 1-4 immediate early genes for functional analysis. They then use dCAS9 fusions to activation domains to show that recruiting activating enzymes to candidate enhancers can increase eRNA expression from the enhancer, as well as activate expression of the nearby presumed target gene. This is consistent with the role of these sequences as enhancers and is a well executed study showing an expected result. Enhancers have specificity for their targets, and act independently in this system. Experiments are generally well powered (n=9). They extend these findings to multiple methods of mRNA induction and show a requirement for Pol2 in cultured neurons using drugs. They also conduct an smFISH study to show eRNA production weakly correlates with target gene production in single neurons, though it is not as clear how this is linked to the rest of the paper. Finally, they use "CRISPR-Display" to determine what happens when they recruit an eRNA sequence to the enhancer using a gRNA, as well as use ASOs to try and suppress eRNA to see if they can suppress induction. These last two experiments are key to demonstrating potentially causality for eRNAs in controlling gene expression. The unstated alternative hypothesis, is that eRNA production is simply an epiphenomena of and not really required for enhancer activity. The key question in evaluating this paper scientifically is 1) How much did they support their hypothesis over the alternate. A related question, for impact, would be how different is this from what is seen using the same tools in other systems (this requires further reading to determine)? That doesn't change the science, but if it is the first time this is shown anywhere, then it is of higher impact than if it is simply showing that eRNA function in neurons is the same as it is in previously characterized systems (i.e. it would go to a neuro journal rather than a general high impact journal).

Recommendation: Major Revision.

Major concerns:

1) One of the major conclusions of the paper is that eRNA alone is sufficient to induce expression of the Fos mRNA, based on Fig 5 (CRISPR-Display). However, the data do not support quite such a strong statement: In 5C left panel, eRNA CRISPR-display is not sufficient, by itself, to significantly induce Fos. It does, however, appear that eRNA CRISPR-display can enhance the induction by KCL (right side of panel C). This suggest that eRNA might indeed have a biological effect. I would like to see the following changes 1) alter title and conclusions to match data. 2) in 5C, show the data both normalized and un-normalized. 3) an additional control that would be interesting is to use an eRNA sequence but from elsewhere in the genome, but with gRNA to the same Fos enhancer. This experiment would ask whether the sequence of the eRNA matters, or merely the presence of RNA at the loci.
2) A lot of experiments (e.g. 5C) are based on qPCR with n=9. However, a potential major concern is that it is not explicitly stated that these are 9 independent biological replicates. Some groups confuse technical replicates (pipetting the same cDNA into 3 wells to account for pipetting or measurement error) with biological replicates (3 independent cultures, assessing biological variability). Three biological replicates, each measured in triplicate on qPCR, are still n=3, not n=9, and statistically should be treated as such. I am not saying the authors made this mistake, but as written it is not clear if they did or did not. This should be clarified and stats reconducted if needed, and/or biological replicates added.

Minor concerns

1) Even if major concern 1 is addressed and there is still an effect of the eRNA Display, I think it would be important for the authors to discuss these results in comparison to the magnitude of the findings seen in 2C. In other words, if eRNAs do a have a role, can one state how significant this role is (20% of the magnitude of activating the enhancer by changing chromatin?) This might change conclusions not as to whether eRNA can have an effect, but the extent to which it matters. Perhaps it plays a role, but is not a major contributor. This discussion would add nuance to their findings, but not negate them. The authors could also consider applying their eRNA ASO in the context repeating experiment 2C with the dCas9 activation. This would allow them to directly assess whether the eRNA is required for all, or just a fraction, of the induction.
2) The single cell RNA ISH is only moderately informative to their author's questions, at least as currently justified. I am not sure it needs to be removed, but if it is included it needs at least one additional control: since signal from eRNAs is so low, I would like to see them also compared to a non-specific ISH probe like a LacZ or something to make sure the results are robust. Also, the panels for 2D might need to have all 3 single colors displayed independently. I cannot see the eRNA signal they are pointing too in either panel.
3) Seeing 4 nuclear loci for eRNA is also confusing. Perhaps an intronic FOS RNA FISH would be informative in defining the transcriptional loci? Then again perhaps not (I'm still not clear on the premise for this experiment).
4) When the authors say "and at these loci, activity-induced expression of eRNAs and mRNAs were significantly correlated". They should make it clear that this is the expected result, as the activity induced eRNA and mRNAs were defined in such a way that this had to be true. It is phrased too much like a finding, but as a finding it is a very circular one.
5) In figure 1E-G authors show counts rather than CPM. If they are using counts, they should specify that libraries had same seq depth. If not, they should normalize (i.e. CPM).
6) Figure 2F needs statistical analysis. * can be added to quadrants where the effect was significant.
7) I would like some discussion of their ASO findings in light of papers showing miRNA repression can lead to a coordinated silencing of targeted loci. Is it the loss of eRNA that matters? Or does Ago2 recruitment to these area of the nuclei silence the chromatin so the ASO effect is actually gain of function rather than loss of function? Their data could be consistent with either model and this should be discussed.