Tips for writing papers

Eran Segal
Weizmann Institute of Science

Working Hypothesis for this Talk

- **You have the results for the paper**
- **You just need to package them into a paper...**

Paper Structure

- Title
- Abstract
- Introduction/Background
- Results
- Conclusions/Discussion
- Figures
- Methods

- **A paper is not a history of your work!**
- **A paper should have one main message**
  - Message should be reflected in each section of the paper

Title

- **Title should convey your main message**
  - "Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization"
- **Alternatively, title can be broader and provoking and not tell the whole story (ok to ask people to read abstract…)**
  - "The role of site accessibility in microRNA target recognition"
  - "A genomic code for nucleosome positioning"
  - "Genome-wide Measurement of RNA Secondary Structure in Yeast"

Test Yourself #1

- **Paper and main message can be read in multiple ways**
  - Title alone (sometimes)
  - Abstract alone
  - Intro alone
  - Paragraph headings alone
  - Conclusion alone
  - Figure titles alone
  - Figure visuals alone
- **Pitches are at various different levels**
  - Elevator and Pubmed pitch (title, abstract, par. Heading,)
  - For the quick glancers (figures, figure titles)
  - For the deep reader (full paper + methods + supp)

Test Yourself #2: Abstract

- **Challenge is in a short space, to structure it by:**
  - Motivation only pertinent to main message
  - The unanswered question that you tackled
  - What you did, the dry facts: “Here, we…”
  - Key results, may be your main message
  - Broader view of the results and impact, can be your main message

Abstract example

- Eukaryotic genomes are packaged into nucleosome particles, which occlude their DNA from interaction with most DNA binding proteins. Nucleosomes have higher affinity for particular DNA sequences, reflecting the sequences' ability to sharply bend, as required by the nucleosome structure. However, it is not known whether these sequence preferences have a significant influence on nucleosome positions in vivo, and thus regulate the access of other proteins to DNA. Here, we isolated nucleosome-bound sequences at high resolution from yeast, and used these sequences in a novel computational approach to construct and experimentally validate a nucleosome-DNA interaction model, and to predict the genome-wide organization of nucleosomes. Our results demonstrate that genomes encode an intrinsic nucleosome organization, and that this intrinsic organization itself can explain ~50% of the in vivo nucleosome positions. This nucleosome positioning code may facilitate specific chromosome functions, including transcription factor binding, transcription initiation, and even remodeling of the nucleosomes themselves.
  - 2 sentences of general background related to our main message
  - 1 sentence of motivation, what was not known before our work?
  - 1 sentence on what we did, the dry facts
  - 1 sentence on what we found
  - 1 sentence on what we found in a broader context and impact

Test Yourself #3: Introduction

- **General background**
  - A paper is not a review!
  - For each background given, test whether it helps to motivate your specific work and message
  - Make the introduction understandable to non-specialists
- **Sentence/paragraph on what is the unanswered question that you will deal with and why is it interesting**
- **What you did**
  - The dry facts at a very high level. Typically, "Here, we..."
- **The key results**
- **No need to discuss broader impact**
  - Enough of this in abstract and conclusions
- **Bad: "In section 1, we show... , in section 2,...**

Test Yourself #4: Results

- **Each paragraph has the following structure**
  - What did you ask or wanted to test next and why is it important?
  - What you actually did: the dry facts
  - Interpretation of the results: what was the answer
  - Scientific test: do your facts support the interpretation?
  - Each paragraph should read coherently even when skipping over the dry facts
  - Note: sometimes this spreads over 2 paragraphs

Paragraphs

- We used several tests to gauge the accuracy of our method. First, we examined the nucleotide distribution over the first bases of the sequenced fragments. We found that the sequence specificity of these bases has little information content, suggesting minimal sequence-dependent bias in our adapter ligation and cDNA conversion process (Supp. Fig. 1). Second, since some cDNA creation protocols exhibit a bias towards reads from the 5' end of the transcript, we examined whether our protocol has such a bias, by plotting the number of reads as a function of position along the transcript, averaged across all of the transcripts. Positions that exhibited the largest deviation from the mean coverage had 18% more reads than the mean coverage, suggesting that our protocol has only a relatively small bias towards particular regions along the transcript (Supp. Fig 2). Next,..., … together with the above comparisons, the significant correspondence between PARS and folding prediction provides a global independent validation for the ability of PARS to provide genome-scale and high-quality measurements of properties of RNA structure at single nucleotide resolution.
- Red highlights demonstrate outer structure of entire paragraph
- Blue and yellow highlights also follow the same structure locally
- **Avoid just stating result after result and always give interpretation**

Paragraph Heading Example

- Constructing and validating a nucleosome-DNA interaction model
- Predicting the nucleosome organization intrinsic to the genomic DNA sequence
- Intrinsic nucleosome organization determines many in vivo nucleosome positions
- Global features of the intrinsic nucleosome organization in yeast
- Intrinsic nucleosome organization varies by chromosomal region type
- The yeast genome encodes low nucleosome occupancy over functional binding sites
- Genomes encode low nucleosome occupancy over transcription start sites

Figures

- Visually striking
- Clear
- Color consistency

Figure Headings

- Good: Our protocol has minimal bias towards particular regions of the transcript
- Bad: Average read density as a function of position along the transcript

- Figure 1. Probabilistic nucleosome-DNA interaction model.
- Figure 2. Genome-wide prediction of intrinsic nucleosome organization and comparison to literature-reported nucleosome positions.
- Figure 3. Higher order features of intrinsic nucleosome organization and comparison with in vivo occupancy experiments.
- Figure 4. Intrinsic nucleosome occupancy varies with genomic location type and is low at functional transcription factor binding sites.
- Figure 5. Genomes encode unstable nucleosomes at transcriptional start sites.

Conclusions / Discussion

- **One paragraph summary**
  - Tell the main message in yet a different way
- **Limitations**
  - Your place to be critical about your results
  - Lets reviewers/readers see that you are aware of issues
  - Allows you to tell people why some issues are non-issues or how you may deal with them
- **1-3 sentences about broader impact of the work**

Writing Stages of the Paper

- **Figures:** Start by getting the main message visually
- **Title**
- **Abstract:** the shortest version of your full story
- **Introduction:** Decide what each paragraph will say
- **Results (1):** Write down all the paragraph headings
- **Conclusions:** What each paragraph will say, discussion point
- **Results (2):** Write part 1 (motivation) and 3 (interpretation) of each paragraph
- **Fill in missing text** (Intro, Results, Conclusions)
- **Finalize technical stuff**: Figure legends, Methods

General Tips

- **Active versus passive**
  - Bad: Read counts were computed for every base along each annotated coding sequence
  - Good: We computed the number of reads at each base along every annotated coding sequence

General Tips

- **Statements must be accurate**
  - Omitting details is ok, but not at the expense of accuracy
  - Test yourself: do they stand to trial?

General Tips

- **Be consistent in colors and annotations**
  - Objects of the same type should always have the same color

General Tips

- **Quantify every assertion or don't make the assertion**
  - Bad: PARS and probing are very similar
  - Good: PARS and probing are highly similar ($R=0.9$, $P<0.001$)

General Tips

- **Be specific whenever possible**
  - OK: PARS results are highly reproducible.
  - Good: Our protocol samples cleaved RNA fragments in proportion to their abundance