20.109 Communication Workshop 4

Research Manuscripts: structure and writing process

Prerna BOOrgava

Sean of the Dead Clarke

MIT BE
BIOLOGICAL ENGINEERING

Communication Lab

There are no explicit models for successful papers.

Collect papers that you like!

Analyze what makes them especially clear & compelling.

Try using their techniques.

Revising papers is essential

- Do not try to write this paper in one day.
- **Outline** or draft the sections, then set aside the paper for several days.
- When you get stuck: write topic sentences, work on the next section, look at examples
- Get feedback:

  peers, instructors, Comm Lab Fellows!

Writing a paper integrates topics we have already covered…

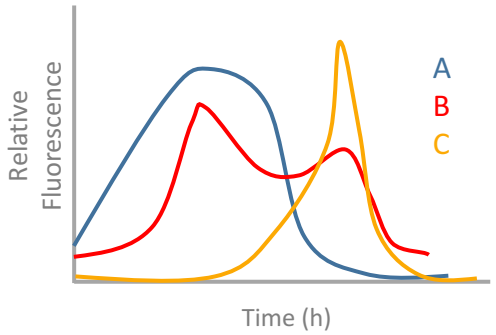
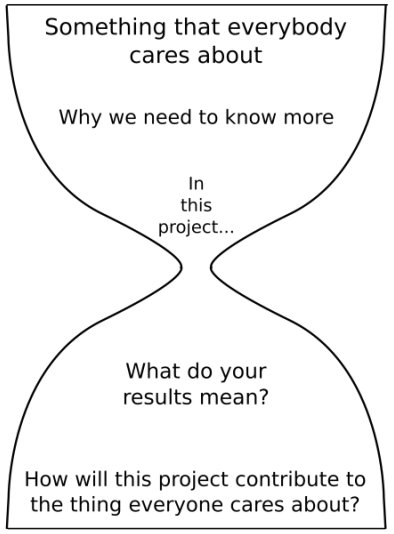


Fig. 1: A, B, and C have different dynamics under Condition X. A, B, and C were sampled using Method 1 and their fluorescence quantified with Method 2. Fluorescence data normalized to negative control.

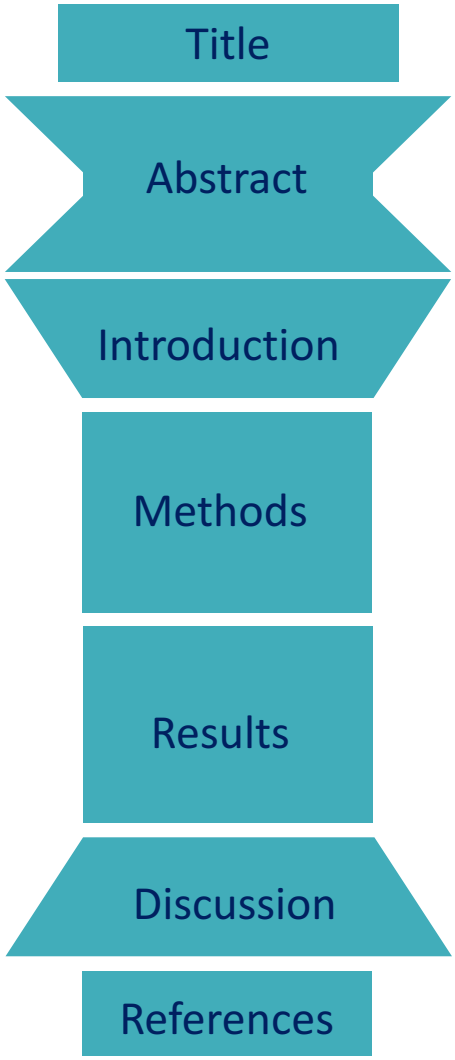
Figures & Captions



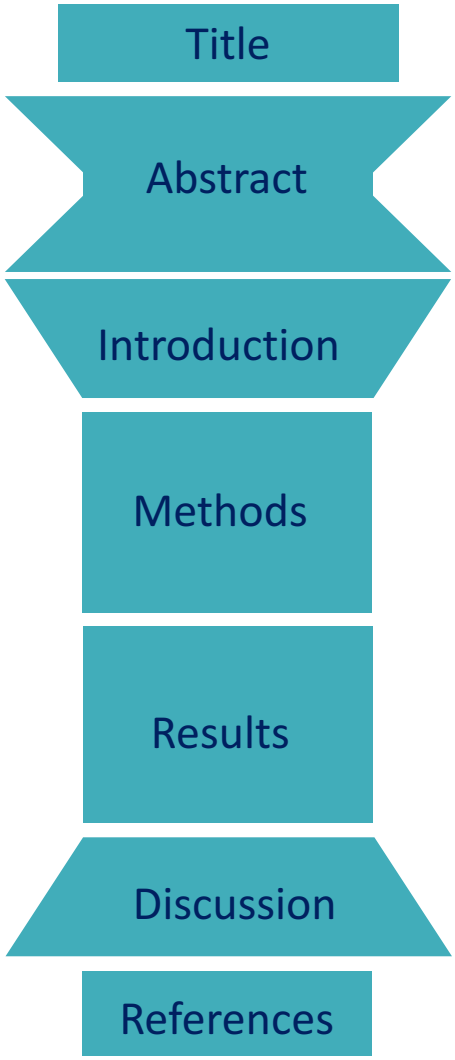
Abstracts & Titles

What do we already know about the structure of a manuscript?

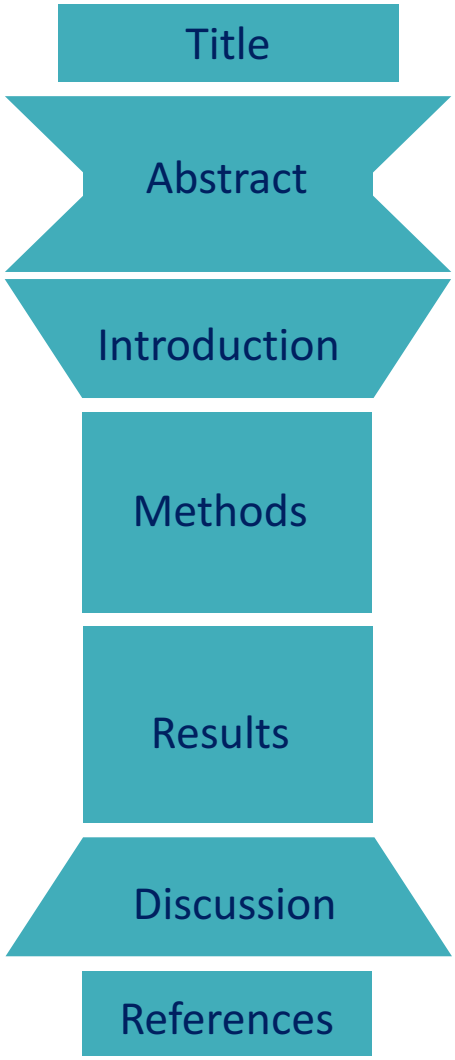
A standard manuscript has sections that guide the reader



Papers are often thought of as linear…



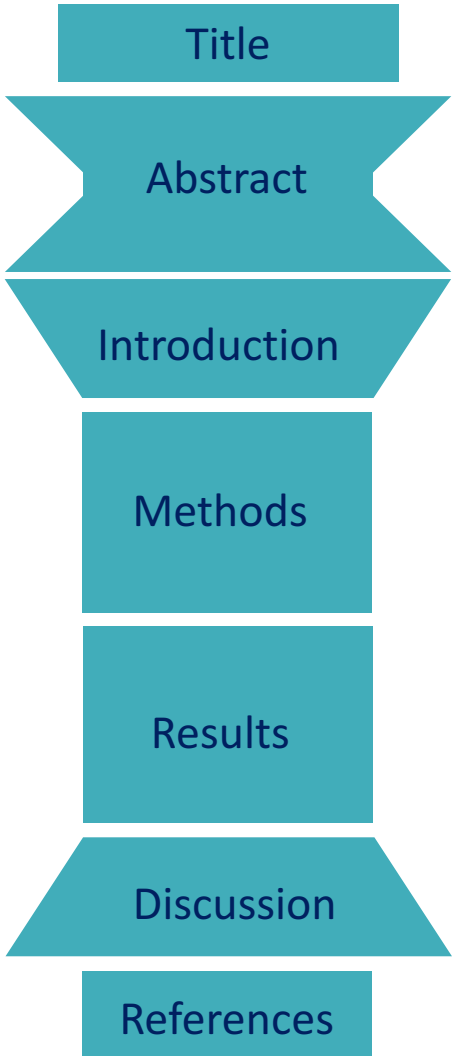
…yet are both **read** and **written** nonlinearly.



In what order do you read a paper?

In what order will you write a paper?

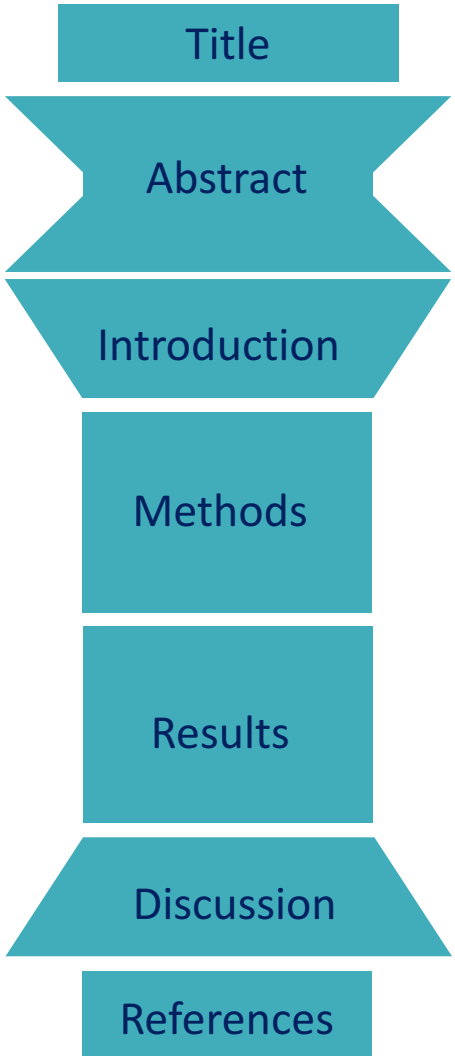
We recommend the following order for writing your paper:



1. Figures + Captions
2. Results
3. Discussion
4. Introduction
5. Abstract
6. Title

7. Methods

A paper is also fractal, building in redundancy to help guide the reader

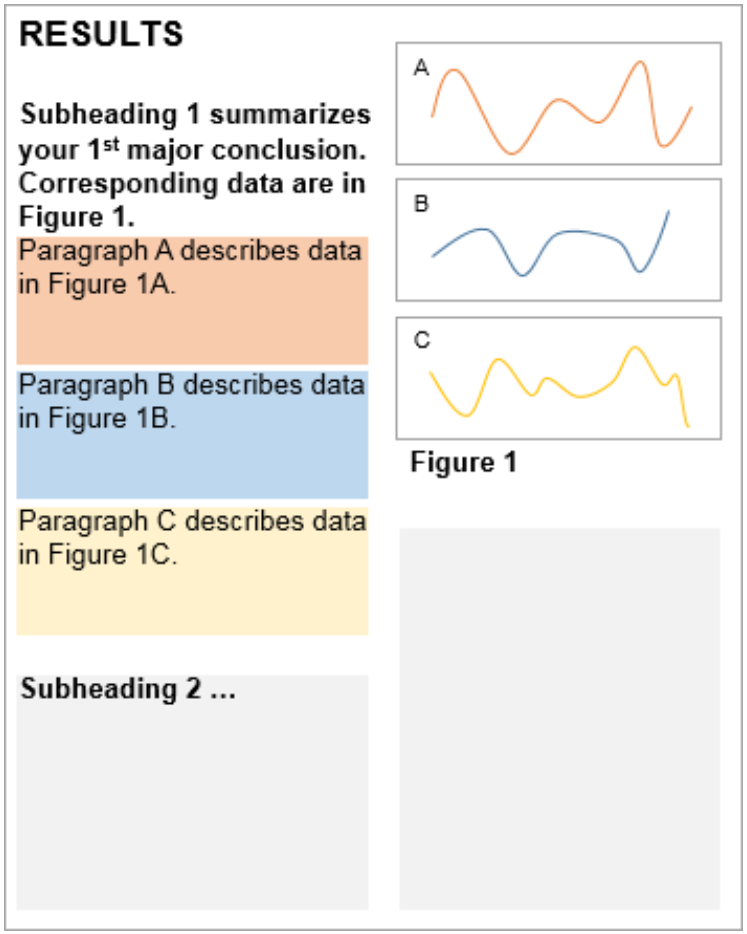


Redundancy in your paper helps your reader find the information they need.

General background	Something everyone in your audience cares about.	Introduction: beginning
Specific background	Zoom in from General Background to the thing you did.	Introduction: middle
Knowledge gap, Unknown	Question that will be answered by your research. Problem, phenomenon that is not understood.	Introduction: end
HERE WE SHOW	Conclusion, answer to the Unknown	Introduction: end Results: end Discussion: beginning
Results	Brief summary of approach + very high-level results. Common pitfall = too much Methods/Results.	Introduction (high level) Results Methods
Implication, Significance	So what? What do your results mean for the thing everyone cares about?	Discussion

Use parallelism: Put all of your content in the same order.

Data || Methods || Results || Discussion

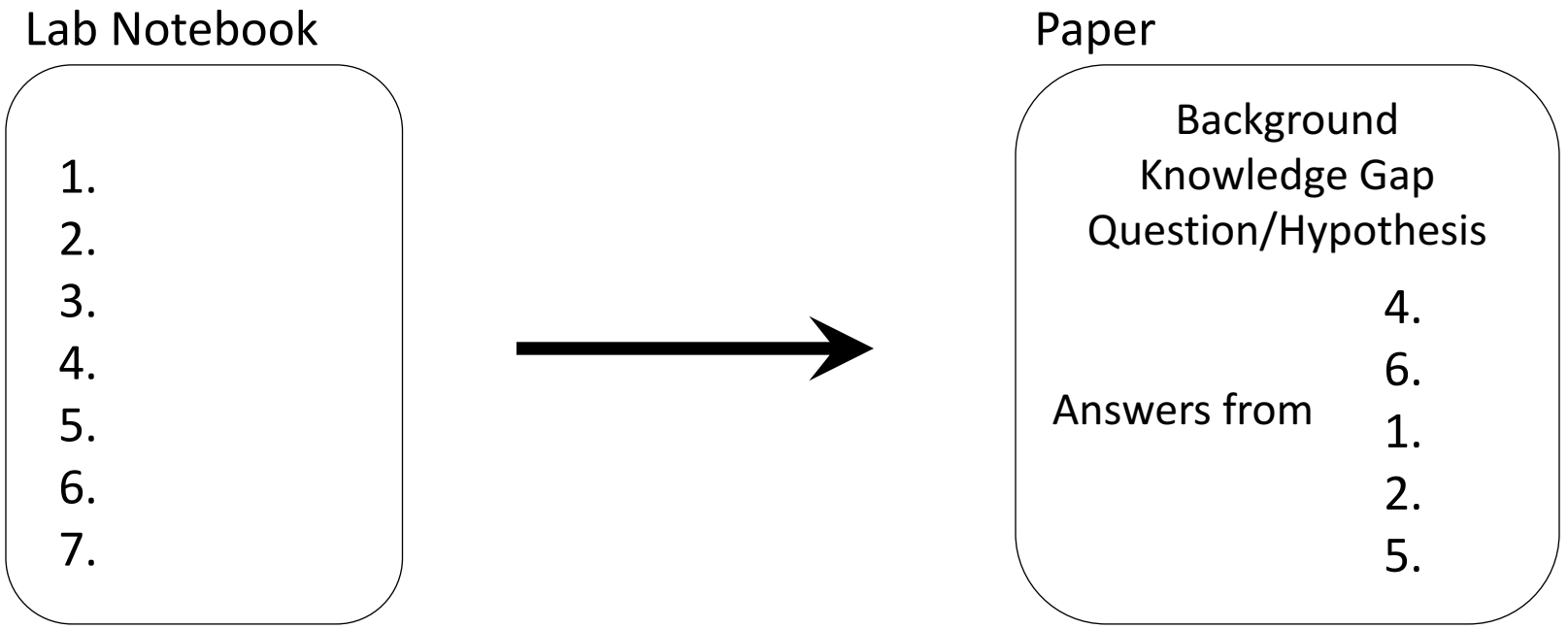


Methods: Most experimental detail

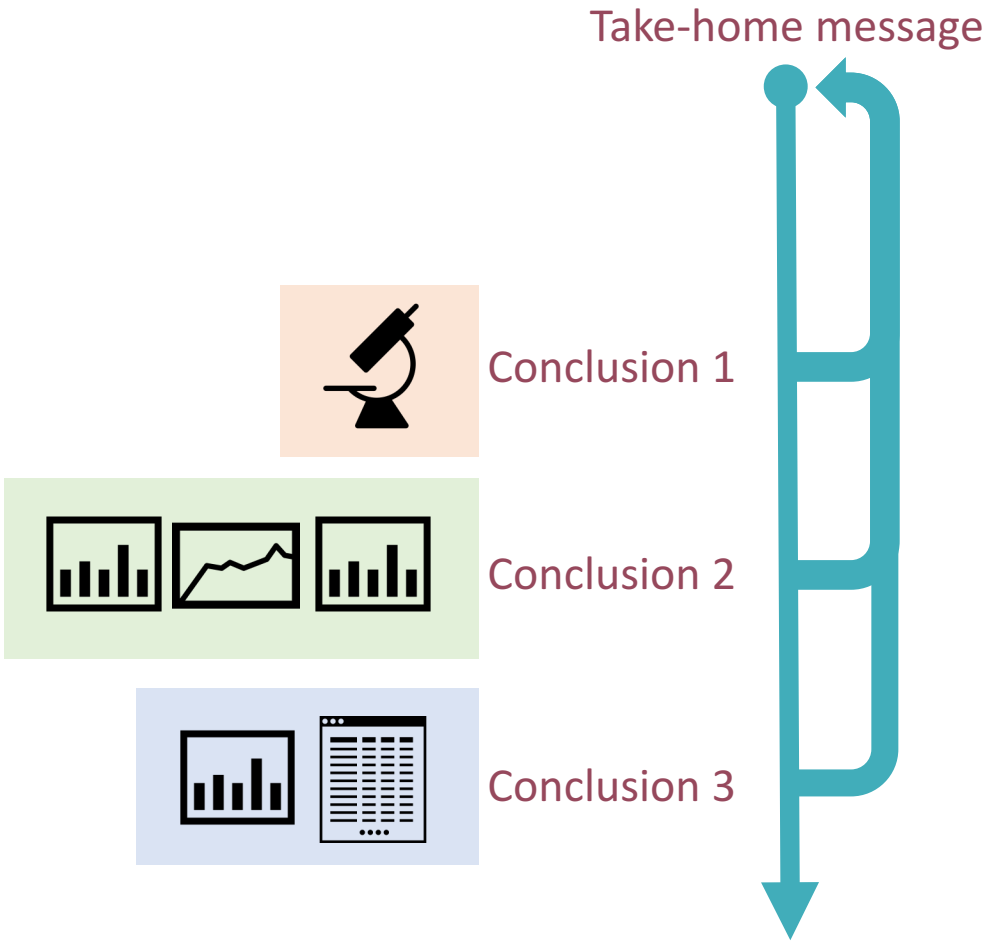
Results: Motivation for key methods you used; high-level summary of methods used to obtain results

Figure captions: high-level description of methods used

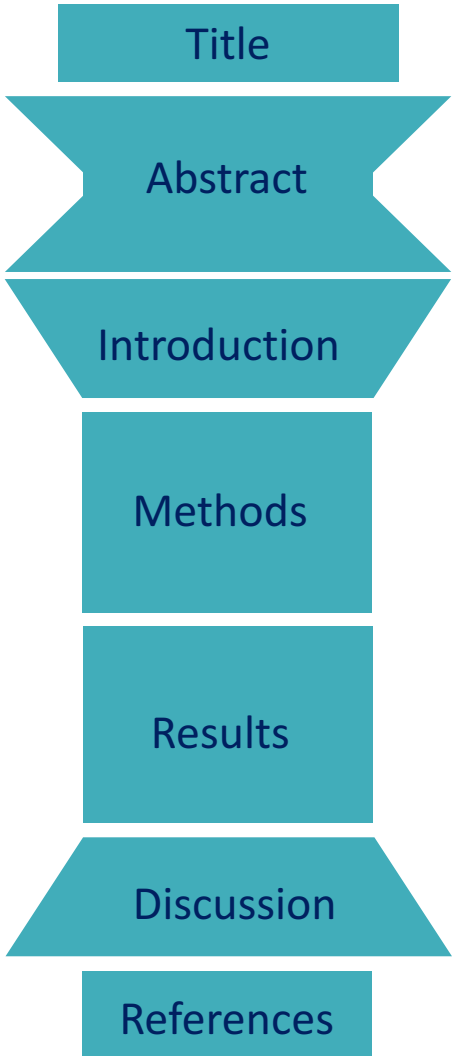
Your paper tells a story about your data



Your story is communicated through your figures



The sections of your paper answer different questions



What is your story?
Why should I care?

What do I need to know to understand your story?

How did you do your experiments?

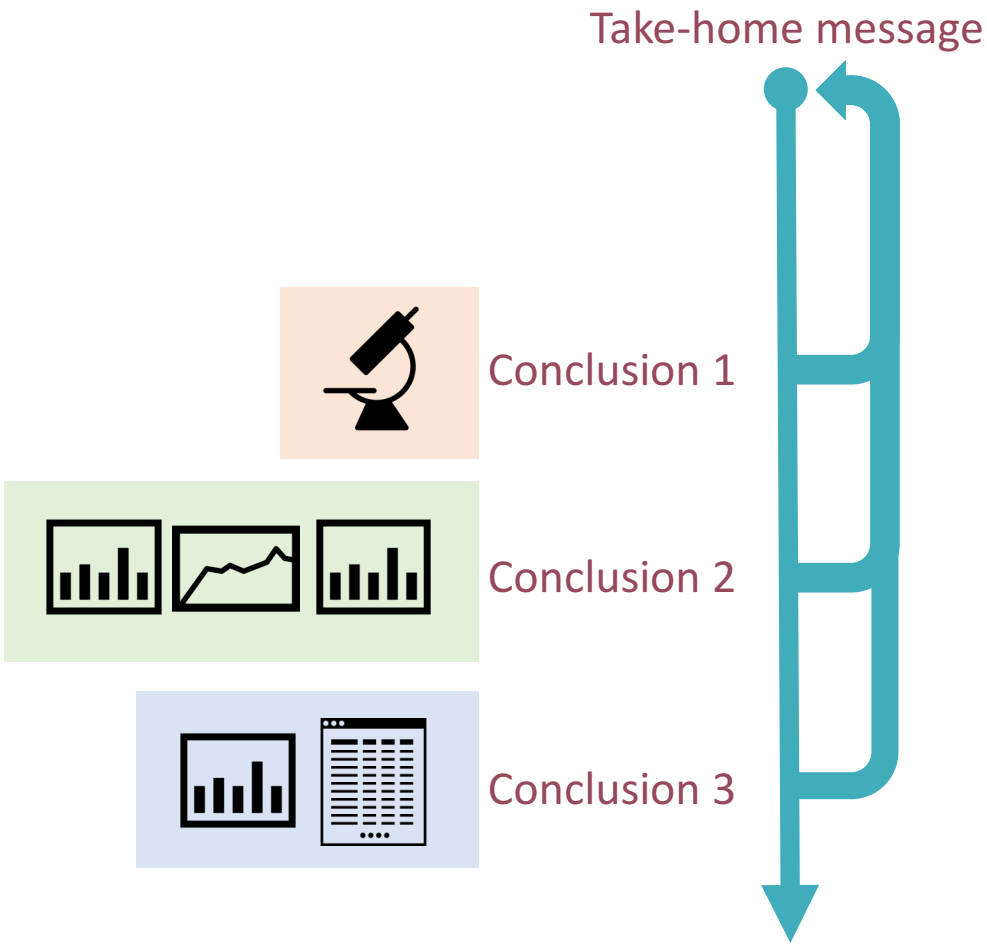
What did you find?

What does it mean?
Where can we go from here?

Results

What did you find?

Results = rationale + data + conclusion



Results = rationale + data + conclusion

Conclusion 1

Conclusion 2

Conclusion 3

RESULTS

Subheading 1 summarizes your 1st major conclusion. Corresponding data are in Figure 1.

Paragraph A describes data in Figure 1A.

Paragraph B describes data in Figure 1B.

Paragraph C describes data in Figure 1C.

Subheading 2 ...

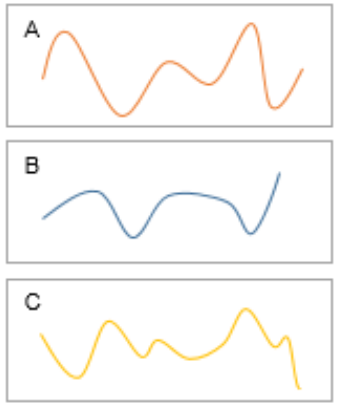


Figure 1

Results = rationale + data + conclusion

Conclusion 2

Title

Paragraphs for each of the key experiments

RESULTS

Subheading 1 summarizes your 1st major conclusion. Corresponding data are in Figure 1.

Paragraph A describes data in Figure 1A.

Paragraph B describes data in Figure 1B.

Paragraph C describes data in Figure 1C.

Subheading 2 ...

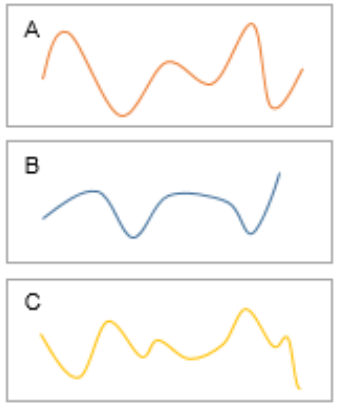
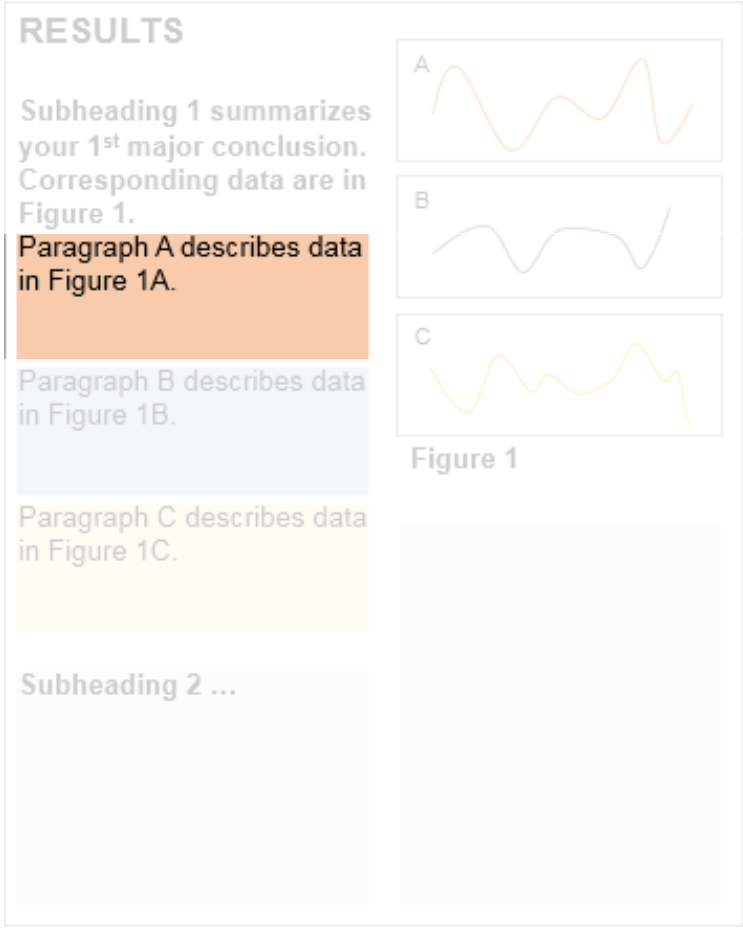


Figure 1

Results = rationale + data + conclusion



Results = rationale + data + conclusion

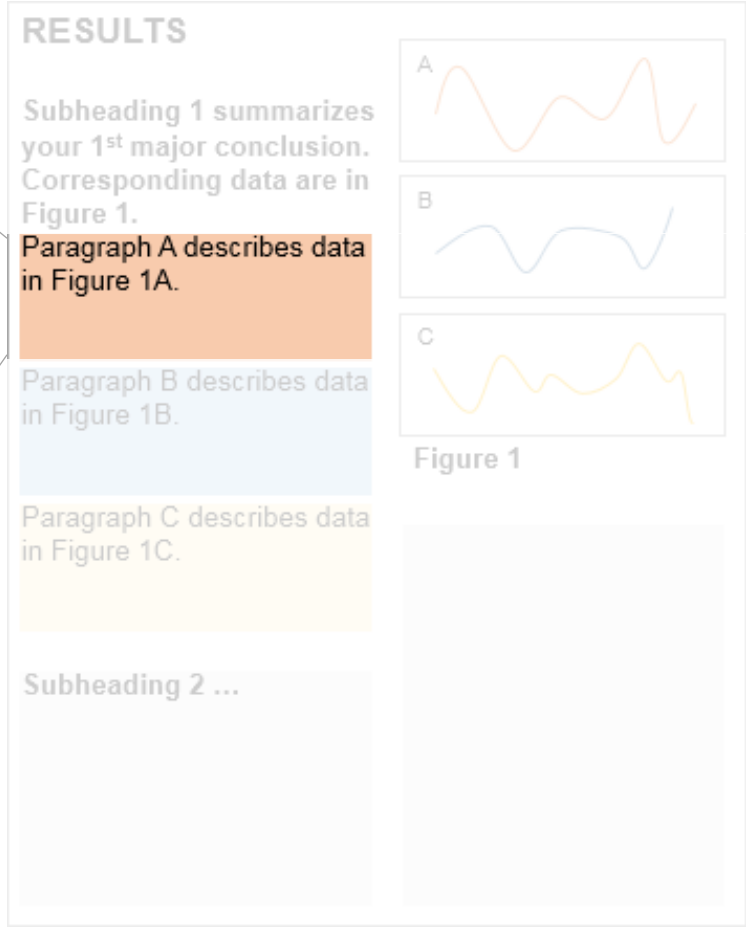
In order to determine *X*, *Y* was performed, showing *Z* major results.

Data + conclusions

- pro, then con
- most to least important
- experiment vs. control

Transition sentence

- re-summarize findings
- justify movement to next experiment or hypothesis



Results: Discuss minimal essential data.

Maximize signal-to-noise.

Include in your paper

- The experiment or dataset that is the **strongest proof** of your conclusion.
- Parts of your chosen dataset might contradict your main conclusion, or support one claim but not another.
- Discuss all parts of a figure in your results section.

Results: Show minimal essential data.

Maximize signal-to-noise.

Include as Supplementary Information

Experiments or datasets that…

- Also support your conclusion but are not the strongest proof

  method is less validated
  data are less statistically significant
  data are less intuitive to interpret

- Were run to validate methods

- Were run to rule out alternative hypotheses

Results: Follow the Herskowitz Rule

time spent describing an individual result $\propto$ **importance** of that result to the paper's main conclusion

Ira Herskowitz, UCSF

Results: The heading of each result section should reflect the message of that figure

The RuvC-like Domain of Cpf1 Mediates RNA-Guided DNA Cleavage

What do you notice about this title?

Figure 4. Catalytic Residues in the C-Terminal RuvC Domain of FnCpf1 Are Required for DNA Cleavage

Zetsche et al, 2015. *Cell*, 163.

Results: Example of motivation of experiment

The RuvC-like domain of Cpf1 retains all of the catalytic residues of this family of endonucleases (Figures 4A and S4) and is thus predicted to be an active nuclease.

What do you think they will test in this experiment?

Zetsche et al, 2015. *Cell*, 163.

Results: Example of motivation of experiment

The RuvC-like domain of Cpf1 retains all of the catalytic residues of this family of endonucleases (Figures 4A and S4) and is thus predicted to be an active nuclease. Therefore, we generated three mutants— FnCpf1(D917A), FnCpf1(E1006A), and FnCpf1(D1225A) (Figure 4A)—to test whether the conserved catalytic residues are essential for the nuclease activity of FnCpf1.

Zetsche et al, 2015. *Cell*, 163.

Results: Example of data and conclusions

We found that the D917A and E1006A mutations completely inactivated the DNA cleavage activity of FnCpf1, and D1255A significantly reduced nucleolytic activity (Figure 4B).

Zetsche et al, 2015. *Cell*, 163.

Results: Example of data and conclusions

We found that the D917A and E1006A mutations completely inactivated the DNA cleavage activity of FnCpf1, and D1255A significantly reduced nucleolytic activity (Figure 4B). These results are in contrast to the mutagenesis results for Streptococcus pyogenes Cas9 (SpCas9), where mutation of the RuvC (D10A) and HNH (N863A) nuclease domains converts SpCas9 into a DNA nickase […].

Zetsche et al, 2015. *Cell*, 163.

Results: Example of data and conclusions

We found that the D917A and E1006A mutations completely inactivated the DNA cleavage activity of FnCpf1, and D1255A significantly reduced nucleolytic activity (Figure 4B). These results are in contrast to the mutagenesis results for Streptococcus pyogenes Cas9 (SpCas9), where mutation of the RuvC (D10A) and HNH (N863A) nuclease domains converts SpCas9 into a DNA nickase […]. These findings suggest that the RuvC-like domain of FnCpf1 cleaves both strands of the target DNA, perhaps in a dimeric configuration. Interestingly, size-exclusion gel filtration of FnCpf1 shows that the protein is eluted at a size of 300 kD, twice the molecular weight of a FnCpf1 monomer (Figure S2B).

Zetsche et al, 2015. *Cell*, 163.

Results: Example of overall structure

In order to determine *X*, *Y* was performed, showing *Z* major results.

Data + conclusions

pro, then con
most to least important
experiment vs. control

Transition sentence

re-summarize findings
justify movement to next experiment or hypothesis

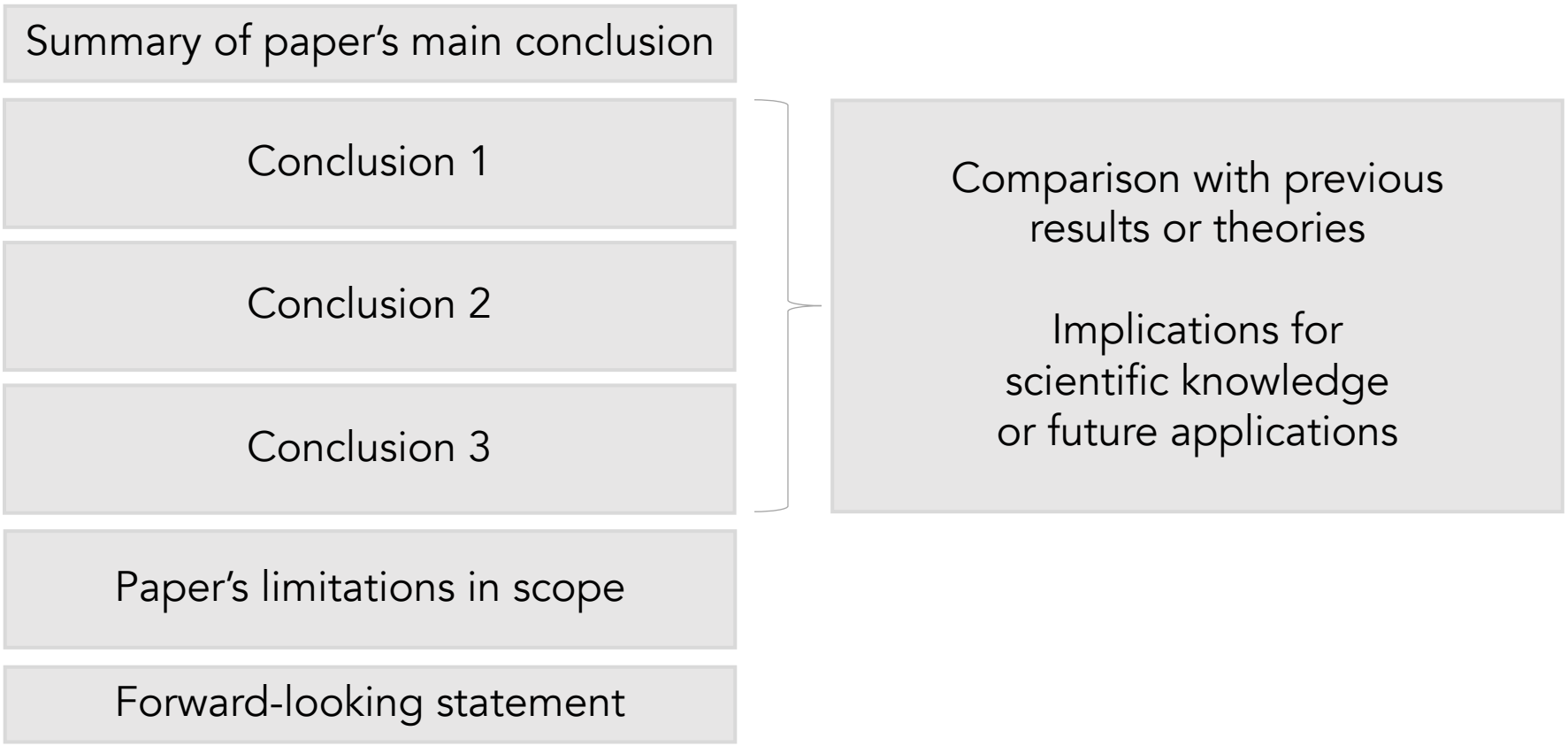
The RuvC-like domain of Cpf1 retains all of the catalytic residues of this family of endonucleases (Figures 4A and S4) and is thus predicted to be an active nuclease. Therefore, we generated three mutants— FnCpf1(D917A), FnCpf1(E1006A), and FnCpf1(D1225A) (Figure 4A)—to test whether the conserved catalytic residues are essential for the nuclease activity of FnCpf1. We found that the D917A and E1006A mutations completely inactivated the DNA cleavage activity of FnCpf1, and D1255A significantly reduced nucleolytic activity (Figure 4B). These results are in contrast to the mutagenesis results for Streptococcus pyogenes Cas9 (SpCas9), where mutation of the RuvC (D10A) and HNH (N863A) nuclease domains converts SpCas9 into a DNA nickase (i.e., inactivation of each of the two nuclease domains abolished the cleavage of one of the DNA strands) (Jinek et al., 2012; Gasiunas et al., 2012) (Figure 4B). These findings suggest that the RuvC-like domain of FnCpf1 cleaves both strands of the target DNA, perhaps in a dimeric configuration. Interestingly, size-exclusion gel filtration of FnCpf1 shows that the protein is eluted at a size of 300 kD, twice the molecular weight of a FnCpf1 monomer (Figure S2B).

Zetsche et al, 2015. *Cell*, 163.

Discussion

What does it all mean?

Speculation and interpretation belongs in Discussion, not Results.



The Discussion should start with a summary of the main message/conclusion

Summary of paper's main conclusion

1 or 2 sentences

Reiterate your "here we show"

DISCUSSION

In this work, we characterize Cpf1-containing class 2 CRISPR systems, classified as type V, and show that its effector protein, Cpf1, is a single RNA-guided endonuclease. Cpf1 substantially

A successful Discussion answers questions for both experts and non-experts.

Comparison with previous results or theories	*How do you account for results that contradict the rest of the field? How does it connect with other work?*
Scientific or engineering implications	*How will this work impact the field or people or the world?* No more than 1 degree of speculation
Paper's limitations in scope	*How do you explain confusing or complicated results?*

Discussion builds from the results

Comparisons?
Implications?
Limitations?

Particular phrases that would not be in other sections?

Cpf1 substantially differs from Cas9 – to date, the only other experimentally characterized class 2 effector – in terms of structure and function and might provide important advantages for genome-editing applications. Specifically, Cpf1 contains a single identified nuclear domain, in contrast to the two nuclease domains present in Cas9. The results presented here show that, in FnCpf1, inactivation of RuvC-like domain abolishes cleavage of both DNA strands. Conceivably, FnCpf1 forms a homodimer (Figure S2B), with the RuvC-like domains of each of the two subunits cleaving one DNA strand. However, we cannot rule out that FnCpf1 contains a second yet-to-be identified nuclear domain. Structural characterization of Cpf1-RNA-DNA complexes will allow testing of these hypotheses and elucidation of the cleavage mechanism.

Comparisons?
Implications?
Limitations?

Differences in language?

Cpf1 generates a staggered cut with a 5’ overhang, in contrast to the blunt ends generated by Cas9. This structure of the cleavage product could be particularly advantageous for facilitating non-homologous end joining (NHEJ)-based gene insertion into the mammalian genome. Being able to program the exact sequence of a sticky end would allow researchers to design the DNA insert so that it integrates into the genome in the proper orientation. Specifically, in non-dividing cells, in which genome editing via homology-directed repair (HDR) mechanisms is especially challenging, Cpf1 could provide an effective way to precisely introduce DNA into the genome via non-HDR mechanisms.

The Discussion often ends with a look at the future

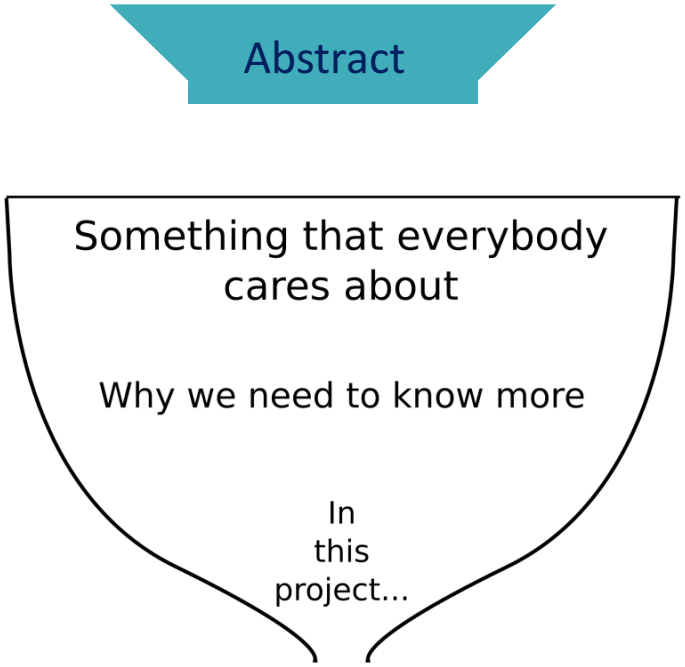
The natural diversity of CRISPR systems provides a wealth of opportunities for understanding the origin and evolution of prokaryotic adaptive immunity, as well as for harnessing potentially transformative biotechnological tools. There is little doubt that, beyond the already classified and characterized diversity of the CRISPR-Cas types, there are additional systems with distinctive characteristics that await exploration and could further enhance genome editing and other areas of biotechnology as well as shed further light on the evolution of these defense systems.

Introduction

What do I need to know to understand your story?

Introduction = Why did you do this research?

- Your research taught you something, right?
- Introduction convinces the reader that this knowledge is worth having
- background + knowledge gap + here we show



Introduction: Clearly establish your central question and take-home message

- Clearly define the knowledge gap/central question of the study and follow with a clear hypothesis.
- Very briefly summarize the key results & conclusions of the paper.

General background
Specific background

Knowledge gap,
Unknown

HERE WE SHOW

Results
Implication, Significance

Introduction: Clearly establish your central question and take-home message

Given the broad applications of Cas9 as a genome engineering tool (Hsu et al., 2014; Jiang and Marraffini, 2015), we sought to explore the function of Cpf1-based putative CRISPR systems.

Here, we show that Cpf1-containing CRISPR-Cas loci of Francisella novicida U112 encode functional defense systems capable of mediating plasmid interference in bacterial cells guided by the CRISPR spacers.

Zetsche et al, 2015. *Cell*, 163.

Introduction: Briefly summarize your key results

Unlike Cas9 systems, Cpf1-containing CRISPR systems have three features. **First,** Cpf1-associated CRISPR arrays are processed into mature crRNAs without the requirement of an additional trans-activating crRNA (tracrRNA) (Deltcheva et al., 2011; Chylinski et al., 2013). **Second,** Cpf1-crRNA complexes efficiently cleave target DNA proceeded by a short T-rich protospacer-adjacent motif (PAM), in contrast to the G-rich PAM following the target DNA for Cas9 systems. **Third,** Cpf1 introduces a staggered DNA doublestranded break with a 4 or 5-nt 50 overhang. **To explore the suitability of Cpf1 for genome-editing applications,** we characterized the RNA-guided DNA-targeting requirements for 16 Cpf1-family proteins from diverse bacteria, and we identified two Cpf1 enzymes from Acidaminococcus sp. BV3L6 and Lachnospiraceae bacterium ND2006 that are capable of mediating robust genome editing in human cells.

Zetsche et al, 2015. *Cell*, 163.

Introduction: Identify the significance of your findings

Collectively, these results establish Cpf1 as a class 2 CRISPR-Cas system that includes an effective single RNA-guided endonuclease with distinct properties that has the potential to substantially advance our ability to manipulate eukaryotic genomes.

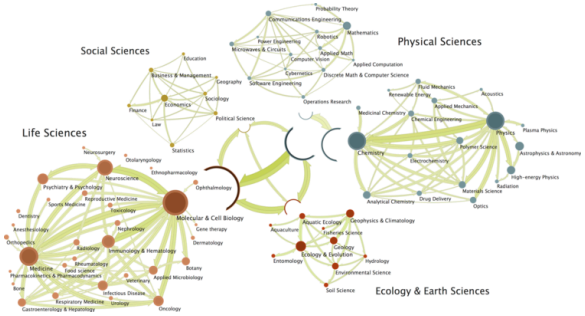
Zetsche et al, 2015. *Cell*, 163.

Additional Tips

Paragraph structure helps you and your reader

- One paragraph = one thought.
- 1st sentence summarizes this thought, last sentence reiterates.
- Elaborate in a logical order:
  - pro then con
  - most to least important evidence
  - chronological (be careful!)

References connect your paper to the research ecosystem



- Built over the course of the paper
- Make sure you include papers that…
  - reach conflicting conclusions
  - are from your competitors
  - were published during the course of your work

  (Reviewers will be looking)
- Your abstract will not have references, all other sections should!

Revising is ESSENTIAL!

- Do not try to write this paper in one day.
- **Outline**/draft the sections, then set aside the paper for several days.
- If you get stuck: outline, write topic sentences, work on the next section, look at examples
- Get feedback:

  peers, instructors, Comm Lab Fellows!

Assignment or paper questions?

20% of course grade (full rubric on wiki)

Title and Abstract		10%
Introduction	2-3 p.	10%
Methods	3-4 p.	20%
Results and Figures	4-5 p.	50%
Discussion	2-3 p.	10%

(12pt., double-space except abstract, max. 14 pages)