

1 Capra Lab Manuscript Template for Cell Press Journals

2 Your Name^{1,2}, Additional Co-Author(s)⁴, and John A. Capra^{1,2,3,5,*}

3 ¹Bakar Computational Health Sciences Institute, University of California, San Francisco, CA, USA

4 ²Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA

5 ³Biomedical Informatics Graduate Program, University of California San Francisco, San Francisco, CA, USA

6 ⁴Additional Co-Author Department, University, City, State, USA

7 ⁵Lead contact

8 *Correspondence: tony@capralab.org

9

10 **SUMMARY**

11 Cell Press articles begin with a “summary” rather than an abstract. This should be 150 words
12 or less. Per *Cell Genomics*: “An effective summary includes the following elements: (1) a
13 brief background of the question that avoids statements about how a process is not well un-
14 derstood; (2) a description of the results and approaches/model systems framed in the context
15 of their conceptual interest; and (3) an indication of the broader significance of the work. We
16 discourage novelty claims (e.g., use of the word “novel”) because they are overused, tend not
17 to add meaning, and are difficult to verify. Please do not include references in the summary.”

18 INTRODUCTION

< 1 page, approx. 3 paragraphs

Begin by framing the broad field in which you are working very briefly. You do not need to provide a full review of this field. Strategically cite a few reviews. Just identify the main relevant work and cite review articles for those who need more context. Get to the critical gap/problem as quickly as possible.

Then elaborate on the specific problem that you are working on and explain why solving it is important (if you haven't already). Presumably, other folks have also looked at this problem. Briefly review what they have found and/or the relevant hypotheses. Set up why there is a gap/problem (e.g., *We need to test this new hypothesis, there are better machine learning methods, etc.*).

Then explicitly detail what questions/hypotheses/etc you are trying to answer/test. Feel free to use a list. Describe your innovative approach—you may introduce such an approach in the paragraph(s) directly before this one. Note major predictions here and reemphasize the significance. Someone should be able to read only this paragraph and have a good sense of what the study/you/the team has tried to accomplish. (You can bullet point or list out the aims/problems/questions if you really want to draw attention to them).

19 Introductions for Cell Press journal articles should not include any subsections. A quick note
20 about citations: they should occur in the form of superscript numbers, numbered in order of ap-
21 pearance¹⁻³. Use the `supercite` command immediately after the relevant text to write citations
22 using the citation keys from the `.bib` file. You can also explicitly refer to the author(s)/paper us-
23 ing the following convention: “Enard et al.¹ quantified gene expression in four primate species”.
24 Cell Press Journals use journal abbreviations in the References section. If you export a `.bib` file
25 from a reference manager (e.g., Mendeley, Zotero), ensure that journal abbreviations are in-
26 voked. Abbreviated components of the journal title should also include a period (see the Sholtis
27 et al. 2010 reference below, Trends in Genetics → Trends Genet.). These can be added man-
28 ually in your reference management software library or code can be added in the `0main.tex` file
29 to do this automatically.

30 **First paragraph(s)** What is the problem? Why do we care? First sentence can be a gen-
31 eral truth or general limitation that your audience accepts. E.g., *While differences between*
32 *archaic and modern humans are well described, there is a poor understanding of the genetic*
33 *mechanisms underlying such phenotypic differences.* or *Most aspects of archaic hominin biol-*
34 *ogy cannot be directly studied due to their lack of preservation in fossils.* Don't waste the first
35 sentence with a definition.

36 **Middle paragraph(s)** What have others found about this specific problem? What are the
37 current hypotheses and support for each? and/or what data have been previously analyzed?
38 What are the gaps? E.g., *One such mechanism may be alternative splicing, but we cannot*
39 *obtain such data from extinct taxa.* This part of the Intro may require a few paragraphs, but you
40 do not need to fully review the history of work on the question. Focus on the current state of
41 the field.

42 **Last paragraph** Overview of the major questions addressed here and why they are impor-
43 tant. *Here, we investigated X using dataset Y and dataset Z. We tested the A hypothesis. We*
44 *also considered this other obscure hypothesis. We predicted x, y, and z.* E.g., *Here, we lever-*
45 *age a new deep neural network that can predict splice altering variants from sequence alone*
46 *to examine such variants in archaic hominins.*

47 RESULTS

As long as needed, but usually 4-7 subsections

Divide this section into subsections with a logical progression from one analysis to another. The subsection titles should provide an outline of the results that enables skimming. Thus, try to make each results subsection title state a brief result. Do not feel constrained to follow the order in which you did the analyses. Tell the most logical and easy to follow story that highlights the most exciting parts of your findings early and ends with a finding that leads most towards future directions (or is a preliminary finding toward the new direction). Make sure to reference and use all data presented in figures and tables. Negative results are useful. Actions should be written in past-tense, while statements/results are in the present-tense. E.g., (*We TESTED the correlation of X vs. Y, and we OBSERVED that they ARE correlated*).

Each results section should contain the following: context, approach, result, details, conclusion (see below for template). Repeat as needed. Depending on the relatedness of results, it is possible to have multiple of these 'modules' under a single heading. Conversely, you may have more headings than figures if, for example, multiple sections are needed to describe a multi-panel figure. (In this case, consider if the figure should be split up or not.)

48 Subsection titles should not end in a period

49 Per *Cell Genomics*: "This section should be divided with subheadings. In our view, good sub-
50 headings convey information about the findings, so we encourage you to be specific. For ex-
51 ample, say "Factor X requires factor Y to function in process Z" rather than "Analysis of factors
52 X and Y using approach Q." We recommend that you use similar language in your figure titles
53 for clarity and structural harmony."

54 Can Include Optional Initial Paragraph

55 Concise overview of what the reader needs to know about the methods or context (but don't
56 take too long to get to the real results). *The goal of this work is to demonstrate "X"*. If you use
57 a lot of complicated or easily confused terminology, you can define it here to set the stage.

58 Declarative result statement

59 **Context** Set up the context and logic for the experiment/analysis. Clearly state the question
60 and hypothesis. Sometimes this will already have been set up by the previous section and can
61 be skipped.

62 **Approach** What was the experiment/analysis, briefly (cite methods section if needed)? De-
63 scribe key factors of the approach, especially details that apply specifically to this result but not
64 to others using the same method (i.e. specific simulation parameters that vary across results
65 sections).

66 **Result** Clearly state the MAIN result of the experiment/analysis and any relevant statistics. Give
67 a quantification of the effect size, and if you performed a significance test, give the quantifica-
68 tion of significance (e.g., p-value or q-value) and test used in parentheses. E.g., *The mean
69 expected 3D divergence is 78% higher than the observed 3D divergence (P = 1.8 × 10⁻⁴⁸,
70 t-test)*. Reference relevant figure panels.

71 **Details** Elaborate on any results that modulate or demonstrate the robustness of your main
72 result. These usually reference supplemental figures that support your main result.

73 **Conclusion** Conservative conclusions drawn from the result. What do these data say about
74 your hypothesis? *This suggests...*

75 **Example** *Given X, we hypothesized that Tony is great. To evaluate/quantify Tony's greatness,*
76 *we did X. We found Y (big picture overview of result) (Main text figure reference). We con-*
77 *trolled for Z. Tony's greatness was robust even when considering W (Supplemental figure(s)*
78 *reference). This suggests that Tony's greatness knows no bounds.*

79 **Example results section to demonstrate figure references and placement**

80 You should insert main text figures and tables into the document as soon as possible after
81 the first mention in the text. We suggest you do this for initial submission even if the journal
82 requests that you put your figures and captions at the end in a separate section. Your reviewers
83 will thank you! Here is an example of how to reference a full figure in-text (**Figure 1**) or just a
84 sub-panel (**Figure 1A**). You might also want to reference two figures at the same time like this
85 (**Figures 1, S1**). Sometimes you might reference a range of figures like these pair of *Capra*
86 shown in **Figures S1–S2**. You can also reference a range of subpanels like **Figures 1A–1B**.
87 You can do the same thing for tables (**Table S1**).

A

- **Title.** The figure title should tell a succinct result and lead the reader to the figure's conclusion (when possible).
- **Visually clear:** Remove all "chartjunk": unnecessary or confusing visual elements found in a figure that do not improve the message (in the best case) or add confusion (in the worst case). For example, chartjunk may include the use of too many colors, too many labels, gratuitously colored backgrounds, useless grid lines, etc.
- **Text style.** Figure and caption should be the same font as the main text. The font should be Helvetica or Arial.
- **Axis labels.** Always include axis labels.
- **Abbreviations.** All abbreviations defined in the caption.
- **Symbols/colors.** All symbols or colors defined - consistent throughout paper when possible
- **Dataset:** Specify data set used if not clear.
- **Sample size** ($N = X$). Some places require n to be lowercase for sample (rather than population) sizes
- **Raw data points.** Show whenever possible (especially if $n < 10$)
 - For scatter plots, point to one dot and have a small description (e.g., "one neuron")
- **Confidence intervals or SE bars.** Clearly labeled, for example: "Data are presented as mean values \pm SEM"
 - If using box plots, tell how they were made ("Boxes represent the IQR and outliers are more extreme than $1.5 \times \text{IQR}$.")
- **Statistical significance** ($P = X$ or $q = X$). Include details about the test used, one-sided vs two-sided (maybe degrees of freedom too)
- **Tables.** Avoid tables or do not overuse

B

- Generally, a single column width measures 85-88 mm and a double column width measures 170-180 mm.
- Max height around 225mm (but that is with caption) (A pdf page is 210 x 276 mm)
- Font size should be 5-14pt
 - Axis labels 8-9pt
 - Axis numbering 7-8pt
 - Subpanel labels = 13pt bold
- Axis lines should be 0.75 pt (ungroup anything that looks weird)
- Axis ticks should be 1mm (short ticks 0.85, long ticks 1.35) (1.1, 0.6)
- All lines should be wider than 0.25pt (should probably be 0.5+)
- If things are weird, embed font
- Check spelling
- Remove masked layers (change to clip)

Figure 1. Figure title which should not end in a period

(A) Succinct caption for Panel A. Note that figure captions will not split across pages.

(B) Caption for panel B. Add additional subpanel captions as needed depending on the figure.

Tips for captions and figures: All figures and tables should be inserted into the document as soon as possible after the first mention in the text. Make sure to number them sequentially as they are referenced in the text and provide a succinct descriptive caption. If there are multiple parts to a figure, give clear capital letter labels (A, B, C, etc.) to each panel. Make sure each figure and sub-figure are referenced somewhere in the text! The images should be saved in vector formats (such as PDF or EPS) to maintain high resolution. The exceptions to this are: 1) if you are including a photograph or 2) if you are plotting a file with thousands of overlapping individual data points that would each be represented in the vector file (use a bitmap format here). The title of the figure caption should state

the overarching result. (This will often be similar to a results section title.) Captions should interpret the results briefly. The figures and captions should be able to be “stand alone” from the main text with enough context and methods for interpretation. (Many readers will only look at the figures and captions.) Subplots should contribute to one major overarching scientific point. If you can’t come up with one conclusion/title for a figure, consider if it should be two figures. If two figures show the same conclusion, one should probably go to the supplement (e.g., demonstrating that a results replicates across cell types or with a different control).

88 DISCUSSION

Approx 1 page

The discussion should be a broad overview of the significance of your findings to the community you are addressing. You can mirror the results section, but you don’t have to. You should start with a brief summary of the main findings and then a paragraph for the accompanying context and interpretation of each of the big picture conclusions or contributions. Then you should address high-level limitations of the study (not just small limitations that you might address in your results/methods) Finally, end with future directions on a positive note. How can people use your model/framework results? This positive-critical-positive is a compliment sandwich.

Everything discussed should be within the frame of reference of your paper’s major conclusions or contributions. You will likely want to address a shortlist of the key papers your paper speaks to, but this is not a literature review. Even if you don’t talk extensively about a paper, you can still cite it if it is relevant to your conclusions. (There is little cost to being free with your citations!)

Contrary to popular belief, you can bring up a new result in the discussion, especially if it is preliminary or a response to reviewer critiques. Save this for findings that are in service to the discussion, rather than just a main finding of a paper. You can also use sub-sections if it is helpful in framing but this is somewhat non-standard.

Optional: Consider proposing a model or framework in your discussion that integrates or contextualizes your results. This can be accompanied by a schematic or “model” figure. This is a great way to emphasize the contribution of your study and guide future work.

89 Per *Cell Genomics*: “The discussion should explain the significance of the results and place
90 them into a broader context. It is often helpful to the reader to indicate the directions in which
91 the work might be built on going forward. It should not be redundant with the results. The
92 discussion may contain subheadings and can be combined with the results section.”

93 **First paragraph** Brief summary of the paper that sets up the discussion of the major contri-
94 butions. You can start by restating the major gap that you addressed (*Despite X, we previously*
95 *had a limited understanding of Y. Here, we show/demonstrate Z. Together, these findings high-*
96 *light/provide...*) This can parallel the last paragraph of your intro (where you outline the big
97 picture questions and their context for the field) or parallel the results, but this is not required.

98 **Middle Paragraph(s)** (1-4 paragraphs) Major contribution/finding/conclusion and its con-
99 text. What does your result/contribution mean? Who does it speak to and why do they care?
100 What are the implications? Were your hypotheses correct, surprising, contradictory? If the
101 results are not conclusive, why not? How would you perform further work to get to a more
102 conclusive result?

103 **Penultimate paragraph(s)** (1-2 paragraphs) Limitations. Identify the broad categories/
104 themes of limitations and don’t get too particular about the details of each limitation. (i.e., need
105 more samples/ data, model misspecification or over-simplification). Keep it succinct, don’t apol-
106 ogize, but show that you have thought about caveats. Should segue into the future directions.
107 *Although our approach provides many novel benefits, it also has limitations that we hope future*
108 *work will address.*

109 **Last paragraph** Future directions and final big picture conclusion. This could be broken out
110 into a separate conclusions section. *As X happens, we anticipate that Y will facilitate under-*
111 *standing of Z which allows for W.* Why is this exciting for your audience? Where are the next
112 steps? Note that this is could possibly result in scooped, but even if you are thinking of working
113 on a next step, consider putting it out there, because science for all!! *We encourage future*
114 *studies to address X.* Be positive about specific future opportunities inspired by your work. It
115 leaves an impression).

116 **METHODS**

No length limit

Goal: Communicate the data and methods used to produce the results. Start writing this as you are doing your analyses. Objective: (1) To write clearly how you produced all parts of the results and (2) promote reproducibility of the findings. Style: Overall, the methods section should be clearly and thoroughly written. In theory, any knowledgeable reader (e.g., a graduate student working in your area) would be able to reproduce any of the findings just by reading your methods. Organization: Write subsections with clear subheadings for each method. These should be descriptive. Ideally, subsections will be structured so that the reader can “look up” the details for any subsection of the results. Ordering subsections: Subsection order should generally match the order of the results. However, if you repeat the same type of analyses in two separate parts of the paper, you can describe this once and refer to the sections those methods address. Voice: Write in the active past tense (E.g. “We computed..., we intersected..., we quantified...”). Note on publicly available datasets: Include a citation or url link to datasets used, along with the last download date. Citations: if you use it, cite it. There is no ambiguity here.

In some journals, the methods section precedes the results section, while in others it follows the Discussion section. If it comes before, then the text should provide some orientation and motivation. If it comes after, then it can be more of a list. Unless you know it will come first, write the methods section as if it were following the discussion section and then add orienting text later if necessary.

Other tips: Think of the methods section is a cookbook, and each subsection is its own chapter. Some chapters will discuss where to get the finest ingredients, some will discuss specific cooking techniques, and the rest will discuss the recipes that combine ingredients and technique. For each recipe (subsection), you must be clear on which ingredients are needed and the order that the ingredients are assembled.

You do not need to divulge every quotidian analytical detail (e.g., *I first loaded .tsv into a data frame using pandas (v14.0) function pd.read_csv(filename) and pivoted it into a table*). But you should provide every analytical detail relevant to reproducing the final data analyzed (e.g. *We removed all loci mapping to sex chromosomes using the subtract function from BEDTools (v2.2.1; Quinlan and Hall 2010).*)

117 **1st Methods Section**

118 Cell Press journals use STAR methods (<https://www.cell.com/star-authors-guide>). One can
119 format these methods ahead of time but this formatting can also wait until acceptance/revisions
120 as most Cell Press journals do not require formatting for initial submissions.

121 **Data and Code Availability**

122 Identify original data sources and provide links to software in this section.

123 **Acknowledgements**

124 Thanks folks here. Note computational resources and look up specific language to use. For
125 example, research completed using Wynton at UCSF can use language from [https://wynton](https://wynton.ucsf.edu/hpc/about/citation.html)
126 [.ucsf.edu/hpc/about/citation.html](https://wynton.ucsf.edu/hpc/about/citation.html). Include any relevant funding agencies and specific grant
127 numbers here. Ask collaborators for information they want to include here.

128 **Author Contributions**

129 Use the CRediT Taxonomy to indicate author contributions. See [https://www.cell.com/pb/asse](https://www.cell.com/pb/assets/raw/shared/guidelines/CRediT-taxonomy-1430242873507.pdf)
130 [ts/raw/shared/guidelines/CRediT-taxonomy-1430242873507.pdf](https://www.cell.com/pb/assets/raw/shared/guidelines/CRediT-taxonomy-1430242873507.pdf).

131 **Competing Interests**

132 The authors declare no competing interests.

133 **References**

- 134 1. Enard, W., Khaitovich, P., Klose, J., Zöllner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe,
135 K., Muchmore, E., Varki, A., Ravid, R., et al. (2002). Intra- and Interspecific Variation in Pri-
136 mate Gene Expression Patterns. *Science* 296, 340–343. [https://doi.org/10.1126/science.](https://doi.org/10.1126/science.1068996)
137 [1068996](https://doi.org/10.1126/science.1068996).
- 138 2. King, M.-C. and Wilson, A. C. (1975). Evolution at Two Levels in Humans and Chim-
139 panzees. *Science* 188, 107–116. <https://doi.org/10.1126/science.1090005>.
- 140 3. Sholtis, S. J. and Noonan, J. P. (2010). Gene Regulation and the Origins of Human Bio-
141 logical Uniqueness. *Trends Genet.* 26, 110–118. <https://doi.org/10.1016/j.tig.2009.12.009>.

142 Supplemental Text

143 Some supplemental text that doesn't fit in the main text or provides extra detail can go here.

144 We will use this space to outline some general tips.

145 Article tips, tricks, and nitpicks

- Formatting/scientific corrections
 - The first time you report a p-value, give the test used.
 - Report p values as ($P = 0.\#$), be careful with ($P = 0$) (use $< \#$ when P is very small)
 - Change “mutation” to “variant”...maybe change “SNP” or “SNV” to “variant(s)” if specificity is not important (i.e., don't forget that indels and CNVs/SVs are also types of genetic variants)
 - For genomic coordinates/distances, the format is # kb or # Mb (space between number and unit, not # Kb or # mb)
 - Use American spelling conventions.
- Organization
 - Put answers to the “why” question in the discussion
 - Bring important findings to the first sentence of a paragraph/section
 - With section headers, figure titles, and paper titles, try to state a result if possible (rather than a method, e.g. “Neanderthals were bald” rather than “Investigating hair patterns of Neanderthals”)
- Wordsmithing
 - Where you can avoid being vague, be specific: instead of “establishes the relationship” say the relationship that was established. Instead of “Is incompletely understood” say what is the specific thing that is incompletely understood
 - Get rid of “metadiscourse”. E.g., Change “Our findings XY and Z support previous findings, including studies by Smith et al and Johnson et al, which reported that Neanderthals might have been bald at one point” to “XY and Z support findings that Neanderthals may have been bald (citation)”.
 - Don't claim you're the first ever (or save that for the cover later)
 - Generally speaking, avoid passive voice. If you can add “by zombies” after the verb and it still makes sense...rethink your phrasing
- Word choice
 - Change utilize to use
 - Change Impacted to influenced
 - Delete uses of “those/these” in reference to previous sentence (make sure the subject of the sentence is clear)
 - “Characterize” connotes that you don't really have a hypothesis (try quantify?)
 - Some favorite transition words: thus, nonetheless, furthermore, indeed, therefore
 - Fewer (discrete quantities) vs less (continuous quantities)
 - (e.g.,...) and (i.e.,...) can be helpful
 - “Data” is the plural of “datum”.
 - Avoid personifying inanimate objects (“the gene wants to...”)
 - Splitting infinitives is ok.
 - Contractions are not acceptable in academic writing.
 - Remove: Interestingly, surprisingly
- Punctuation
 - Parenthetical punctuation: clauses and sentences
 - Use an Oxford comma to separate the last and second-to-last elements in a list.
 - Beware of consistency between Hyphens(-), n-dash(–), m-dash(—)
 - With reporting intervals be consistent in how you use dashes and spaces (1-2 vs 1 - 2 vs 1 – 2, etc)
 - Use a comma between independent clauses joined by a conjunction.
 - Use “that” for restrictive clauses (no commas). Use “which” for non-restrictive clauses (use commas).
 - Pluralization rules for abbreviations.



Figure S1. This is a picture of Tony Capra Nobody puts Tony in a supplement...

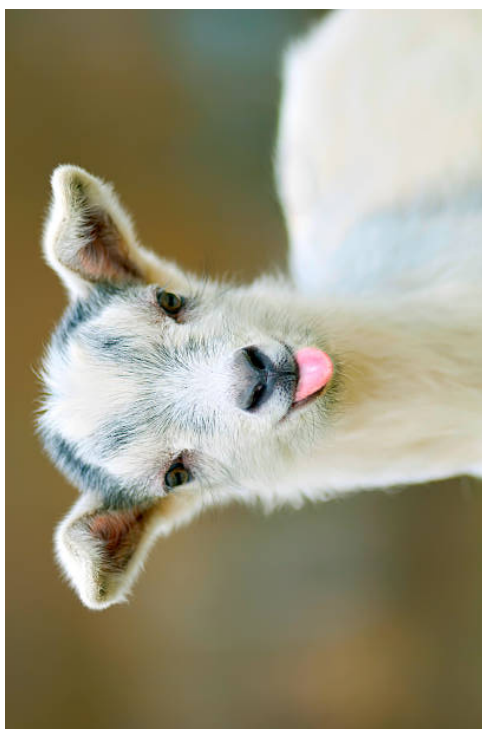


Figure S2. This cute sideways goat wishes you good luck with your manuscript This is how you rotate a figure but not the caption. To rotate the caption with the figure, use the function(s) `generateSidewaysFig` or `generateSidewaysFigSubpanels`.

147 **Supplemental Tables**

Models	Metric 1			Metric 2
	precision	recall	F-score	R@10
model 1	0.67	0.8	0.729	0.75
model 2	0.8	0.9	0.847	0.85

Table S1. Example table. I use `Excel2LaTeX.x1a` to create easy tables. There are also other online websites. Look for one that is booktabs enabled. Here's good advice about making simple readable tables: <https://users.umd.edu/~jbg/images/tables.gif>