Overview

The single purpose of a Methods section in a paper is to explain the data acquisition, QC, processing, and analysis clearly enough so that a sufficiently sophisticated scientist could **replicate your study perfectly and find identical results**.

Because this is the unique purpose of a Methods section, these sections are boring and dry to someone who is not interested in the methods. To those of us who care a lot about Methods, however, these are often the only section we read carefully. This is a recipe book.

Writing style

Writing style for a Methods section is *declarative* and *in past tense*. Say what you did, with no embellishment or motivation:

- "We used the hg18 reference genome sequence, removing non-autosomal chromosomes. We then mapped all reads to this genome sequence using MAQ v0.6.8 (Li et al., 2008) using the default parameters (the default settings allow two mismatches in the first 24 bases of a read)." *Note the version number of the software and the citation.*
- "For each alignment, for each of all six possible reading frames (three on the (+) strand and three on the (-) strand), we calculated a likelihood ratio for a model under which dN/dS is estimated versus a model in which dN/dS is 1 using PAML (Yang, 2007). To format alignments for PAML, gaps in the human sequence were removed, gaps in non-human sequences were converted to Ns, and each of the six possible frames was trimmed to be a multiple of three bases." *Note the clear explanation of method input and output.*

Methods for the experiment should include the following:

- State clearly how your data were acquired (download from what website with what ID number on what date; generated from what experimental protocol?)
- Describe data set "The data set contained 2156 samples and 7,943,454 SNPs on autosomal chromosomes."
- State clearly how your data were QC'd and processed, including citing software and other resources.
- State clearly if you had to reformat or modify any of the underlying data for any reason.

Methods for the analysis should include the following:

- For analysis: statistics used in validation
- Software used for validation and analysis (R packages, Python packages)
- Describe how significance thresholds for p-values and FDR were identified

Methods development

A description of developed methods should include the following:

- A clear definition of the random and observed variables, including dimension and type: "For n samples across p genes, we have a matrix $Y \in \mathbb{R}^{n \times p}$."
- A description of the generative model for a method:
  For $x_i \in \Re^p$ for sample $i \in \{1, \ldots, n\}$ and p genes, and latent variables $z_i \in \{0, 1\}$, we have

$$\begin{aligned} z_i \mid \boldsymbol{\pi} &\sim \quad \text{Mult}(\boldsymbol{\pi}) \\ \mathbf{x}_i \mid z_i = k &\sim \quad \mathcal{N}_D(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \end{aligned}$$

  where k is the number of latent components, and $\boldsymbol{\Sigma}_k$ is a diagonal matrix with σ^2 on the diagonal.
- For parameter estimation: the starting point of the parameters, the updates to each of the parameters, the convergence criteria
- For simulations: as in the generative model, describe how each random variable was generated.

A comparison against related methods should include the following

- For related methods: version, date of download, parameter settings
- how data were processed (e.g., "We thresholded all parameters below 0.05 to 0 to evaluate results against the sparse simulation.")
- always run related methods as smartly as possible without spending too much time fiddling with the parameters (cross validation over a hyperparameter grid is always well-appreciated).
- always set parameters to give them the benefit and process the results accordingly (e.g., "We set the number of clusters to match the known number of clusters in the simulation.")
- describe all (hyper)parameter settings of all methods, including your own.

A key element of this, not included in this text, is to make your data, code, and (I would suggest) your results analysis and processing iPython or KnitR scripts available when this manuscript is sent out. This way you can ensure that other researchers can replicate your data precisely.

Metrics for validation

Although not the focus of your paper, you should be very clear about what metrics you used to quantify or compare results. Writing out equations here removes all uncertainty, or citing specific packages. The application of these should also be stated clearly (e.g., gene-wise, across samples, within each population, within each cluster).

For example: "Pearson's Correlation between gene expression levels x and y was computed as $cor(x, y) = cov(x, y)/\sqrt{var(x)var(y)}$."

Methods resources

I googled "How to write a Methods section for a scientific research paper" and found these helpful documents:

- Purdue Online Writing Lab: https://owl.english.purdue.edu/owl/resource/670/04/
- Dartmouth Professor's ideas: http://cancer.dartmouth.edu/documents/pdf/methods_section.pdf
- Elsevier: http://www.elsevier.pt/en/revistas/revista-portuguesa-pneumologia-320/artigo/how-to-write-scientific-paper-writing-the-methods-section-90026890