Appendix A: Peer-reviewed papers and thesis chapters

Figure A.1 shows how my three peer-reviewed papers informed the research questions that arose while working on this thesis. The papers include a gene ontology (GO) paper [106], a literature search paper [105], and a mouse paper [119].

A.1 Gene ontology paper

The work done on my gene ontology paper [106] addressed open Gene Ontology Enrichment Analysis (GOEA) questions:

- How correct are the GOEA results?
- How many expected Gene Ontology (GO) results are missing?
- What affects how many results are missing?
- Should I use the "propagate counts" mode?
- How do I summarize the GO results?
- How do I visualize the GO results?

After conducting the research required to write this paper, I felt confident about the portions of my thesis that use GOEAs.

A.2 Literature search paper

Knowledge acquired through work on the literature search paper [105] helped inform the scant literature search results at the beginning of my thesis. It was interesting to find that many highly esteemed authors that appeared in the final literature search of the thesis did not appear in the early search using Google Scholar.

Scientific literature search results should be reproducible [22] [80] [79]. PubMed provides reproducible literature searches, while Google Scholar does not. My literature search results for this thesis have dramatically improved through my PhD time by combining citation data from the National Institutes of Health Open Citation Collection [92], the PubMed web search experience, programmatic access to PubMed's data using NCBI's E-Utils libraries [146], and revision management of the literature search results using git, all coordinated with my Python package [105] found at https://github.com/dvklopfenstein/pmidcite.

A.3 Mouse paper

Any analysis or plot in this thesis can also be performed with mice and flies and can be expanded to additional species. My work on the Cyclin-D1/G9a paper [119] provided practical data for working with the mouse genome and comparing mouse clusters to clusters in the human genome. It also provided experience with comparing LAD region location data annotated on an older mouse genome to the cyclin-D1/G9a molecule binding point on a newer mouse genome.

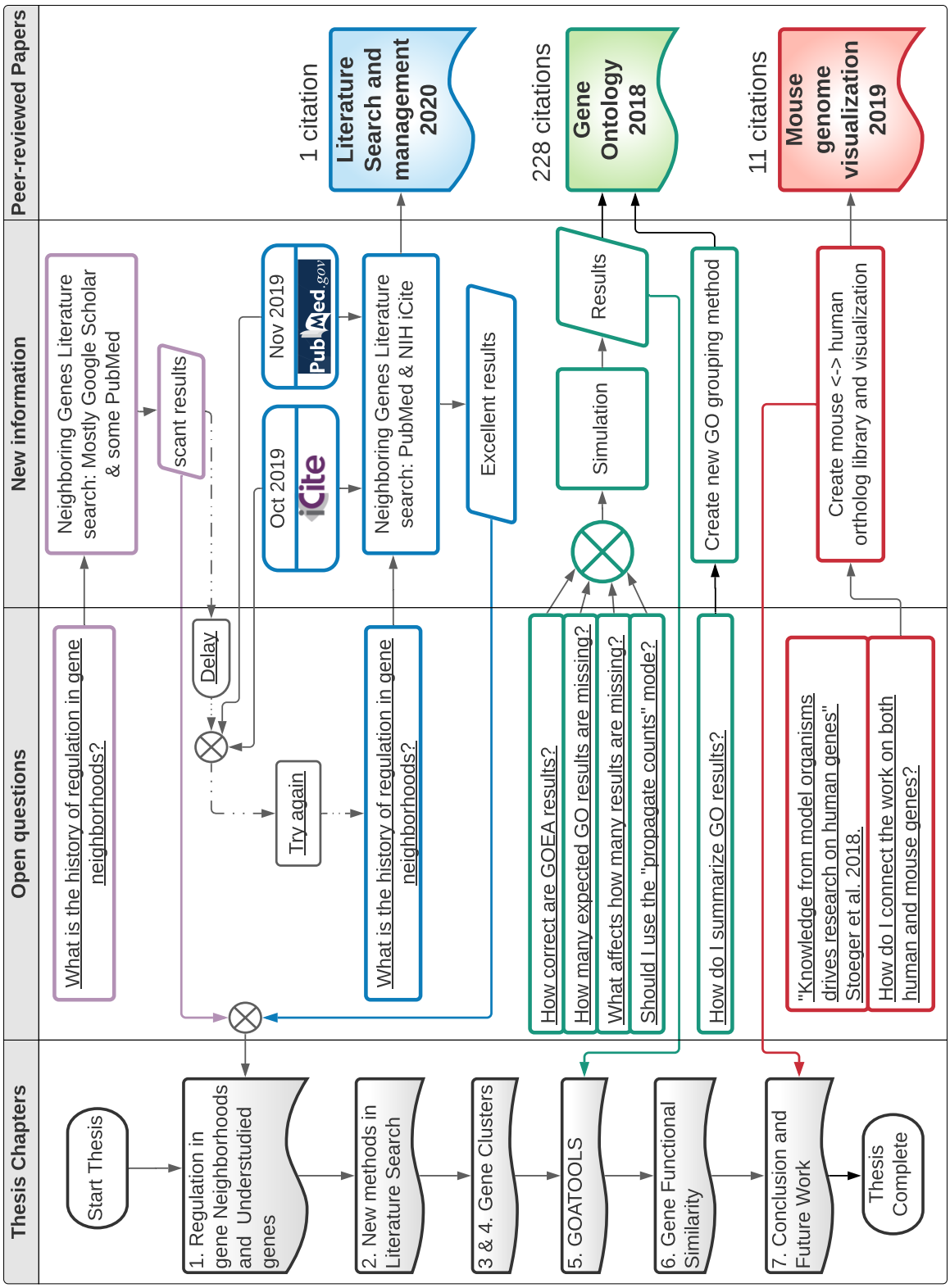

Figure A.1: **How the peer-reviewed papers fit into the thesis.** The PhD thesis chapters appear as gray shaded boxes in the left-most column. The peer-reviewed papers appear as blue, green, and red shaded boxes in the right-most column. The citation counts were obtained from Google Scholar in June 2021. The research question process flow is in the middle columns.