

1 Legend

• Question or comment by jury member. *Description of dissertation.*

• **Important matter** →

• Matter which will take some time to reply to / fix →

▷ Response or statement by candidate.

□ Unfinished task.

✓ Finished task.



2 Dries De Maeyer

2.1 Chapter 1

• The scope of the complete thesis is to define state transitions in cell types and the definition of cell type/state transition is not clear cut. I would frame the definition of a cell type more based on our current understanding and technology of biology. I.e. where do you draw the line of a new cell type?

□ Define cell type in introduction.

• Could you make the main outline of the thesis could even be more clear by providing a graphical overview of the different topics addressed in this thesis?

□ Add outline to thesis

2.2 Chapter 2

• Would you be able to extend this method to include Citeseq data? How would you see this?

▷ Yes. dyngen already generates protein abundance information, but more work is needed to compare characteristics of distributions of genes of citeseq or cytometry data to dyngen data, to ensure that the outputted information is realistic enough. As such, this feature is not included in the list of features in the current version of dyngen.

□ Add this to discussion in dyngen?

• Can you use the generated models for validation with existing cellular heterogeneity from scRNA-seq data?

□ Ask for explanation pertaining this question.

2.3 Chapter 3

• Since an ensemble of methods is proposed as the best approach to perform TI, how would this be addressed? Is there an objective means of selecting good performing methods?

▷ An internal measure for quantifying how well a given trajectory fits the original data would be extremely useful but also non-trivial. One internal measure could be: to what extent is the expression of some of the genes smooth along the trajectory. Variations of this approach should be tried and compared to gold standard metrics.

• Can TI methods be used for detecting and quantifying technical variation in between different samples? Be used as a metric for assessing batch integration? ▷ In some cases, yes. If two similar but disconnected trajectories are detected, it is possible to integrate those samples by performing trajectory alignment. Up until now, I have not encountered datasets where this approach would likely yield acceptable results, and instead would recommend performing batch correction separately from trajectory inference.

2.4 Chapter 7

- *Based on the method presented here, we could “learn” the topology of biological networks. What is to your insight the transposability of these topologies?*

□ Ask for explanation pertaining this question.

2.5 Chapter 8

- *The eight chapter is a manuscript in preparation about the assessment of TI methods. The main discussion topic is the lack of a quantitative measure for the scoring of new TI methods. This leads to the publication and development of new methods that do not show a clear performance improvement compared to other methods. The main topic touches upon some sore points that are inherent in bioinformatics. The main thought I have with this chapter is that defining clear goals in an emerging technology is difficult and therefore is sometimes ignored by fellow bioinformaticians. This because the lack of standardized performance metrics, missing public datasets that are well described and an ad hoc problem definition. I follow the idea of defining a clear cut approach to benchmarking (as described in chapter nine) and I like the proposition that is started in this manuscript for a discussion on good TI metrics. Would approaches like e.g. DREAM challenges be a means of tackling these issues to your opinion?*

▷ Yes and no. I believe any discussions into quantifying performance in any subdomain of bioinformatics through efforts such as benchmarking studies or challenges are crucial. However, ...chicken and egg problem; *in silico* simulators.

□ Finish explanation.

- *Do you expect that all methods which are mainly developed to answer specific biological questions would adhere to the same biological problem definition? Is the fact that people are trying to answer different question not a prior to define more generalized questions?*

▷ Indeed, methods designed to answer specific biological questions might have (slightly) differing problem definitions. This is one of the central issues we encountered in performing our benchmark, because the differences in problem definitions resulted in very different data structures in the inputs and outputs of these methods.

3 Joeri Ruysinck

▷ Small changes and typos have been fixed, unless noted otherwise.

3.1 Strengths of the thesis

- *The research domain of this thesis is young and technical evolutions in the encompassing field move fast. This makes it especially challenging to do research in the field, as often the original PhD thesis subject is obsolete at the end. The candidate has done exceptionally well to not only suggest and create novel algorithms which are at the state-of-the-art but also to grasp and tackle the greater high level challenges in the field, resulting in a potential seminal benchmarking paper.*

- *The thesis as a whole contains many relevant contributions, of which several have been published in high quality journals.*

- *The figures in the thesis are of exceptional quality. They are creative and able to convey large volumes of information to the reader in a visually pleasing format.*

- *The introduction chapter of the thesis demonstrates that the candidate understands and is able to convey the background of both (computational) biology and computer science.*

- *The passion of the candidate to bring modern good software development practices to the field is present throughout the work and might contribute to a more mature algorithm development in computational biology.*

3.2 Weaknesses of the thesis

- **The quality of writing, layout and graphical presentation is of a lower quality in the non-published parts of the thesis. Certain unpublished chapters feel rushed and unfinished.**



- Rewrite sections of dyno and of bred.

- Furthermore, for some chapters it feels that they could be merged into a broader discussion chapter.

- Reorder chapters in a more logical order. Rewrite abstracts to better introduce the context of each of the chapters.

- **The Introduction and Discussion chapter is missing a more broad context description. After reading the work, I gained very little insight how other researchers are using these tools to advance science.**



- Rewrite introduction in terms of the three main research questions. Discuss how answering these PhD questions can change the world. Better motivate the significance of certain technological advancements.

- Rewrite discussion in terms of the three topics: Trajectory inference, network inference and benchmarking. Discuss the practical implications of this work, and future perspectives in the field.

- *The main contribution of this work is the excellent benchmarking study published in Nature Biotechnology. However, this also effects the overall cohesiveness of the work, as information is repeated throughout the book at several levels of detail. This however a common side-effect of grouping published manuscripts into a thesis and I'm aware there is no (quick) solution to this.*

- ▷ Indeed, this is a result of bundling multiple manuscripts into one thesis. While there is a redundancy in information presented to the reader, these redundancies are largely confined to the introduction sections of each chapter.

- *It is very rare to see that research groups (or individual researchers) invest this much effort and time in creating qualitative software. I agree with the ideas of the candidates that these practices to adopt software to accompany research ideas should be encouraged. Yet I wonder how the current academic system can support such concepts. Academics often hold different positions in different research groups around the world, who should be the maintainer in these cases? The original PhD students that developed the work? The original research group?*

- ▷ I would argue that in order to develop software projects of this scale, adapting good software development practices (e.g. using code revision systems, unit testing, automated integration testing) 'pays for itself'. That is, making use of these practices requires some time investment at the start. However, developing and maintaining a large code base without these practices would require much greater effort, as it can lead to more frequent unexpected debugging expeditions as a result of long undetected bugs.

- *Will such higher standards not lead to barriers which might be harder to overcome for PhD students which might be constrained by limited support or funding?*

- ▷ Most of the good programming practices which are advocated for throughout this dissertation (Table 3.1, Figure 8.5, Figure 9.1) are skills that are not difficult to acquire given that the PhD student in question is able to program. A two-day crash course in good software development practices is likely sufficient to obtain the basics of developing qualitative software for a given programming language. If a PhD student is aiming to develop software on which other research will depend, they should familiarise themselves with the proper 'quality control' tools of software development.

3.3 Summary and chapter 1

- *As an interdisciplinary thesis, it is very much enjoyable that the introduction chapter contains a concise, yet precise overview of the concepts which are needed to understand the work of the candidate. I believe the chapter is nicely written and especially the flow of information is well chosen. However, the candidate only briefly touches upon the societal value of the technological innovations he describes. For a PhD thesis, I think it would be an added value if the candidate could briefly describe the practical implications of the technological advances (e.g. precision medicine, new treatments) and motivate the research from a non-technical view point (How is/can this PhD change the world?)*



- ▷ See section 3.2.

- Figure 1.9 does not contain SCORPIUS (probably published before, but still a bit weird for a PhD containing the manuscript)

▷ Indeed, this figure was adapted from Cannoodt et al. 2016 and serves as an illustration that TI methods often have common building blocks. Extending this figure to >75 TI methods would decrease the interpretability of the figure.

- Bold is only used in 1.1.2 and research challenges

▷ Usage of the bold font style in 1.1.2 and 1.3 serve clear, but different, purposes. In 1.1.2, three words are highlighted to establish a link between each term's introduction in the first paragraph, and their explanations in the second paragraph. In 1.3, important sections of each research objective are highlighted in order to highlight the structure of this section.

- Use of we/I in summary

▷ Usage of first person pronouns in this dissertation is very limited. In the summary of my contributions to the field, I believe usage of the first person improves legibility in comparison to when the phrases in question are converted to a passive form.

3.4 Chapter 2

- It is not clear to me what is meant with 'such as sampling a cell at a certain time point and once more at a later point' and why this is currently not possible?

- Figure 2.1.A. What does the figure in 'Combine simulations' represent? Are these single simulations which are mapped continuously to the backbone?

- Figure 2.5. It is not explained what Subfigure F represents. Only after the reader has completed the chapter, it becomes clear that the red lines probably represent snapshots and the lines, curves of gene expression.

- Figure 2.6: Consider changing the dashed lines, as they are not (clearly) visible printed

- Figure 2.6 and the example text for the cyclical example is unclear.

- While I understand that the instructions result in the desired state change, I also see many other configurations which would result in a cyclical example.

- Why is it currently needed to explicitly keep certain module expressions constant to determine the backbone? Would the same result not be achieved by simply simulating X times and averaging expression levels?

- Figure 2.7. The figure legend could be extended

- The introduction of FANTOM5 is abrupt and it would be better if the context would be briefly explained to the reader.

- The paragraph describing Step 3 is unclear: -> Target genes ARE? Regulated ... but is-> are ?

- Figure 2.8 can incorrectly suggest that target genes can only have a single regulator assigned

- Table 2.2. contains many default 'magic' numbers? It would be nice to know how in practice you ended up selecting them and which problems you encountered during design.

▷ The default values (or rather, default distributions) are determined based on our understanding of biology; for example transcription occurs at a much faster rate than splicing or translation, and this is reflected in the default distributions.

- Since these are default values, it would suggest that the user can modify them? Why would a user do that and what would be the impact?

- Am I correct that currently splicing offers little extra value as the formula's would imply that alternative splicing is not supported?

- **During a first read, it was confusing at times to read specifically about the strategy to determine the backbone while it had not been described how the simulation was performed.** At a certain time I was under the impression that dyngen did not simulate cell evolution end-to-end but only between states. Would the readability be improved if the order was changed in which the concepts



are explained? The backbone is only needed for the visualisation.

▷ The backbone is an essential part of defining the gene regulatory network that is used for the simulation, not just for visualisation purposes.

□ I will modify the text to clarify the purpose of the backbone better.

• 2.4.6. is a bit worrying. Could I reproduce dyngen without knowing these details? Does the sampling process not deserve more explanation?

□ I will better describe the different sampling procedures, with the help of one or more illustrations.

3.5 Chapter 3

• There are some artefacts of converting the paper into a PhD chapter: e.g. references to Supplementary tables/figures should be changed to the actual table number.

▷ Certain supplementary figures and files had not been included in the dissertation, but instead the reference contained a clickable URL which directs the reader to a downloadable file. This information is lost to readers of the printed dissertation.

✓ Explicit captions have been added in a section for supplementary figures and tables, and contain a printable URL which direct the reader towards the file.

• Some words are marked with an asterisk but the footnote? is missing.

▷ Section 3.4.2 uses an asterisk to denote special cases and does not refer to a missing footnote. This notation is introduced in the paragraph leading up to the usage thereof. An actual asterisk has been added to the introduction of the notation to improve readability of this section.

• You mention consensus predictions, is this something you investigated more in depth? Would 'ensemble' predictions result in better performance? Is there a good way to aggregate the predictions?

• Monocle DDRTree seems to overestimate the topology based on Figure 3.5, is this somehow related to the way you have to force the algorithm into the common format?

• The dyngen described here seems to be less advanced than the one from Chapter 2?

• Consistency: this chapter is US English



3.6 Chapter 4

• Figure 4.7: The figure is lacking what the colors represent.

• Figure 4.7: I see that even if you use the same dimensionality reduction; there is still a lot of variation between the methods on a very simple example. Could you clarify how often these methods disagree on trivial examples?

• Dyno offers a lot of guidance with respect to how to choose your methods, run them, etc. Did you consider support for features such as interpretation or automated error detection after prediction?

• **This chapter does not really read as a scientific paper (yet).** There are few comments I have since this is mainly an instruction manual + feature overview.

□ I will add a methods section to this chapter, as well as rewrite the results section to make it read more like a manuscript.

• Figure 4.7 is missing slingshot and paga label in part B

▷ Figure 4.7 does contain labels for Slingshot and PAGA. ✓ small gap between part A and part B has been added, to improve interpretability.



3.7 Chapter 5


• Comparing the preprint with the chapter in the thesis. Some minor changes seem to have been made to SCORPIUS: e.g. I believe in the pre-print classical Torgerson MDS is mentioned, while the more optimized Landmark MDS is used in the chapter and final version. There is also no mention of outlier removal. The main result Figure 5.4 is however identical. **Did this change or other potential**



changes have no (measurable) effect on the initial analysis?

▷ The SCORPIUS software was originally developed and the preprint published in 2016. Since then, the source code has been modified to be more scalable (LMDS + changes to principal curves), but these changes do not have a noticeable effect on the results of the method. The outlier removal was removed from the methodology because this step has since become a standard practice in single-cell omics preprocessing pipelines.

□ I will generate a comparison between the old and new version of SCORPIUS on the Schlitzer dataset, and quantify the differences in outcome.

• ***I also noticed that the biological conclusions or hypothesis have been reduced in claim strength in the paper or have been removed.*** It would be interesting to know if this is due to new insights that have been found in the period 2016-→ 2019 or mainly because the paper's target audience has shifted. 

▷ Section 5.2.2 regarding the functional modules found by SCORPIUS was reduced in character count, yet it makes the same claims. The main differences are that gene names and explanations of the biological functions are not discussed in detail, but are instead summarised in figure 5.4. The reasoning for these changes are indeed to adapt the text to the target audience. For example, detailed explanation of what actin polymerisation is, was removed; because immunologist readers already know this information and data scientist readers are presented with sufficient information when told that particular genes recovered by SCORPIUS are related to a biological process that is highly relevant to dendritic cell development, that is, actin polymerisation.

• This chapter would offer information on how the developed algorithms are being used to discover new insights, so I'm unsure if it is beneficial if this chapter evolves more into a technical description of Scorpius as the former is missing from the thesis.

• It would help to list the title of the manuscript on the pre-print server

✓ I fixed the DOI. This should be sufficient to find back the manuscript on the pre-print server.

3.8 Chapter 6

• The abstract seems to be missing a sentence.

• ***A main challenge in network inference has always been choosing a suitable cut-off to determine if something is interacting or not.*** There are examples of regions in the network which are accurate but come at a much later stage in the ranking, at which time many spurious edges have been added to other parts of the network. I can only assume that this problem becomes even more challenging if one wishes to create case-wise GRN.

• The added value of the last paragraph in 6.3. seems small

• Why is Scorpius used to generate Figure 6.3? Is this not confusing, since I assume we are not working with single cell data?

• I'm afraid to ask, but I don't get the reference to the name bred.

• ***I don't understand how the approach in figure 6.5 is different from GENIE3 and how it results in a tensor instead of matrix/ranking of 'shuffle' variable importance scores.*** I believe I'm missing important information on how SCENIC works to be able to understand the modifications. I really struggle with this chapter: is the method executed for each profile or for all profiles together? Do you not end up with a single GRN for each of the profiles and as such why would bred be a true case-wise NI method and the Scenic approach would not? It would be nice to discuss this more during the defense.

▷ bred indeed produces a single GRN for every individual sample in the dataset, and this by running more-or-less the same computations as a typical execution of GENIE3 with permutation importances. The distinction between bred and SCENIC is that SCENIC infers a single GRN for the whole dataset using GENIE3, and then uses a separate metric to predict to what extent an interaction was active in one of the samples.


□ I will expand improve figure 6.5 to better explain the workings of bred and the differences with GENIE3.

• The readability of Figure 6.1 could be improved by using a black font

3.9 Chapter 8

- Does this contribution not better fit in an overall discussion of the thesis?

3.10 Chapter 9

• **As the candidate is not first author on this paper, I'm not sure if this paper should be included in the PhD book verbatim.** Perhaps it should not be included or it could be merged with chapters 8-9-10 into a single discussion chapter. Maybe at the very least, the candidate should state a bit more explicitly why the content was included in the PhD book and what the personal contribution was. 

▷ This chapter was a collaboration between multiple labs. Each lab had independently had the idea of developing a set of guidelines on how to benchmark computational tools. Similar to other authors, I had written my own set of guidelines and reasoning behind the guidelines. The first author, Lukas M. Weber, pooled all of our inputs together and generated a first draft of the manuscript, which was subsequently reviewed and edited by all authors. I will not modify the contribution list as modifying it would make it contradict the author contributions list of the publication.

□ I will modify the abstract to better discuss and motivate how this chapter fits in the dissertation.

- Author contributions is not correct (CS)

▷ Unless I am mistaken, I believe they are correct.

3.11 Chapter 10

• The chapter contains a concise recap and critical view on the contributions in literature. It however does not discuss Chapter 9 (see comment above).

• **The discussion could be improved by stating how the tools are allowing new biological insights.** Only a small hint is provided in the last sentence of 10.4.

▷ See section 3.2.

4 Peter Dawyndt



• Not all chapters have been published and some are still in stage where some work needs to be done before they are publication-ready.

• **Not always a logical connection between the successive chapters**

□ I will better discuss the connection between the chapters in the introduction. I will reorder the chapters according to the new overview figure that was featured in the presentation.

• Drop the paper X prefix from most of the chapters

4.1 Chapter 2

• chapter does not seem completely finished; **may need a working example to demonstrate the steps a users needs to take to use the software, and to illustrate how the results look like;** for now, only the inner workings of the simulator are discussed  

□ I am actively working on adding more detailed use cases to the results section of this chapter.

• no information about implementation language (since Gillespie SSA uses a C++ library, this may hint the framework is also implemented in C++)

• case study may also serve as user documentation

• no references to user documentation, source code (open source?)

• no discussion about the performance of the framework

- since this is a PhD in computer science, I might also have expected a class diagram on the software, as a proof that the framework has been designed with extensibility in mind, as is claimed in the introduction of the chapter



- p24-25 how are networks specified by user (ergonomy & standards); support for SBML?
- p26: has the BBL-ontology been invented for this application or an existing ontology?
- maybe this chapter should come after the next chapter

4.2 Chapter 3

- what reactions have the authors received from the developers of the tools that were reviewed in this benchmark, or from users of the TI tools that have used the benchmark results to decide what tools to use
- could this study also yield some standard ways for data formats that tools could use to input/output data (already hinted in the text); what approach would the authors choose towards a standards proposal?

4.3 Chapter 4

- Some users filed an issue on GitHub reporting problems with running Docker on Windows; a promise was made by the authors that this issue would be looked at, but no solution has been provided so far; any progress in resolving this issue?
- as with CH2, this chapter does not seem completely finished; especially the results section (4.2) read more as an incoherent listing of features of dyno, where an overall flow of this section is missing for the readership; adding some details about the case study, e.g. how to run the analysis on the command line, might benefit the readership

4.4 Chapter 5

- **fundamental question: how objective is a benchmark of a tool, if developers have designed both the tool and the benchmark?**



▷ The software and original preprint of tool pre-dates the benchmark. Since then, two modifications have been made to improve the scalability without changing the methodology, and one step in the methodology was removed because it was decreasing the stability of the method without improving the accuracy significantly. As such, the benchmark did not have a (strong) impact on the methodology of the tool.

▷ Throughout the development of the benchmark, we were acutely aware that we could be unintentionally biasing our benchmark to favour SCORPIUS. We paid attention to the following matters when designing our benchmark, hoping to have mitigated the largest sources of bias in this way.

- Our main hypotheses and methodology were determined before implementation of any code or performing any of the experiments.
- When including synthetic datasets we also included synthetic datasets generated by simulators other than dynogen.
- We performed extensive testing of our metrics to verify that each metric quantifies what we aim for it to quantify.
- When wrapping SCORPIUS vs. TI methods from other authors, we could be biased in our interaction with the tool simply because we better understand our own software. To mitigate this, we contacted every author of every TI method for comments on our usage of their software and also on the quality control scores that we assigned to their software based on our question checklist. Several authors made contributions to the wrappers, and also improved their own software based on certain QC criterion that we proposed.

4.5 Chapter 6

- *paper needs finalisation to make it publication-ready*

4.6 Chapter 7

- *performance measurements show that method is faster than state of the art, but no formal time complexity is derived and no formal proof is given that the method still computes the correct number of graphlets; strange behaviour in time measurements is not diagnosed in depth*

4.7 Chapter 8

- *I would argue against having this as a separate chapter, but would use its content to supplement the introduction and CH3*

4.8 Discussion

- *points of discussion, open source: upon publication or release-early-release-often?*
- ***reads more like a summary*** + *listing some use case that made use of contributions made in the manuscript, rather than a real discussion*
▷ [See section 3.2.](#)
- *10.7 might better go into a preface of the manuscript*

5 Pierre Geurts

5.1 Remarks alongside summaries


- *Chapter 1 does not go into the technical details of the computational tools, but it gives a gentle and nicely written introduction to the field.*
- *The results section of Chapter 2 is however limited. I would have appreciated to see more experiments to illustrate the quality of the generated data and also to have a comparison with other simulators from the literature. But to be fair, it is not clear how data simulators could be compared to each other or even how the quality of a given simulator could be quantitatively evaluated.*
- *Chapter 4 is a description of the features offered by the toolkit, more than a technical presentation or an evaluation of the tools, but more details about every step can be found in the other papers.*
- *Regarding Chapter 6: Further experiments would be needed to really evaluate the performance of the method, notably on artificial datasets where the true case-wise networks would be known, but also on real single-cell omics dataset, which seemed to be the original objective of the authors.*


5.2 Questions / remarks


- *The focus of the comparison in Paper 2 is on ranking the different methods. This is extremely useful obviously. **I think it would have been interesting to take this opportunity to also assess the methods on a more absolute scale.** I don't think the paper provides answer to such questions for example: Is a given method good enough, in general or for a given type of trajectory? Is the problem of inferring a given type of trajectory solved in a satisfactory way? On which kind of trajectories should the research community focus? What are the cases that no method can tackle well?*
- *Similarly, in Paper 4, the SCORPIUS method is only assessed relatively to other methods. There is*






no doubt that SCORPIUS is a very good method but I'm missing an absolute assessment on how well it performs. It would have been interesting to report the correlation between the predicted and true pseudo-time values, in addition to the more general score of Paper 2 and to show datasets where it performs poorly, if any.


- **In Paper 2, I was surprised that no statistical test was performed to assess whether methods are statistically different from one another** (or better than random), as this is commonly done in machine learning for example. In addition to method ranking, it would have been interesting to be able to highlight groups of methods that are not statistically different from each other in terms of accuracy. 

- In several papers, the authors use Random Forests variable importances to find differentially expressed genes (along a trajectory, between branches, at some branching point, etc). **I'm wondering why they use RF instead of more standard statistical tests for finding these genes.** RF are useful when some multivariate effects need to be found but I'm not sure it is the case in all situations where RF is used in the thesis. In Paper 2 for example, it is said that mtry is set to 1% to make sure that highly correlated features are not suppressed in the ranking. In Paper 4, Figure 5.5, there is a mention that "shallow" random forests are used, which suggest that they don't want to find complex interaction effects. What is the expected benefit of using RF instead of statistical tests? 

- In Paper 4, **I'm missing a discussion of the computational complexity of the SCORPIUS method.** In Section 5.2, the authors say that MDS and principal curves have been adapted so that they scale linearly but given that the approximated principal curves algorithm is initialized by k-means, isn't the algorithm quadratic? This is confirmed by Figure 3.7 in Paper 2 that shows that SCORPIUS scales quadratically with the number of cells. 

- In Paper 5, it is argued that most case-wise network inference methods consist in post-processing a static GRN, unlike the bred method which is a true case-wise NI method. My feeling is that bred is not that different in spirit from existing post-processing methods, as it starts from a model trained on the full dataset of samples, which is trying to find regulators that are relevant for most of the samples. **What makes bred a true case-wise NI method?** More fundamentally, **I'm missing a more formal definition of what a case-wise regulatory network is.** A network can obviously not be derived from a single sample (as interactions between genes cannot be retrieved from a single static expression vector). 

- As mentioned in the conclusion of Paper 5, the bred method needs to be assessed on more problems and against ground truth networks. One baseline experiment that I would have been interested in at this stage is **a comparison of the clustering in Figure 6.1 obtained from the regulomes with a clustering obtained from the original expression profiles.** Does using the regulomes really make a difference in terms of the final clustering?  

- In Paper 7, the improvement obtained with IncGraph on the two gene networks is interesting. However, **I would have liked to see a comparison with a very simple heuristic that goes through the ranked list of edges and only adds a new edge if it is not redundant with a previously added edge** (as in minimum spanning tree construction algorithms). 

5.3 Suggested modifications

▷ Small suggestions have been accepted, unless noted otherwise.

- During my first read of the thesis, I had difficulty understanding what the author meant by "milestone" and by "region of delayed commitment". Maybe introduce a definition of these notions somewhere in Chapter 1, if it fits.

- I don't understand footnote 2 on page 108 explaining why the method in Chapter 4 is called 'bred'. Maybe clarify this (sorry if I miss something obvious).

- The title of Chapter 6 is misleading as the method is actually not applied to single-cell data. I would suggest to maybe change it into "inferring case-wise regulatory networks".

- In Paper 5, Section 6.4.1: how is the minimum threshold on average importance set? This is important as this threshold conditions the number of interactions found.

- Page 5, the sentence "IHC realizes the visualization...to specific antigens" seems to be redundant

with the next one.

6 Stein Aerts

- The introduction is relatively brief, but well written.
- The methods of Chapter 2 are extensive, but the results are very brief. I'm wondering whether some parts that are now in the Methods section would not fit better in the Results section. Testing how representative the generated data are is of course a huge challenge, this could have been addressed perhaps in the Discussion.
 - I am working on providing use-cases in the results section of this chapter.
- **It would also be interesting to evaluate the impact of the gene regulatory models in the simulation. How crucial is it to use such an extensive model of mRNA production and degradation, and how do these parameters affect the simulated single-cell data?** ⚠
- Chapter 4 is a bit opposite of chapter 2 - here the results are described, but without methods. For example, how are features (genes) selected on the branch points, in a branch, etc (which methods are used for this).
 - I will elaborate on the methodology of this chapter.
- The marker gene identification is very interesting, but is not described in detail in the Methods, how is this RF built, and trained? Is it RF regression or classification? What is the topology, why use a RF, has it been compared to other techniques?
- In the benchmark of SCORPIUS compared to other methods, it is mentioned that many data sets are used, but there are no details on which data sets.
- **In Chapter 6, it would have been interesting to compare the results of bred with such methods (for example GENIE3 + regulon specificity score).** ⚠
 - I will attempt to finish a quantitative comparison between bred and GENIE3 + regulon specificity score. If not in the PhD dissertation, this will definitely be included in the publication of this work.
- **The methods could be a little bit more clear on the specific differences between bred and GENIE3.** ⚠
 - I elaborate on the difference between bred and GENIE3 better.
- In Chapter 7, it may have been interesting to discuss the relationship with the classical network motifs from Uri Alon.
- Chapter 9 could have contained a bit more in-depth discussion about trajectory inference in general, biology, and discussion of the results in the context of other work.

7 Vanessa Vermeirssen


7.1 Chapter 1

- The introduction is concise and to the point. This is also the case for the introduction sections of the different paper chapters. Therefore, some sections could have been more elaborated on e.g. briefly describe and include the differences between the different dimensionality reduction methods and define different network inference approaches. At the end of the introduction, the objectives are clearly put into the research context.

7.2 Chapter 2

- In the introduction the PhD candidate mentions **other scRNA-seq profile generators i.e. the state of the art, but in this chapter no comparison is made with those in order to analyze how well dyngen performs on the generation of scRNA-seq profiles.** In the introduction I also lack a short description ⚠

of the workflows/underlying algorithms of those scRNA-seq profile generators as opposed to that of dyngen. Can you also comment on how dyngen differs from simulators of gene regulatory networks such as GeneNetWeaver and Neural Gene Network Constructor?

- Dyngen is put forward as a multi-modality simulator as opposed to other scRNA-seq profile generators. **This multi-modality is mostly reflected in the results as different simulations of experiments and extracted data, all based on transcriptomics, whereas in the introduction this multi-modality is defined as the profiling of another data type than scRNA-seq.** Please clarify this in this chapter. 

- p. 26 The backbone is defined as a network of cellular states. The connection between gene modules and those cellular states could be made more explicitly clear. Do I understand correctly that in a specific cellular state specific gene modules of the module network become/are active?

- A legend is missing for Fig. 2.7

- p. 27 Step 1: "from the module IT belongs to".

▷ This has been fixed.

- Write more clearly in step 2: "These interactions..." Specify.

- And step 3: "Target genes..." Verb missing in sentence.

▷ ???

7.3 Chapter 3

- p.37 Last part missing in sentence? "The overall score between the different dataset sources was moderately to highly correlated ..."

▷ ???

- dyngen is somewhat differently explained here in the Methods section. Please integrate with paper 1 chapter and remove redundancy.

7.4 Chapter 4


- It is recommended to clarify and extend the discussion on future improvements that could be added to the tool, e.g. alignment and differential expression. What is the biological significance? What is meant by feature importance tools/score?

7.5 Chapter 6

- Is this really the first algorithm for inferring case-wise regulatory networks?

▷ No, as mentioned in the introduction: "Thus, while several case-wise NI methods have already been proposed, their implementation consisted of post-processing a static GRN to arrive at a case-wise GRN." I thus argue that bred is the first *true* method for inferring case-wise regulatory networks.

- p. 107 Delete in Results: We generate.

- p. 108 "While accuracy of the network of predicted regulatory interactions is lower in comparison to experimental validation techniques, network inference methods offer an unbiased and high-throughput insight into the regulatory dynamics of a biological system." I do not agree. Network inference methods do have a bias, depending on the underlying algorithm (cfr Marbach et al. 2012, ...), but so do specific experimental techniques that measure only an aspect of gene regulation. It is true that precision-recall for network inference predictions is not high, so they can only be viewed as hypothesis generators, but are nevertheless valid since they offer, relatively easy, a genome-wide view on gene regulation. 

▷ I meant to use the term 'data-driven', rather than 'unbiased'.

□ Change 'unbiased' into 'data-driven'.


- There are significant differences between the methods of Kuijer and Liu, please describe these methods in more detail in the introduction.

▷ I respectfully disagree. In essence, the core concept of how a case-specific GRN is derived is exactly the same (See Kuijjer Fig 1B and Liu Fig 1C), namely by performing classical (static) NI twice, once all of the samples and once more after removing one of the samples. The differences between the two NI executions results in the desired case-specific network.

□ Regardless, I will revisit the description of these methods and expand the explanation, if necessary.

- What about case-specific random forests like in the R package ranger, `csfr()` ?


▷ case-specific random forest is a similar sounding but entirely different concept. From the documentation of the `csfr()` function: In case-specific random forests (CSRF), random forests are built specific to the cases of interest. Instead of using equal probabilities, the cases are weighted according to their difference to the case of interest.

- The sentence “As such, these methods will most likely recover the interactions that are prevalent in the whole population, and will miss interactions that are specific to only a subpopulation” is incorrect. 
I would suggest you read the relevant method papers again, see above.

▷ I respectfully disagree. I believe detecting interactions which are only prevalent in a portion of the population will make it harder for classical (static) NI methods to detect them. Similar to how it is possible to detect differential expression in a sub-population when using single-cell transcriptomics vs. bulk transcriptomics, sample-specific NI will make it easier to detect context-dependent interactions.

□ An experiment would need to be performed to provide an answer to this issue.

- p. 114, 6.4.3. Can you include a motivation for the specific clustering algorithms used here?

- **The discussion is very short, please elaborate on:** possible applications in the field of personalized precision medicine and comparison to state of the art. 

□ I elaborate on the implications and application of the method and refer to the current state of the art.

7.6 Chapter 7


- Please define graphlet and orbit more precisely in the introduction and differentiate between the concepts network motif, subgraph and graphlet.

- Can you also briefly introduce the non-incremental approach Orca?

7.7 Chapter 8

- A large part of this chapter is already incorporated in previous chapters, especially in paper2. I would suggest to integrate both and avoid redundancy. A large part of this chapter would fit in the discussion of paper2.

7.8 Chapter 10

- **The general discussion on the thesis is only 3 pages and consists of small, but critical, discussion paragraphs on each of the different papers. The general discussion is too brief and lacks future perspectives on the field. It provides some links between the different chapters, but this bridging could be increased further. I miss an overarching discussion on where the field of modelling single cell dynamics is going.** 

▷ See section 3.2.