Chapter 8

Discussion

8.1 Overall summary

In this thesis I have introduced computational methods to unravel the life history of a tumour from massively parallel sequencing data and applied them to the 2,778 cancer genomes in the ICGC PCAWG project. To aid with the subclonal reconstruction of tumours I introduced algorithms to estimate subclonal copy number, multiplicity and cancer cell fraction values for SNVs and indels and to infer the subclonal architecture.

The methods have been extensively validated using different simulated data sets and cross-compared to those from other labs. The validation shows that, even though DPClust is one of the best performing methods on the simulated data within the PCAWG consortium and that performance is consistent with the results on real data, it is not perfect. The observation that methods are not perfect lead to the development of consensus procedures for both copy number and subclonal architectures. I showed that the output of the consensus is more consistent across methods, and for the subclonal architectures that corresponded well to the patterns observed on simulated data.

The copy number consensus procedure first combines breakpoints from six CNA calling methods with SV calls into a single, complete consensus. By synchronising the segmentations across the six callers we achieve six profiles that are complete and directly comparable. There is substantial complete agreement between the callers, but agreement on the near full genome is only achieved through a strict majority vote. This highlights that there is still substantial disagreement between some of the methods, which has already resulted in method improvements. Consensus subclonal architectures were obtained across 11 individual subclonal inference methods. Three consensus approaches were developed based on different representations of the input data and validation showed that they are consistently comparable

to the best individual caller, while assigning all SNVs, indels and SVs for which allele frequencies were available.

When these methods are applied to the PCAWG data set we find that nearly all tumours contain at least one subclone, that subclones contain driver mutations in known cancer genes and those drivers are under positive selection. These findings show that tumours are still evolving at the point of diagnosis. Furthermore, the PCAWG data set covers 36 histologically distinct cancer types and donors cover a wide range of ages, while recruitment has occurred in Europe, North America and Asia, resulting in a diverse population that covers many germline genetic backgrounds. The fact that we observe subclones in nearly all tumours across this range of variables may suggest that tumours are in a continuous process of clonal expansion.

Recent work suggests that tumours could evolve neutrally through genetic drift. We applied the concepts from Williams et al. (2016) and found signs of positive selection in subclonal mutations from tumours identified as evolving neutrally and these tumours can contain driver mutations for which we are highly confident that they are subclonal. However, dN/dS ratios were used to estimate the amount of positive selection, and dN/dS analysis pools mutations across samples. It is therefore possible that a subset of tumours is indeed no longer under the influence of positive selection, or that the VAF tail that is used to detect neutrality contains subclones that are under positive selection and those that don't. But that subset is sufficiently small to not affect the dN/dS ratios obtained on the pooled mutations.

Analysis of clinically actionable driver mutations reveals that the detected clones can be informative in the clinic. 11% of tumours contain at least one subclonal actionable event, which means a prescribed treatment is inherently flawed as it would not target all tumour cells. However, over half these tumours also contain a clonal actionable event. That may provide a route to apply more effective treatment as clones could be targeted separately.

Evaluation of mutational signatures shows that activity of life-style associated signatures decreases during tumour evolution, although this signal could also be explained by the increase of a combination of other signatures. Meanwhile, APOBEC activity typically increases. Finally, cancer timelines were created by combining the evolutionary histories of tumours within a cancer type. The analysis confirms classic knowledge and suggests that cancer types follow distinct patterns of tumour evolution.

8.2 Future directions

8.2.1 A more in-depth view of intra-tumour heterogeneity

The single sample whole genome sequencing data presented in this thesis ultimately provides only a high level view of intra-tumour heterogeneity. Work by others based on multiple samples from the same cancer (Jamal-Hanjani et al., 2017) shows that there is more high level heterogeneity, and our analysis about additional heterogeneity corroborates that. It's possible that the view obtained by these bulk sequencing approaches is just the tip of the iceberg and that tumours consist of many 100s to 1000s of clones. Bigger and deeper sequencing studies will reveal the true extend of intra-tumour heterogeneity.

Larger sequencing studies, that cover more cases of the same cancer type, such as the Pan Prostate Genomics Consortium, are needed to extend our knowledge of late drivers. The work in this thesis suggests that tumours become more diverse as they evolve, with a larger set of genes acting as late drivers. A complete overview of late drivers may lead to new treatment options and it may shed light on drivers that are rarely early.

A currently mostly unexplored angle in these kinds of heterogeneity studies is expression data, which could be overlayed onto the subclonal architectures. Subclonal inference is focussed on establishing the genotype of evolution. What effect these additional mutations contained within the cells in the subclone have on the expression profile remains unclear. Using matched RNAseq data it should be possible to observe expressed transcripts with a subclonal mutation. This kind of analysis will however not provide a complete overview of the expression profile of a subclone. Single cell sequencing technology, especially sequencing of the genome and transcriptome from the same cell, does give access to that information (Macaulay et al., 2015).

Single cell technology also provides access to the lower levels of heterogeneity that are not visible with current bulk sequencing approaches. One could paint the 3D landscape of tumour heterogeneity by carefully sampling cells from the tumour environment (Mamlouk et al., 2017). However, single cell genome and transcriptome technology is still too expensive and cannot easily be scaled up to the numbers required to paint a comprehensive picture, while the genomic data is hampered by factors such as allele dropout and sequencing errors (Van Loo and Voet, 2014).

Higher sample counts per tumour can not only lengthen the branches of the evolutionary trees, it may also provide a more fine grained picture of the tree trunk. SNVs that appear as clonal in one tumour region, may in fact appear subclonal in another. A more fine grained trunk will provide a better view of the very early events that have given rise to the tumour

and provides a clearer sequence of events. A clearer picture can then lead to cancer type evolutionary histories with smaller confidence intervals.

8.2.2 Tumour evolution

Recent work on normal tissue from healthy individuals reported many small clones at an in-depth view of epithelial tissue and showed complex clonal dynamics, containing clones with driver mutations in well known cancer genes (Martincorena et al., 2016). Given these observations, and assuming they can be extrapolated to other tissues, it is surprising that the cancer evolutionary timelines reported in this thesis suggest that cancers develop over decades. The process of clonal expansion may appear similarly in healthy tissues and in tumours. More detailed experiments are required to understand the dynamics of the environment in which the tumour grows and why the malignant growth can escape while many early clones cannot. Ultimately, we cannot easily observe the micro-environment *in vivo* when a driver mutation has its selective advantage. Normal tissue observational experiments will be required to better understand the dynamics of clones that provides the breeding ground for malignant lesions.

More detailed experimental data will also facilitate those who work on mathematical models of tumour evolution. So far this field has been held back by the lack of a bridge between models of early evolution and the tumours observed in the clinic that represent a much later evolutionary phase. Projects are needed where model development is provided with measurements of input variables, and longitudinal follow-up of clonal dynamics of the populations that provided the input variables could provide the ground truth for intermediate predictions made by developed models (for example, clone sizes and distributions at various time-points). Scenarios should include the introduction of driver mutations, through CRISPR for example, or population bottlenecks as created by the application of drugs. A combination of these two should lead to closing of the gap between those that study tumour evolution top-down (as is reported in this thesis) and those that study it bottom-up.

8.2.3 Towards clinical application

The findings reported in this thesis reveal that tumour evolution continues up to (shortly before) diagnosis. This has profound clinical implications as it suggests every tumour can in principle become resistant to treatment (Holohan et al., 2013), which can be unlocked via a single mutation (Nazarian et al., 2010; Zaretsky et al., 2016). In cancer types with a high mutation burden it is therefore not implausible that every conceivable somatic mutation is available. The resistance mechanism may therefore already be available and will be

selected for via treatment application. It is therefore important to assess how ubiquitous the resistance mechanism has to be before treatment becomes inevitably unsuccessful. If a large subpopulation of cells with the resistance mechanism is required (as opposed to a single cell with a single mutation), then a routine screening could be facilitated to rule out certain types of treatment. Large scale collection and deep profiling of pre- and post-resistance cancers is required, along with a more complete list of resistance mechanism markers, is required to see whether heterogeneity profiling can be a useful clinical application.

The findings also suggest that broad timelines exist across cancers of a certain type. It has recently been suggested ctDNA provides a complete picture of the major clonal populations within a tumour, and a blood test could therefore in principle be used to not only detect a tumour, but to also assess and follow its progression (Abbosh et al., 2017). A non-invasive blood test could help detect cancers more easily and potentially earlier. If one can classify a tumour as to be on a evolutionary 'pathway', then it may be possible to predict how it will develop further and choose treatments that slow tumour progression by closing possible evolutionary routes. However, even though evidence is beginning to emerge that overall timelines may exist within cancer types (Fearon and Vogelstein, 1990; Gerstung et al., 2017; Makohon-Moore et al., 2018), it is currently not clear whether individual tumours truly follow such a pathway, or whether the observed order of driver mutations is one out-of-many ways to malignancy. A more detailed picture is required of evolutionary paths that tumours take, but it is equally important to assess which other evolutionary paths arise during a tumours' life time and are outcompeted.

The emergence of ctDNA based tumour tracking potentially allows for much earlier diagnosis via routine testing. However, it will increasingly emphasise that better understanding of the difference between normal and malignant tissue evolution is required. Driver mutations in *TP53* in morphologically normal epithelial have been described and are suggesting that just the acquisition of these driver mutations is not enough for a cancer to arise (Martincorena et al., 2018). A more comprehensive overview of normal somatic evolution is required to demarcate the crossover point to malignancy more clearly, and to show that somatic mutations alone (as measured via ctDNA) can clearly differentiate between the two states.

8.2.4 Methods

Within the PCAWG consortium there has been a great emphasis on consensus strategies that build confident calls by combining evidence from multiple methods. These ensemble methods perform well, however they come at considerable computational cost. Our copy number consensus procedure required six methods to run across the full data set to obtain a complete set of breakpoints, followed by another full run to obtain calls on the consensus

segmentation. It is interesting that the review sessions, where copy number was discussed in great detail, has already yielded improvements to methods and has sparked additional development.

There is not only a need for more accurate methods, there is also a need for public benchmarking data sets where performance can be validated and to aid the development process. The extensive comparison of subclonal architecture callers that was performed internally and lead to the development of three consensus procedures also lead to improvements to individual methods. Efforts like the DREAM somatic variant calling - heterogeneity project are an important part of this, as are simulators like BAMsurgeon (Ewing et al., 2015) and SimClone.

More real data with multiple samples of the same tumour are also required to aid further benchmarking. As is discussed in this thesis, multi-region approaches can be more powerful to detect intra-tumour heterogeneity, as it allows for a larger area of the tumour to be sampled and mutation clusters with a similar CCF in one sample can be more easily separated if their CCF differs in another sample. The richer view that is obtained via multi-region sequencing can be used for validation purposes by applying subclonal architecture callers on just a single sample. So far however, comparatively little multi-region whole genome sequencing data has been published. The PCAWG project did include cases where multiple samples (primary-primary, primary-metastasis, primary-relapse) were available. However, these cases span few cancer types, each cancer type is represented by low numbers of cases and the cases are further split by multi-focality and whether they were obtained before or after start of treatment. A large cohort, with multi-region whole genome sequencing is required to help further validate performance of subclonal architecture callers.

My analysis of performance of DPClust on the SimClone1000 data set highlights limits of what can be detected. Subclones that are within 0.25 CCF of each other stand a good chance of being merged. Other methods may have a lower threshold for disentangling subclones, but there is always a limit. The subclone distance limit however is much less likely to affect cases with multi-region sequencing where a pair of mutation clusters may appear in similar proportions in one sample, but different in another. It is important to understand these limits when interpreting the output of subclonal architecture callers. Simulations will help to discover the limits and could ultimately be used to understand their impact on interpretation.

There is considerable scope to develop subclonal inference methods that use SNVs, indels, CNAs and SVs. There should be more signal to detect subclones by considering evidence across all types of mutations. In this thesis I have presented an approach that includes CNAs as pseudo-SNVs, based on CCF values estimated by Battenberg. This approach however does not take into account the characteristics of the underlying data upon which the copy

number estimates are based. A combined approach should take those into account, should have a suitable error model for the CNA CCFs and should include a step that detects whether the assumptions required to estimate CCF values for CNAs are violated.