Chapter 6

Conclusions and future perspectives

6.1 Summary of the main findings

The research presented in this dissertation leveraged naturally occurring somatic mutations as passive marks of cell division and hence as a direct bar coding of the developmental history of a cell. By reconstructing phylogenies from those mutations, I was able to investigate fundamental processes of human development as well as the emergence of cancer via a precursor lesion. Briefly, the main results per chapter are:

1. **Extensive phylogenies of normal human development reveal asymmetries in embryonic lineages, spatial patterning of mosaicism and clonal expansions in adult life.** While the asymmetry of contribution of the first cell division is approximately 2:1 and is conserved in most tissues, the discrepancy in asymmetry between different samples from the same patients points at a role for genetic bottlenecks after blastulation. This is strongly exemplified by a bulk brain sample from one patient, which is almost exclusively derived from one of two zygotic lineages, unlike the bulk colon sample from the same patient. From targeted re-sequencing of embryonic variants, I was able to show a large-scale spatial pattern in human brain corresponding to heterogeneous mosaicism. The phylogenies and spatial data revealed the typical embryonic patch size for human colonic crypts, as well as later clonal expansions in normal tissues such as the prostate, colon and appendix, in some cases due to cancer driver mutations. In addition, the pattern of sharing between seminiferous tubule sections and samples derived from other tissues indicates an extraembryonic origin for human germ cells, most likely the amnion.

2. **Wilms tumours often arise from an embryonal precursor lesion, residing in the normal kidney.** By interrogating tumour variants in normal renal tissues, I identified

a detectable, pre-neoplastic clone in over half of the Wilms tumour. This precursor appears to be driven by hypermethylation of *H19*, effectively resulting in a mosaic Beckwith-Wiedemann syndrome. However, kidneys containing cells with this hypermethylation appear functionally and morphologically normal. From timing the occurrence in bilateral tumours, it is likely that the initial hypermethylation event happens early in development.

3. **The human placenta consists of large clonal patches, which can segregate from the inner cell mass-derived lineages at the earliest stages.** Large bulk biopsies of the placenta contain clonal populations, which can be traced back to its trophectodermal component. Trophoblast clusters excised from the same biopsy are closely related to one another. Comparing early embryonic mutations between placental lineages and umbilical cord DNA, which is derived from the inner cell mass, revealed that in approximately half of the cases a placental lineage has no post-zygotic relation with the umbilical cord. Furthermore, in a quarter of cases, the umbilical cord is entirely derived from a progenitor later than the zygote, facilitating a natural segregation and pathway to generate confined placental mosaicism.

6.2 Future perspectives and ongoing work

Taken together, these three lines of research illustrate the potential of somatic mutations to answer fundamental questions about human developmental processes and the origins of cancer. The findings from the individual chapters have profound implications and lead to interesting future work.

Lineage tracing of normal development has highlighted the existence of mosaic patterns of embryonic mutations on a tissue- and organ-level, which reflect lineage commitments and segregations during embryogenesis. Future studies can use such patterns to reconstruct the growth of individual organs, such as exemplified by the placental study in Chapter 5, but also to investigate expansions and tissue renewal in case of injury. It is plausible that, as the cost of whole-genome sequencing decreases even further, that future lineage tracing studies can be based on even more samples per patient and reveal patterns of human embryogenesis currently obscured to us, such as the precisely localising timing the emergence of the primordial germ cells and the providing evidence for the existence of extraembryonic intercalation in humans. Furthermore, to complement the phylogeny of the embryo proper, genomes derived from extraembryonic tissues, such as the placenta or umbilical cord might further elucidate the lineage commitments between the trophectoderm and inner cell mass, and the epiblast and hypoblast, respectively.

The emergence of Wilms tumour from large precursor clones in the normal kidney has profound implications on the screening, treatment and further surveillance of Wilms tumour patients, especially if aberrant *H19* methylation status is detectable non-invasively. In particular, the discovery of large fields of normal renal cells carrying *H19* hypermethylation leads to important questions: (1) do differentiated renal cells still have the potential to transform into Wilms tumours, (2) how are these cells able to overcome the loss of imprinting, (3) what role do nephrogenic rests have in the origin of Wilms tumour and do they differentiate or disappear naturally over time and (4) how widespread is a mosaic Beckwith-Wiedemann syndrome in the human population.

The results from the placental sequencing project emphasise the common presence of aneuploidy in the human placenta, best exemplified by a case of trisomic rescue of chromosome 10. It is plausible that a significant proportion of the human population has been subjected to trisomic rescue without knowing, given the high prevalence of confined placental mosaicism and the observation that trisomic rescues could explain a significant part of this phenomenon. This naturally leads to the question of how often this affects the general population, and whether we can detect any chromosome-level biases in frequency of aneuploidy. If the trisomic rescue event would lead to a completely normal chromosomal landscape, where the disomy consists of one maternal and one paternal copy, the copy number profile appears completely normal and the reversal of aneuploidy is undetectable. However, if the the trisomic rescue leads to two copies from the same parent being retained, the meiotic recombination leads to regions of haploidisation in the genome, which are easily detectable through conventional whole-genome sequencing. The latter should happen in approximately a third of all cases, if all three chromosomes have an equal probability of being lost.

Beyond the immediate ramifications and direct follow-ups to the research presented here, the methodology of lineage tracing using somatic mutations can be applied to answer a wide variety of other biological questions. The section below will cover a few of the exciting avenues of research following this principle, some of which are ongoing work.

6.2.1 Other childhood cancers, bilateral tumours and secondary malignancies

The initial discovery of an early clonal expansion predisposing to Wilms tumours prompts questions about the generality of this phenomenon. That is, do other childhood cancers arise similarly? Can early *H19* hypermethylation lead to other tumours? One piece of ongoing work focuses on malignant rhabdoid tumour (MRT) and the associated normal tissue, i.e. nerves of the kidney hilum for renal MRT or spinal nerve roots in the case of spinal MRT

(Custers et al., 2021). Preliminary analysis has indicated that normal nervous tissue carries large precursors clones to the MRT, with a mutation burden and clonal composition even more pronounced than the clonal nephrogenesis observed in Wilms tumour. In addition, both normal cells and tumour have inactivated both copies of *SMARCB1*, a tumour suppressor unequivocally implicated in MRT pathogenesis (Margol and Judkins, 2014). No discernible genetic events with a phenotypic consequence distinguish the normal clone from the tumour, suggesting that a non-genetic event might trigger the final transformation from normal cell to tumour.

Beckwith-Wiedemann syndrome predisposes to a number of different embryonal childhood cancers besides Wilms tumour, leading to the possibility that the *modus operandi* of tumour formation is equivalent to clonal nephrogenesis. A preliminary analysis of a single case of hepatoblastoma (an embryonal childhood liver cancer) from a patient diagnosed with Beckwith-Wiedemann syndrome, confirmed the presence of a precursor clone in normal liver, but many more cases are required to investigate the recurrence and robustness of these patterns. This is somewhat hampered by the rarity of hepatoblastoma.

Somatic mutations can be particularly effective when studying the origins of bilateral cancers. Comparison of the mutational patterns between the two lesions can distinguish whether (1) one is a metastasis of the other, (2) both tumours emerge from a common clone or (3) the two tumours have emerged independently (Foulkes and Polak, 2020). For example, the bilateral Wilms tumour cases in this dissertation represent the second scenario, as they arise from a common root of clonal nephrogenesis. However, two cases of bilateral neuroblastoma studied during my PhD revealed that the bilateral lesions only share a few SNVs, all of which shared with bulk blood (Coorens et al., 2020). This indicates an early divergence of tumour lineages during the first few cell divisions of life and hence, an independent emergence of both tumours. This is likely a consequence of germline predisposition mutations. Interestingly, the two tumours often exhibited similar genomic events, such as loss of the second copy of *SMARCA4*.

A last piece of ongoing work studies the origins of secondary acute myeloid leukaemia (AML) or myelodysplastic syndrome (MDS) in two patients after chemotherapy treatment for neuroblastoma and an autologous haemotopoietic stem cell transplant (Coorens et al., 2021a). From the patterns of somatic mutations, it was possible to discern that the primordial clone of the AML and MDS was already present in the stem cell populations prior to transplant and constitutes *bona fide* clonal haematopoiesis. Moreover, the somatic mutations were caused by mutational signatures attributable to platinum-based chemotherapy (Pich et al., 2019), further underpinning the role of the treatment in emergence of these secondary cancers.

This ongoing work further illustrates the versatility and ubiquity in which somatic mutations can be used to answer questions about the origins of human cancers. While the work of my PhD has focused on childhood cancers, these approaches are directly generalisable to adult cancers.

6.2.2 Further methodological advancements

The ongoing decrease in costs for DNA sequencing can facilitate deeper and more extensive sampling strategies to reconstruct phylogenetic trees of normal and aberrant human development. To improve the detection of early somatic variants and the building of phylogenies in such large data sets, two areas of substantial and necessary methodological advancement some to mind.

Firstly, the approach currently taken to call somatic variants is based on the old paradigm of a comparison between a tumour and a normal sample. As an organic extension to this, the somatic variant calling in normal tissues is currently built on algorithms with that paradigm in mind, resulting in the use of an *in silico*-created unmatched normal sample. However, this completely neglects the potential for leveraging the many samples acquired from the same patient. Instead, this wealth of information is only used in the filtering steps I devised, after combining the calls from all pairwise comparisons. In the long run, it might be more time-efficient and computationally powerful to directly combine the genomic information and patterns of mutations across the many samples from the same patient in the initial variant calling stage. Such an amalgamation of whole-genome sequencing data might result in calling of variants with more sensitivity to low VAF mutations by combining variant reads across samples, while simultaneously better handling recurrent artefactual calls and increasing the specificity of calling.

Secondly, while the infinite sites assumption still holds and the maximum parsimony approach for phylogeny was the optimal choice within the confines of this dissertation, the increasing size of data sets and called somatic variants might lead to insurmountable violations of that assumption. As explored in Chapter 3, section 3.10, even at the current scale of sampling and sequencing, mutations re-occur independently within the phylogeny by chance. In addition, algorithms based on maximum parsimony do not scale optimally and very large sets of samples might render this approach intractable. Hence, in the near future, it might be desirable to develop a novel phylogenetic tree building algorithm based on maximum likelihood or Bayesian inference that is specifically tailored to the unique aspects of somatic phylogenetics. Since these approaches require explicit models of genomic sequence evolution, it would provide an opportunity to incorporate our prior knowledge on mutational signatures.

The proposed methodological advancements are by no means straightforward exercises, but they will aid in maturing the field of 'somatic phylogenetics' by further adding robustness to the approaches employed throughout this dissertation.

6.2.3 Closing remarks

This dissertation has explored the vast potential of using somatic mutations to trace the lineages of individual cells to elucidate fundamental processes of human development and the emergence of cancer. The studies presented here, however, only scratch the surface of the myriad problems that can be addressed with the approaches outlined in this thesis. The ubiquity of somatic mutations lends an enormous versatility to the unresolved questions they can answer. Without a doubt, the future of this field holds many exciting opportunities.