DOKTORAND:	
GRAD:	Philosophiae doctor
FAKULTET:	Det matematisk-naturvitenskapelige fakultet
INSTITUTT:	Institutt for informatikk
FAGOMRÅDE:	Bioinformatikk
VEILEDERE:	Geir Kjetil Sandve, Lex Nederbragt, Ingrid Glad, Geir Olve Storvik
DISPUTASDATO:	26. september 2019
AVHANDLINGENS TITTEL:	*Graph-based Reference Genomes – Approaches and Applications*

Are graph-based reference genomes the future of reference genomes? Through three projects, this thesis explores graph-based reference genomes and how they can be used as alternatives to current linear reference genomes.

First, we define coordinate systems on graph-based reference genomes, and propose a
method for representing genomic intervals on these references. Second, we address how graph-based reference genomes can be used to improve the accuracy of read mapping, an important procedure which often is the first step when analysing genomic reads. We compare the existing graph-based read mapping tools and propose a novel two-step approach that we argue represents a promising direction for graph-based read mappers. Finally, we show how graph-based reference genomes can be used to improve detection of regions associated to transcription factor binding from ChIP-seq data, by developing the first peak caller, Graph Peak Caller, that works with graph-based reference genomes.

This thesis shows that graph-based reference genomes can be used to improve detection of regions associated with transcription factor binding from ChIP-seq data. Specifically, our findings show that Graph Peak Caller can trace variants within peaks that are not in the linear reference genome, and that it finds peaks that are more motif-enriched than peaks found by MACS 2. While we have seen that graph-

based reference genomes can improve the analysis of genomic reads – through read mapping and peak calling – we have also experienced that linear reference genomes have certain benefits, e.g. by providing a simple way of representing the locations of genomic features. Based on our findings, it is reasonable to assume that graph-based reference genomes might not fully replace linear reference genomes, but act as good alternatives in many settings.