Accurate genotyping across variant classes and lengths using variant graphs

Genotype estimates from short-read sequencing data are typically based on the alignment of reads to a linear reference, but reads originating from more complex variants (for example, structural variants) often align poorly, resulting in biased genotype estimates. This bias can be mitigated by first collecting a set of candidate variants across discovery methods, individuals and databases, and then realigning the reads to the variants and reference simultaneously. However, this realignment problem has proved computationally difficult. Here, we present a new method (BayesTyper) that uses exact alignment of read k-mers to a graph representation of the reference and variants to efficiently perform unbiased, probabilistic genotyping across the variation spectrum. We demonstrate that BayesTyper generally provides superior variant sensitivity and genotyping accuracy relative to existing methods when used to integrate variants across discovery approaches and individuals. Finally, we demonstrate that including a ‘variation-prior’ database containing already known variants significantly improves sensitivity.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Metagenomics, Integrative Systems Biology, Genomic Epidemiology, Disease Intelligence and Molecular Evolution, University of Copenhagen, South China University of Technology, BGI Europe A/S, BGI-Shenzhen, Technical University of Denmark, University of Oslo, University of Bergen.
Number of pages: 11
Publication date: 2018
Peer-reviewed: Yes

Publication information
Journal: Nature Genetics
ISSN (Print): 1061-4036
Ratings:
BFI (2019): BFI-level 3
Web of Science (2019): Indexed yes
BFI (2018): BFI-level 3
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 21.12 SJR 22.434 SNIP 5.867
Web of Science (2017): Impact factor 27.125
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 20.83 SJR 21.979 SNIP 6.709
Web of Science (2016): Impact factor 27.959
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Web of Science (2015): Impact factor 31.616
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 22.76 SJR 23.98 SNIP 6.332
Web of Science (2014): Impact factor 29.352
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 24.17 SJR 24.193 SNIP 6.287
Web of Science (2013): Impact factor 29.648
ISI indexed (2013): ISI indexed yes