Genetics in Medicine Open (2025) 3, 103433

ELSEVIER

Genetics in Medicine OPEN
An Official Journal of the ACMG

www.journals.elsevier.com/genetics-in-medicine-open

ACMG LAB QA BULLETIN

Laboratory considerations for GRCh37 to GRCh38 reference genome transition: A laboratory quality assurance bulletin of the American College of Medical Genetics and Genomics (ACMG)

Vimla S. Aggarwal[1], Saurav Guha[2], Matthew S. Lebo[3,4], Kyle Retterer[5], Curt Scharfe[6], Lora J.H. Bean[7]; on behalf of the ACMG Laboratory Quality Assurance Committee[8,*]

Disclaimer: This Lab QA bulletin is designed primarily as an educational resource for laboratory geneticists to help them provide quality medical services. Adherence to this Lab QA bulletin is completely voluntary and does not necessarily assure a successful medical outcome. This Lab QA bulletin should not be considered inclusive of all proper procedures, treatments, and tests or exclusive of other procedures, treatments, and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific procedure, treatment, or test, laboratory geneticists should apply their own professional judgment to the specific clinical circumstances presented by the individual patient or specimen.
Laboratory geneticists are encouraged to document the reasons for the use of a particular procedure, treatment, or test, whether or not it is in conformance with this Lab QA bulletin. Laboratory geneticists also are advised to take notice of the date this Lab QA bulletin was adopted and to consider other medical and scientific information that becomes available after that date. It also would be prudent to consider whether intellectual property interests may restrict the performance of certain tests and other procedures. Where individual authors are listed, the views expressed may not reflect those of authors' employers or affiliated institutions.
The mention of any testing approach, product, or sponsor in this Lab QA bulletin does not constitute endorsement or sponsorship by the American College of Medical Genetics and Genomics. The American College of Medical Genetics and Genomics does not endorse or recommend any specific testing approach or product mentioned in this Lab QA bulletin.

ARTICLE INFO

Article history:
Received 14 April 2025
Accepted 15 April 2025
Available online xxx

Keywords:
Clinical next-generation sequencing
Genome assembly
Human genome reference
Sequence alignment tools
Variant annotation

Background

Since the initial publication from the Human Genome Project, successive human genome reference versions, called assemblies or builds, have been released. Each new version has benefited from technological advancements and additional data, improving both the overall quality and the representation from diverse populations. The most recent build, Genome Research Consortium (GRC) Human Build 38 (GRCh38), also known as hg38, was released in 2013, 4 years after the previous build, GRCh37 (hg19). Because of the timing of this release and that of the broad availability of next-generation sequencing, many clinical bioinformatics

This article was a work product of the Laboratory Quality Assurance Committee of the ACMG and the Article Publishing Charge (APC) was waived. No industry sponsorship was received for this work.
The Board of Directors of the American College of Medical Genetics and Genomics approved this bulletin on 24 February 2025.
*Correspondence: ACMG. *Email address:* documents@acmg.net
Affiliations are at the end of the document.

doi: https://doi.org/10.1016/j.gimo.2025.103433

pipelines were developed and validated using the GRCh37 reference sequence. Thus, clinical laboratories are unlikely to have previous validation schemes for a full validation of a reference genome transition. The complexity of genomic data and the software systems that manage that data make designing a comprehensive validation plan for core feature changes challenging.

GRCh37 vs GRCh38

A primary goal in the development of GRCh38 was to provide a more accurate and comprehensive representation of the human genome. Major improvements from GRCh37 include closure of sequence gaps, correction of sequence errors, removal of rare pathogenic variants, and the introduction of alternative loci to better represent population-level variation.[1] Although these changes have favored selecting the most common alleles for the reference, some clinically relevant variants, such as those associated with pharmacogenetic response or risk alleles, are now the reference base (Table 1[1-7]). The internal assumptions or exception handling of informatics pipelines may not correctly handle such differences.

GRCh38 also introduced changes to the representation of the population-level structural complexity of the human genome that can affect secondary and tertiary analysis pipelines and processes. Naive use of existing pipelines without consideration for multiple potential mappings via "alt-aware," graph alignment, or strategic masking of GRCh38 will lead to analytical dropout of these regions (Table 1[1-7]). Informatics processes for selecting alternate contigs and/or duplicated genes will need to be developed and validated.

Validation and production considerations

Clinical validation typically focuses on accurate detection and annotation of clinically relevant variants. When changing a core feature of the variant calling algorithm, identification and investigation of all variants called and annotated from the larger, unfiltered data set should be considered. Unexpected behavior after a reference set change may be systematic or manifest only as rare events.[8] Therefore, any discrepancy in base calling may be informative. It is critical for the laboratory to understand all recognized and potential implications of a reference set change and remain vigilant even after validation.

Besides improved variant detection resulting from more accurate and complete reference genome representation, laboratories switching to GRCh38 can benefit from additional external resources that are not available on the older GRCh37 reference. Importantly, pipeline and process updates will be needed to take advantage of the following external resources available on GRCh38:

- Transitioning to standardized transcripts, such as the MANE Select and MANE Plus Clinical transcripts, which are based on GRCh38,[9] will require careful reconciliation of variant annotations when working with legacy GRCh37 data. Although mapped versions of MANE transcripts are available for GRCh37, these may differ due to sequence or alignment differences between genome builds and should be interpreted with caution, especially in clinical contexts.
- gnomAD v4.1, mapped to GRCh38, contains approximately 5x more genomes than previous versions and pipelines must be prepared for larger data sets.
- Relevant databases and new data sets prepared for publication converted to GRCh38, may introduce

Table 1 Technical challenges associated with migrating reference sequences

Type	Key Challenges	Example
Primary sequence	Variant annotations may differ between genome builds due to sequence changes and alignment shifts. Use of secondary validation tools such as VariantValidator or ClinGen Allele Registry is recommended to confirm equivalence	*F5* (HGNC:3542), NM_000130.5:c.1601G>A; NP_000121.2:p.(R534Q): alternate and reference switch (GRCh37:1:169519049:T:T; GRCh38:1:169549811:C:T)[2,3,a]
	Introduction of new reference content including from previously uncharacterized regions	GRCh38 includes additional centromeric, telomeric, and other previously unsequenced or misassembled regions[1,4,b]
	Changes in annotation	Chr1 vs 1, centromeres are not labeled as N's[1,4]
Copy-number and structural variants	Copy-number variation between GRCh37 and GRCh38 includes both additional complexity and correction of assembly errors	Variants not called in regions affected by false duplication in GRCh38. GRCh38 correction of internal inversion (eg, PTPRQ)[3,5-7]
	GRCh38 includes previously unsequenced genomic regions and large structural variants, either of which may be complex or represented on alternate (alt) contigs	Major histocompatibility loci (MHC) GRCh37: 9 alt contigs at 3 genomic loci; GRCh38: 261 alt contigs across 60Mb[1,4]
	Correction of GRCh37 structure impacts liftover of clinically relevant regions	*INPP5D*: Sequence correction in GRCh38; *TMEM236* and *MRC1*: Correction of false intergenic gap[1]

[a]Tools such as VariantValidator, Mutalyzer, ClinGen Allele Registry, and the hgvs Python library can assist with verifying and harmonizing variant descriptions across genome builds and transcript versions. Incorporating such tools for secondary validation is especially important when classifying variants as potentially pathogenic or reporting them clinically.

[b]The mitochondrial genome is unchanged between GRCh37 and GRCh38, but UCSC's hg19 uses an outdated mitochondrial reference (NC_001807), which can affect variant alignment and annotation.

inconsistencies or errors that affect variant identification and nomenclature.
- GRCh38-specific features of vendor-supplied software that may affect secondary and tertiary analysis processes.
- Masking false gene duplications introduced into GRCh38, including medically relevant genes (CBS [HGNC:1550], KCNE1 [HGNC:4260], and CRYAA [HGNC:2388]).[5]

Finally, successful transition may require some or all information systems to support both GRCh38 and GRCh37, potentially indefinitely. Laboratory functions from reporting to client services must be prepared to handle familial testing, reanalysis, and updated reports that span the reference sequence transition, as well as the potential for identification of variants initially missed because of limitations in technology or knowledge. Liftover tools[10] may be validated and incorporated into (semi)-automated pipelines or validated processes to accurately translate older nomenclature to the current reference.

Future considerations

The transition to GRCh38 will not be a 1-time activity. The GRC has and continues to update GRCh38 (https://www.ncbi.nlm.nih.gov/grc/human) and, although many issues have been resolved (>11,000 to date), this is an evolving process. Several unresolved issues, including additional collapsed regions in the GRCh38 reference, have been identified, and efforts are underway to address these through initiatives such as the Telomere-to-Telomere CHM13 genome sequence and the human pangenome.[11,12] Similar validation methods will be needed as data from advancing technologies, such as long-read sequencing, are added to both references and external databases.

Acknowledgments

The authors thank Sandor Roberts and Honey Reddi from the American College of Medical Genetics and Genomics Lab QA Committee who helped with the selection of the authors and the review of the final manuscript. The authors were involved with the review of the data and the development of the manuscript.

Conflict of Interest

All workgroup members receive a salary for providing clinical services that may be relevant to the content of this document in either the laboratory or patient care setting at their listed affiliations.

Affiliations

[1]Columbia University Irving Medical Center, New York, NY; [2]New York Genome Center, New York, NY; [3]Mass General Brigham, Brigham and Women's Hospital, Harvard Medical School, Cambridge, MA; [4]Broad Institute of MIT and Harvard, Cambridge, MA; [5]Geisinger, Danville, PA; [6]Yale School of Medicine, New Haven, CT; [7]Revvity Omics Inc, Pittsburgh, PA; [8]American College of Medical Genetics and Genomics, Bethesda, MD

References

1. Schneider VA, Graves-Lindsay T, Howe K, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res*. 2017;27(5):849-864. http://doi.org/10.1101/gr.213611.116
2. Lansdon LA, Cadieux-Dion M, Yoo B, et al. Factors affecting migration to GRCh38 in laboratories performing clinical next-generation sequencing. *J Mol Diagn*. 2021;23(5):651-657. http://doi.org/10.1016/j.jmoldx.2021.02.003
3. Lansdon LA, Cadieux-Dion M, Herriges JC, et al. Clinical validation of Genome Reference Consortium human build 38 in a laboratory utilizing next-generation sequencing technologies. *Clin Chem*. 2022;68(9):1177-1183. http://doi.org/10.1093/clinchem/hvac113
4. Guo Y, Dai Y, Yu H, Zhao S, Samuels DC, Shyr Y. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics*. 2017;109(2):83-90. http://doi.org/10.1016/j.ygeno.2017.01.005
5. Wagner J, Olson ND, Harris L, et al. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat Biotechnol*. 2022;40(5):672-680. http://doi.org/10.1038/s41587-021-01158-1
6. Li H, Dawood M, Khayat MM, et al. Exome variant discrepancies due to reference-genome differences. *Am J Hum Genet*. 2021;108(7):1239-1250. http://doi.org/10.1016/j.ajhg.2021.05.011
7. Miller CA, Walker JR, Jensen TL, et al. Failure to detect mutations in U2AF1 due to changes in the GRCh38 reference sequence. *J Mol Diagn*. 2022;24(3):219-223. http://doi.org/10.1016/j.jmoldx.2021.10.013
8. Jia T, Munson B, Lango Allen H, Ideker T, Majithia AR. Thousands of missing variants in the UK Biobank are recoverable by genome realignment. *Ann Hum Genet*. 2020;84(3):214-220. http://doi.org/10.1111/ahg.12383
9. Morales J, Pujar S, Loveland JE, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature*. 2022;604(7905):310-315. http://doi.org/10.1038/s41586-022-04558-8
10. Park KJ, Yoon YA, Park JH. Evaluation of liftover tools for the conversion of Genome Reference Consortium human build 37 to build 38 using ClinVar variants. *Genes (Basel)*. 2023;14(10):1875. http://doi.org/10.3390/genes14101875
11. Aganezov S, Yan SM, Soto DC, et al. A complete reference genome improves analysis of human genetic variation. *Science*. 2022;376(6588):eabl3533. http://doi.org/10.1126/science.abl3533
12. Liao WW, Asri M, Ebler J, et al. A draft human pangenome reference. *Nature*. 2023;617(7960):312-324. http://doi.org/10.1038/s41586-023-05896-x