Dear editors of Genome Research and reviewers,

Thank you very much for reviewing our manuscript once again and for providing final important suggestions and criticism. Please find the new version of the manuscript with changes and improvements that address them.

We enhanced the telomere mapping approach to incorporate the suggestions by Reviewer 2, which allowed us to unambiguously assign telomeric reads to specific chromosomal arms. To that end, we (1) refined the construction of the extended reference genome by masking subtelomere-telomere boundaries that were previously identified as inexact and (2) incorporated a step of the pipeline that filters out ambiguous candidates. Please also find point-by-point replies to the comments made by Reviewer 2 at the end of this document.

We are glad to report that these actions, in fact, improved the analysis without affecting the conclusions of the paper. While the identities and the amount of the candidate reads changed due to the disambiguation step, the overall results remained the same: motif variants and spectra of haplotypes are still captured, and the measure of interpopulation similarity has even grown slightly larger.

Moreover, several of the concerns we previously raised in the Discussion about incorrect annotation of the human reference genome were addressed at the early stage of the extended reference construction. Importantly, these changes also allowed us to capture a set of well-supported *p*arm telomeric reads and tobe able to analyze and describe haplotypic variation on *p*arms, which provides additional results.

The supplemental files containing the refined extended genome have been updated, and all numerical and qualitative results that were affected by these changes (e.g., exact values of sequence percentage explained by specific motifs), along with the figures, have been updated in the main text and in the supplemental materials as well.

We also addressed all suggestions regarding the formatting of the manuscript and the availability of code: (1) modified the capitalization of terms (10x, Chromosome), (2) re-formatted all citations and references to adhere to the journal’s format, (3) re-rendered figures to be more legible upon reduction, (4) supplied Python scripts as Supplemental File S3, and (5) adjusted the flow of the narrative as much as possible to include the updated results and methods in the appropriate sections of the manuscript.

**Responses to comments by Reviewer #2:**

A: We would like to thank the referee for the careful assessment of our manuscript and the constructive criticism.

*Q: With the revision, it is clear that edgeCase has features that will make it a very useful tool for analyzing telomere reads and telomere motifs in sufficiently long reads. However, the inherent limitations of subtelomeric reference sequences still make interpretation of many results in the datasets used in this paper problematic. […] Ideally, the authors would: (1) identify all boundary-spanning reads in a high depth-of-coverage Genomic PacBio dataset by virtue of their long telomere/telomere-like repeat tracts adjacent to 3-5 kb or more of non-telomere repeat DNA (distal subtelomere DNA). This sequence organization should only occur at the boundaries of subtelomere/telomere repeat in the human genome, with the one exception of 2q fusion site which would be readily identifiable by the head to head orientation of telomere repeats that has been identified previously. All other known internal telomere-like repeat sequences are short (<500 bp) and degenerate.*

A: Thank you for drawing attention to this important point; we updated the edgeCase pipeline by implementing your suggestions. Specifically, we incorporated a disambiguation step that filters out reads mapping to more than one chromosomal arm, and only retains the candidates overlapping the subtelomere-telomere features by at least 3Kbp. At the same time, we refined the construction of the extended genome reference by once more carefully considering the existing subtelomere assemblies, and selecting for analyses only those that are precisely placed in the genome and terminate at the subtelomere-telomere boundary. Combined, these two actions allowed us to narrow down the groups of candidate reads that represent sequences of specific telomeres in each subject; of note, the 2q arm was not present in the resultant set, making this particular point of disambiguation straightforward.

Q: *[Ideally, the authors would:] (2) Use the distal subtelomere DNA to group the boundary reads into distinct sets with identical or near-identical distal subtelomere sequence. At the highest resolution, there would theoretically be 92 discernable groups, each representing a unique boundary allele at the start of a terminal (TTAGGG)n tract. Based on what we already know about distal subtelomere sequences, the actual number of discernable groups would be much smaller due to several families of distal subtelomeric segmental duplications, that have within-family nucleotide sequence similarities of 97-100%. However, there should also be a number of distal subtelomere sequences lacking the highly similar duplicons and these could be uniquely mapped to some subtelomeres (e.g., 2p, 7q, 8q, 11q, 12q, 14q, 17p, 18q, Xp/Yp, possibly some others) and analyzed using EdgeCase. Edgecase may also enable subdivision of individual duplicon families based on patterns of telomere repeats and interspersed telomere-like repeats occurring in the linked telomere repeat tracts, which would also be a valuable use of it.*

A: Thank you again for this important suggestion. Owing to the improved selection of unambiguously mapping reads in the step we describe in the answer to the previous question, we were now able to classify the candidate reads into groups mapping to their respective unique distal subtelomere sequences. Eight out of nine of the expected arms listed in the suggestion were captured in this manner (2p, 7q, 8q, 11q, 12q, 14q, 17p, 18q), with the only exception of Xp/Yp, as the Xp/Yp subtelomeric contig mapped poorly to the reference genome and was excluded from the extended reference. Six additional arms were also captured in this manner (3p, 4p, 5p, 9p, 12p, 15q). Moreover, further clustering of per-arm haplotype variation by pairwise edit (Levenshtein) distances considered both the telomeric and the subtelomeric sequences of reads, thus factoring in the similarity of the distal subtelomeric sequence as well. As a result, we were now able to present a refined picture of sequence variation on *q-* as well as *p*-arms, in which patterns of canonical and non-canonical telomeric motifs describe intra- and inter-subject, intra- and inter-population variation.

We hope that addressing your comments, suggestions, and criticism with these changes is satisfactory to the reviewers and the editors.