

## Supplementary information

---

# The Human Pangenome Project: a global resource to map genomic diversity

---

In the format provided by the authors and unedited

**The Human Pangenome Project: A Global Resource to Map Genomic Diversity**  
**Supplementary Information Guide**

**Item 1: Supplemental Table 1. 47 genomes comprise the initial pangenome data resources.**

**Item 2: Current Membership of the Human Pangenome Reference Consortium**

**Supplemental Table 1. 47 genomes comprise the initial pangenome\*.**

Sample	Cohort	HiFi	ONT	ONT >100kb	Bionano	HiC	Child	Mat.	Pat.
HG00438	HPRC	41X	44X	16X	162X	93X	36X	36X	33X
HG00621	HPRC	40X	34X	14X	243X	95X	35X	38X	44X
HG00673	HPRC	40X	37X	13X	163X	98X	34X	32X	36X
HG00735	HPRC	43X	33X	12X	167X	82X	35X	37X	36X
HG00741	HPRC	38X	47X	18X	127X	77X	37X	36X	36X
HG01071	HPRC	35X	43X	16X	167X	78X	37X	35X	55X
HG01106	HPRC	47X	45X	15X	164X	73X	37X	45X	33X
HG01123	HPRC	39X	17X	6X	164X	62X	-	45X	35X
HG01175	HPRC	36X	48X	18X	161X	78X	35X	36X	34X
HG01258	HPRC	36X	23X	10X	105X	74X	37X	36X	39X
HG01358	HPRC	38X	21X	6X	163X	63X	38X	34X	37X
HG01361	HPRC	42X	35X	11X	169X	54X	38X	36X	35X
HG01891	HPRC	37X	39X	16X	165X	48X	37X	33X	36X
HG01928	HPRC	35X	27X	11X	109X	81X	38X	38X	36X
HG01952	HPRC	41X	33X	12X	163X	70X	37X	34X	33X
HG01978	HPRC	38X	41X	16X	167X	103X	40X	36X	33X
HG02148	HPRC	40X	20X	6X	164X	72X	38X	37X	33X
HG02257	HPRC	36X	27X	10X	167X	59X	38X	34X	36X
HG02486	HPRC	40X	31X	9X	163X	45X	-	35X	37X
HG02559	HPRC	41X	26X	8X	162X	56X	-	37X	35X
HG02572	HPRC	31X	17X	7X	162X	133X	37X	37X	37X
HG02622	HPRC	46X	15X	7X	110X	61X	39X	34X	36X
HG02630	HPRC	49X	21X	9X	104X	64X	40X	35X	36X
HG02717	HPRC	45X	29X	10X	107X	85X	39X	37X	39X
HG02886	HPRC	44X	34X	14X	161X	67X	41X	47X	34X
HG03453	HPRC	52X	12X	5X	109X	70X	35X	43X	35X
HG03516	HPRC	36X	46X	17X	170X	57X	33X	35X	36X
HG03540	HPRC	49X	31X	13X	168X	62X	40X	36X	36X
HG03579	HPRC	48X	17X	7X	165X	83X	35X	36X	35X
HG002	HPRC_PLUS	178X	710X	66X	102X	382X	33X	33X	33X
HG005	HPRC_PLUS	47X	58X	5X	-	-	51X	115X	113X
HG00733	HPRC_PLUS	33X	78X	5X	90X	64X	36X	35X	35X
HG01109	HPRC_PLUS	32X	71X	9X	158X	24X	37X	36X	39X
HG01243	HPRC_PLUS	35X	60X	6X	155X	30X	35X	36X	36X
HG02055	HPRC_PLUS	40X	65X	7X	156X	32X	37X	33X	34X
HG02080	HPRC_PLUS	34X	56X	6X	144X	32X	37X	34X	33X
HG02109	HPRC_PLUS	31X	-	-	222X	40X	-	36X	45X
HG02145	HPRC_PLUS	36X	-	-	142X	37X	36X	33X	47X
HG02723	HPRC_PLUS	46X	-	-	155X	25X	39X	36X	36X
HG02818	HPRC_PLUS	37X	-	-	-	-	38X	35X	38X
HG03098	HPRC_PLUS	36X	57X	2X	114X	26X	36X	37X	34X
HG03486	HPRC_PLUS	39X	-	-	-	-	37X	34X	33X
HG03492	HPRC_PLUS	34X	51X	3X	155X	34X	34X	37X	36X
NA18906	HPRC_PLUS	42X	-	-	-	-	35X	36X	36X
NA19240	HPRC_PLUS	40X	-	-	-	-	35X	45X	35X
NA20129	HPRC_PLUS	38X	-	-	-	-	36X	34X	36X
NA21309	HPRC_PLUS	39X	-	-	-	-	45X	32X	33X
HPRC Median		40X	31X	11X	163X	72X	37X	36X	36X
All Median		39X	35X	10X	162X	64X	37X	36X	36X

\*Of the listed samples, 29 had data produced by the consortium including: (i) Pacific Biosciences HiFi long-read data, (ii) Oxford Nanopore Technologies long-read data (11X >= 100KB), (iii) Hi-C contact maps, and (iv) Bionano physical maps. All DNA was collected from early passages of

LCLs. Seven of the samples also have Strand-Seq data for phasing and detecting inversions. Each assembled sequence in the data release is from the offspring of a trio. Trio sequencing data (Illumina NovaSeq from TruSeq PCR-free libraries) from the New York Genome Center's High Coverage 1KG dataset<sup>1</sup> were used for assembly and QC and included in the release. The remaining 18 assembled samples in the release are aggregated from data produced outside of the HPRC consortium including GIAB and the Human Genome Structural Variation Consortium (HGSVC), listed above as HPRC\_PLUS.

- 1 Byrska-Bishop, M. *et al.* High coverage whole genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *bioRxiv*, 2021.2002.2006.430068, doi:10.1101/2021.02.06.430068 (2021).

## Supplementary Information

### The Human Pangenome Project: A Global Resource to Map Genomic Diversity

#### Current Membership of the Human Pangenome Reference Consortium

##### The Human Pangenome Reference Consortium Coordination Center

Lucinda Antonacci-Fulton<sup>1</sup>, Eddie Belter<sup>1</sup>, Sarah Cody<sup>1</sup>, Changxu Fan<sup>1,2,3</sup>, Paul Flicek<sup>4</sup>, Ira M. Hall<sup>5</sup>, David Haussler<sup>6,7</sup>, Heather A. Lawson<sup>1,2,3</sup>, Daofeng Li<sup>1,2,3</sup>, Joshua F. McMichael<sup>1</sup>, Karen H. Miga<sup>6</sup>, Benedict Paten<sup>6</sup>, Chad Tomlinson<sup>1</sup>, Deepak Purushotham<sup>1,2,3</sup>, Ting Wang<sup>1,2,3</sup>, Ann Zhang<sup>1,2,3</sup>

##### Sample Working Group including Teams for Population Genetics and Ethical, Legal, and Social Issues

Carlos Bustamante<sup>8</sup>, Judy Cho<sup>9,10,11</sup>, Robert Cook-Deegan<sup>12</sup>, Jean-Francois Deleuze<sup>13</sup>, Richard Durbin<sup>14,15</sup>, Simon Easteal<sup>16</sup>, Evan E. Eichler<sup>17,18</sup>, Xiaowen Feng<sup>19,20</sup>, Nanibaa Garrison<sup>21,22,23</sup>, Nadine Gassner<sup>6</sup>, Mary Goldman<sup>6</sup>, Ed Green<sup>6</sup>, David Haussler<sup>6,7</sup>, Erich D. Jarvis<sup>24,25</sup>, Eimear E. Kenny<sup>9,11</sup>, Barbara A. Koenig<sup>26</sup>, Bastien Llamas<sup>27,28</sup>, Nicole C. Lockhart<sup>29</sup>, Bartha M. Knoppers<sup>30</sup>, Ann M. McCartney<sup>31</sup>, Karen H. Miga<sup>6</sup>, Jessica Mozersky<sup>32</sup>, Hardip Patel<sup>27,28</sup>, Alice B. Popejoy<sup>33</sup>, Charles Rotimi<sup>34</sup>, Charmaine Royal<sup>35</sup>, Yassine Souilmi<sup>27,28</sup>, Nathan O Stitzel<sup>1,2,36</sup>, Lisa Wang<sup>9,11</sup>

##### Technology and Production Working Group

Mark Akeson<sup>6</sup>, Brandy Baird<sup>6</sup>, Giulio Formenti<sup>24,25</sup>, Robert S. Fulton<sup>1</sup>, Ed Green<sup>6</sup>, Miten Jain<sup>6</sup>, Brittany Kerr<sup>37</sup>, Chris Markovic<sup>1</sup>, Matthew W. Mitchell<sup>37</sup>, Katy Munson<sup>17</sup>, Hugh Olsen<sup>6</sup>, Sadye Paez<sup>24,25</sup>, William Rowell<sup>38</sup>, Sam Sacco<sup>39</sup>, Lauren Shalmiyev<sup>24,25</sup>, Arvis Sulovari<sup>17</sup>

##### Assembly, T2T, and Pangenome Working Group

Mobin Asri<sup>6</sup>, Pete Audano<sup>17</sup>, Paolo Carnevali<sup>40</sup>, Mark Chaisson<sup>41</sup>, Shubham Chandak<sup>42</sup>, Xian Chang<sup>6</sup>, Haoyu Cheng<sup>19,20</sup>, Vincenza Colonna<sup>43</sup>, Daniel Doerr<sup>44</sup>, Peter Ebert<sup>44</sup>, Jana Ebler<sup>44</sup>, Evan E. Eichler<sup>17,18</sup>, Jordan Eizenga<sup>6</sup>, Olivier Fedrigo<sup>24,25</sup>, Xiaowen Feng<sup>19,20</sup>, Christian Fischer<sup>45</sup>, Stacey Gabriel<sup>46</sup>, Yan Gao<sup>47</sup>, Shilpa Garg<sup>19,20,48</sup>, Kiran Garimelle<sup>46</sup>, Erik Garrison<sup>45</sup>, Ed Green<sup>6</sup>, Stephanie Greer<sup>49</sup>, Andrea Guarracino<sup>50</sup>, Ira M. Hall<sup>5</sup>, William Harvey<sup>17</sup>, Marina Haukness<sup>6</sup>, David Haussler<sup>6,7</sup>, Simon Heumos<sup>51</sup>, Glenn Hickey<sup>6</sup>, Kerstin Howe<sup>15</sup>, Eric D. Jarvis<sup>24,25</sup>, Hanlee Ji<sup>49</sup>, Sergey Koren<sup>31</sup>, Hojoon Lee<sup>42</sup>, Heng Li<sup>19,20</sup>, Wen-Wei Liao<sup>5</sup>, Ryan Lorig-Roach<sup>6</sup>, Ernesto Lowy<sup>4</sup>, Tony Tsung Yu Lu<sup>41</sup>, Shuangjia Lu<sup>5</sup>, Julian Lucas<sup>6</sup>, Rebecca Serra Mari<sup>44</sup>, Dmitri Pavlichin<sup>49</sup>, Pierre Marijon<sup>44</sup>, Charles Markello<sup>6</sup>, Tobias Marschall<sup>44</sup>, Melissa Merediths<sup>6</sup>, Karen H. Miga<sup>6</sup>, Jean Monlong<sup>6</sup>, Njagi Mwaniki<sup>45,52</sup>, Eugene W. Myers<sup>53,54,55</sup>, Adam M. Novack<sup>6</sup>, Sergey Nurk<sup>31</sup>, Benedict Paten<sup>6</sup>, Dmitri Pavlichin<sup>42</sup>, Trevor Pesout<sup>6</sup>, Adam M. Phillippy<sup>31</sup>, Brandon Pickett<sup>31</sup>, David Porubsky<sup>17</sup>, Pjotr Prins<sup>45</sup>, Mikko Rautiainen<sup>31</sup>, Arang Rhie<sup>31</sup>, Kishwar Shafin<sup>6</sup>, Jonas Sibbesen<sup>6</sup>, Jouni Siren<sup>6</sup>, Varsha Sreekanth<sup>6</sup>, Arvis Sulovari<sup>17</sup>, Kedar Tatwawadi<sup>42</sup>, Flavia Villani<sup>41</sup>, Mitchell Vollger<sup>17</sup>, Alexander Wait Zaranek<sup>48</sup>, Tsachy Weissman<sup>42</sup>

##### Annotation, Maintenance and Improvement Working Group

Derek Albracht<sup>1</sup>, Eddie Belter<sup>1</sup>, Shelby Bidwell<sup>56</sup>, Konstantinos Billis<sup>4</sup>, Caryn Carson<sup>1,2,3</sup>, Karen Clark<sup>56</sup>, Mark Diekhans<sup>6</sup>, Sarah Dyer<sup>4</sup>, Susan Fairley<sup>4,57</sup>, Paul Flicek<sup>4</sup>, Adam Frankish<sup>4</sup>, Nadine Gassner<sup>6</sup>, Carlos Garcia Giron<sup>4</sup>, Mary Goldman<sup>6</sup>, Tina A. Graves-Lindsay<sup>1</sup>, Marina Haukness<sup>6</sup>, Kevin Howe<sup>15</sup>, Sarah Hunt<sup>4</sup>, Paul Kitts<sup>56</sup>, Milinn Kremitzki<sup>1</sup>, Fergal Martin<sup>23</sup>, Terence Murphy<sup>30</sup>, Valerie Schneider, Francoise Thibaud-Nissen<sup>30</sup>, Sergey Nurk<sup>13</sup>, David Thybert<sup>4</sup>, Thomas Walsh<sup>4</sup>, Ting Wang<sup>1,2,3</sup>, Chunlin Xiao<sup>56</sup>, Daniel Zerbino<sup>4</sup>, Xiaoyu Zhuo<sup>1,2,3</sup>

##### National Human Genome Research Institute

Adam L. Felsenfeld<sup>58</sup>, Carolyn M. Hutter<sup>58</sup>, Nicole C. Lockhart<sup>58</sup>, Baergen I. Schultz<sup>58</sup>, Michael W. Smith<sup>58</sup>, Heidi J. Sofia<sup>58</sup>

#### Associate Members

Omar Eduardo Cornejo Ordaz<sup>59</sup>, Simon Easteal<sup>16</sup>, Obed Amar Garcia<sup>33</sup>, George Liu<sup>60</sup>, Hardip Patel<sup>27,28</sup>, Jeffrey Rosenfeld<sup>61</sup>, Steven Salzberg<sup>62</sup>, Alaina Shumate<sup>62</sup>

#### **Affiliations:**

1. McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA
2. Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA
3. Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO, USA
4. European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK
5. Yale School of Medicine, New Haven, CT, USA
6. UC Santa Cruz Genomics Institute, University of California, Santa Cruz, CA, USA
7. Howard Hughes Medical Institute, University of California, Santa Cruz, CA, USA
8. Department of Biomedical Data Science, Genetics and Biology, Stanford University School of Medicine, Stanford, CA, USA
9. Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
10. Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
11. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
12. Arizona State University, Barrett & O'Connor Washington Center, Washington DC, USA
13. Centre National de Recherche en Génomique Humaine, Institut François Jacob, Direction de la Recherche Fondamentale, CEA, Evry, France
14. Department of Genetics, University of Cambridge, Cambridge, UK
15. Wellcome Sanger Institute, Cambridge, UK
16. Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, National Centre for Indigenous Genomics, Australian National University, Canberra, Australia
17. Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA
18. Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA
19. Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA
20. Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
21. Institute for Society & Genetics, University of California, Los Angeles, CA, USA
22. Institute for Precision Health, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
23. Division of General Internal Medicine & Health Services Research, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
24. Vertebrate Genome Lab, The Rockefeller University, New York, NY, USA
25. Laboratory of Neurogenetics of Language, The Rockefeller University, New York, NY, USA
26. Division of Genetics, University of California, San Francisco, San Francisco, CA, USA
27. Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, Australia

28. National Centre for Indigenous Genomics, Australian National University, Canberra, Australia
29. National Institutes of Health (NIH)-National Human Genome Research Institute, Bethesda, MD, USA
30. Centre of Genomics and Policy, McGill University, Montreal, Québec, Canada
31. Genome Informatics Section, National Human Genome Research Institute, Bethesda, MD, USA
32. Division of General Medical Sciences, Washington University School of Medicine, St. Louis, MO, USA
33. Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA
34. Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, MD, USA
35. Department of African & African American Studies, Duke University, Durham, NC, USA
36. Department of Medicine, Washington University School of Medicine, St Louis, MO, USA
37. Coriell Institute for Medical Research, Camden, NJ, USA
38. Pacific Biosciences, Menlo Park, CA, USA
39. Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA
40. Chan Zuckerberg Initiative, Redwood City, CA, USA
41. Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA
42. Department of Electrical Engineering, Stanford University, Stanford, CA, USA
43. Institute of Genetics and Biophysics, Consiglio Nazionale di Ricerche, Naples, Italy
44. Heinrich Heine University, Medical Faculty, Institute for Medical Biometry and Bioinformatics, Düsseldorf, Germany
45. Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN, USA
46. The Broad Institute of Harvard and MIT, Cambridge, MA, USA
47. Department of Critical Care Medicine, The First Affiliated Hospital of Harbin Medical University, Heilongjiang Province, China
48. Department of Genetics, Harvard Medical School, Boston, MA, USA
49. Department of Medicine, Stanford University, School of Medicine, Stanford, CA, USA
50. IIGM - Italian Institute for Genomic Medicine, c/o IRCSS, Candiolo, Italy, Department of Biology, University of Rome "Tor Vergata", Rome, Italy
51. Quantitative Biology Center (QBiC), University of Tübingen, Tübingen 72076, Germany
52. Kemri Wellcome Trust, Kenya
53. Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
54. Center for Systems Biology, Dresden, Germany
55. Faculty of Computer Science, Technical University Dresden, Dresden, Germany
56. National Center for Biotechnology Information (NCBI), National Library of Medicine, Bethesda, MD, USA
57. Global Alliance for Genomics and Health, Toronto, Canada
58. National Institutes of Health (NIH)-National Human Genome Research Institute, Bethesda, MD, USA
59. School of Biological Sciences, Washington State University, Pullman, WA, USA
60. US Department of Agriculture (USDA) Agricultural Research Service (ARS), Beltsville, MD, USA
61. Rutgers Cancer Institute and the Department of Pathology and Laboratory Medicine at the Robert Wood Johnson School of Medicine, USA
62. Johns Hopkins University, Baltimore, MD, USA