

# Package ‘kgp’

December 21, 2022

**Type** Package

**Title** 1000 Genomes Project Metadata

**Version** 1.1.1

**Description** Metadata about populations and data about samples from the 1000 Genomes Project, including the 2,504 samples sequenced for the Phase 3 release and the expanded collection of 3,202 samples with 602 additional trios. The data is described in Auton et al. (2015) <[doi:10.1038/nature15393](https://doi.org/10.1038/nature15393)> and Byrsk-Bishop et al. (2022) <[doi:10.1016/j.cell.2022.08.004](https://doi.org/10.1016/j.cell.2022.08.004)>, and raw data is available at <<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/>>. See Turner (2022) <[doi:10.48550/arXiv.2210.00539](https://doi.org/10.48550/arXiv.2210.00539)> for more details.

**URL** <https://github.com/stephenturner/kgp>,  
<https://stephenturner.github.io/kgp/>

**License** Apache License (>= 2)

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.2.2

**Depends** R (>= 2.10)

**Suggests** tibble

**NeedsCompilation** no

**Author** Stephen Turner [aut, cre] (<<https://orcid.org/0000-0001-9140-9028>>)

**Maintainer** Stephen Turner <[vustephen@gmail.com](mailto:vustephen@gmail.com)>

**Repository** CRAN

**Date/Publication** 2022-12-21 12:00:02 UTC

## R topics documented:

allmeta . . . . .	2
kgp3 . . . . .	3
kgpe . . . . .	4
kgpmeta . . . . .	5

<b>Index</b>	<b>6</b>
--------------	----------

---

allmeta

*1000 Genomes, SGDP, HGDP, and GGVP metadata*

---

### Description

Population metadata from 212 populations from the 1000 Genomes Project (kgp), Simons Genome Diversity Project (sgdp), Human Genome Diversity Project (hgdp), and Gambian Genome Variation Project (ggvp).

### Usage

allmeta

### Format

A tibble with 212 rows and 8 columns:

**pop** Short population code

**reg** Short region code

**population** Long population description

**region** Long region description

**regcolor** Color for plotting this region on a map

**lat** Population latitude

**lng** Population longitude

**dataset** Which dataset (kgp = 1000 Genomes Project; ggvp = Gambian Genome Variation Project; hgdp = Human Genome Diversity Project; Simons Genome Diversity Project).

### References

Byrska-Bishop, Marta, et al. "High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios." *Cell* 185.18 (2022): 3426-3440.

1000 Genomes Project Consortium. "A global reference for human genetic variation." *Nature* 526.7571 (2015): 68.

Clarke, Laura, et al. "The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data." *Nucleic acids research* 45.D1 (2017): D854-D859.

License information is available at [https://github.com/igsr/1000Genomes\\_data\\_indexes/blob/master/LICENSE](https://github.com/igsr/1000Genomes_data_indexes/blob/master/LICENSE). The 1000 Genomes data is made publicly available according to the Fort Lauderdale Agreement (<https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>).

---

kgp3

*1000 Genomes Project sample data (Phase 3)*

---

### Description

Sample, pedigree, and population data for 2,504 samples in the Phase 3 release of the 1000 Genomes Project data.

### Usage

kgp3

### Format

A tibble with 2504 rows and 10 columns:

**fid** Family ID

**id** Individual ID

**pid** Paternal ID

**mid** Maternal ID

**sex** Sex (1=Male, 2=Female)

**sexf** Sex as a factor

**pop** Short population code

**reg** Short region code

**population** Long population description

**region** Long region description

### Source

<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/>

### References

Byrska-Bishop, Marta, et al. "High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios." *Cell* 185.18 (2022): 3426-3440.

1000 Genomes Project Consortium. "A global reference for human genetic variation." *Nature* 526.7571 (2015): 68.

License information is available at [https://github.com/igsr/1000Genomes\\_data\\_indexes/blob/master/LICENSE](https://github.com/igsr/1000Genomes_data_indexes/blob/master/LICENSE). The 1000 Genomes data is made publicly available according to the Fort Lauderdale Agreement (<https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>).

---

kgpe

*1000 Genomes Project sample data (Expanded)*

---

### Description

Sample, pedigree, and population data for 3,202 samples in the expanded 1000 Genomes Project data.

### Usage

kgpe

### Format

A tibble with 3202 rows and 11 columns:

**fid** Family ID

**id** Individual ID

**pid** Paternal ID

**mid** Maternal ID

**sex** Sex (1=Male, 2=Female)

**sexf** Sex as a factor

**pop** Short population code

**reg** Short region code

**population** Long population description

**region** Long region description

**phase3** Logical; indicates whether this sample is included in the Phase 3 release data

### Source

<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/>

### References

Byrska-Bishop, Marta, et al. "High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios." *Cell* 185.18 (2022): 3426-3440.

1000 Genomes Project Consortium. "A global reference for human genetic variation." *Nature* 526.7571 (2015): 68.

License information is available at [https://github.com/igsr/1000Genomes\\_data\\_indexes/blob/master/LICENSE](https://github.com/igsr/1000Genomes_data_indexes/blob/master/LICENSE). The 1000 Genomes data is made publicly available according to the Fort Lauderdale Agreement (<https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>).

---

kgpmeta

*1000 Genomes Project population metadata*

---

### Description

Population metadata from 26 populations across five continental regions.

### Usage

kgpmeta

### Format

A tibble with 26 rows and 7 columns:

**pop** Short population code

**reg** Short region code

**population** Long population description

**region** Long region description

**regcolor** Color for plotting this region on a map

**lat** Population latitude

**lng** Population longitude

### Source

<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/>

### References

Byrska-Bishop, Marta, et al. "High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios." *Cell* 185.18 (2022): 3426-3440.

1000 Genomes Project Consortium. "A global reference for human genetic variation." *Nature* 526.7571 (2015): 68.

License information is available at [https://github.com/igsr/1000Genomes\\_data\\_indexes/blob/master/LICENSE](https://github.com/igsr/1000Genomes_data_indexes/blob/master/LICENSE). The 1000 Genomes data is made publicly available according to the Fort Lauderdale Agreement (<https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>).

# Index

## \* datasets

allmeta, [2](#)

kgp3, [3](#)

kgpe, [4](#)

kgpmeta, [5](#)

allmeta, [2](#)

kgp3, [3](#)

kgpe, [4](#)

kgpmeta, [5](#)