Ferret: a user-friendly tool to extract data from the 1000 Genomes Project

R. Ba, A. Taverner, N. Vince, P.-A. Gourraud, C.-A. Winkler, M. Servières, Sophie Limou

▶ **To cite this version:**

R. Ba, A. Taverner, N. Vince, P.-A. Gourraud, C.-A. Winkler, et al.. Ferret: a user-friendly tool to extract data from the 1000 Genomes Project. Labex IGO meeting 2018, Apr 2018, Nantes, France. inserm-02161817

HAL Id: inserm-02161817
https://inserm.hal.science/inserm-02161817

Submitted on 21 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ferret: a user-friendly tool to extract data from the 1000 Genomes Project

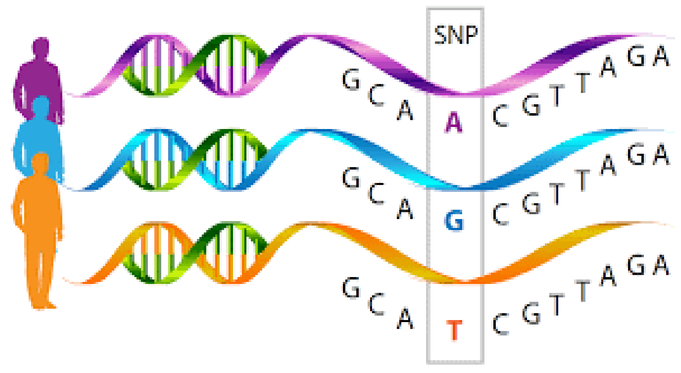
Contact:
sophie.limou@univ-nantes.fr
ferret@nih.gov

R. Ba[1], A. Taverner[2], N. Vince[1], PA. Gourraud[1],
CA. Winkler[2], M. Servières[3], S. Limou[1,2,3]

[1]ATIP-Avenir team, CRTI UMR 1064, INSERM, Nantes, France
[2] Basic Research Laboratory, NIH/NCI, Leidos Biomedical Research, Inc., Frederick MD, USA,
[3]Ecole Centrale de Nantes, Nantes, France

About the 1000 Genomes Project*

International consortium to establish the largest public catalogue of human genetic variations: >84M SNPs for 2504 individuals



HaploView files
data visualization, LD pattern, haplotypes, tagSNP design, etc.

PLINK files
can be merged with user's dataset, etc.

Allele frequency file

CHR	VARIANT	POS	A1	A2	NB_1KG_CHR	1KG_A1_FREQ	ESP6500_EA_A1_FREQ	ESP6500_AA_A1_FREQ
6	rs1065711	31236853	G	A	1006	0.1213	.	.
6	rs1049853	31236900	G	A	1006	0.8867	0.8726	0.9209
6	indel_rs41548117_TC/T	31237100	TC	T	1006	0.9553	0.9594	0.9183
6	rs41542414	31237766	A	T	1006	1	0.9995	0.9897
6	rs9264623	31237950	T	C	1006	0.2773	.	.
6	indel_rs9281300_C/CA	31239170	C	CA	1006	0.4195	0.535	0.4125

Unique advantages of Ferret#

- User-friendly interface
- Input flexibility
- Handles SNPs and indels
- Outputs suitable for pre-existing tools
- Computes allelic frequency
- Retrieves allelic frequency from ESP (Exome Sequencing Project)
- **Dec. 2017 update**: sync with HaploView + gene borders

Perspectives (Aug. 2018)

- Grant access to *HLA* alleles data§
- Functional annotations
- Bi-directional filter for frequencies
- Increase runtime speed

References

*The 1000 Genomes Project Nature 2015
#Limou *et al.* Bioinformatics 2016
§Gourraud *et al.* PLoS One 2014

Publicly available at http://limousophie35.github.io/Ferret/