B



Sep 21, 2020

nf-100GMX-variant-summarizer

Israel Aguilar Ordoñez¹

¹Instituto Nacional de Medicina Genómica (INMEGEN)

1 W

Works for me

dx.doi.org/10.17504/protocols.io.bkv6kw9e

Whole genome variation in 27 Mexican indigenous populations, demographic and biomedical insights



Judith Ballesteros Villascan

Centro de Investigación y de Estudios Avanzados del IPN (Cin...

ABSTRACT

Nextflow pipeline used to count variants for the 100GMX project

'nf-100GMX-variant-summarizer' is a pipeline tool that counts variants in a VEPextended annotated VCF file. This pipeline generates 3 outputs:

- 1) a TSV file with the total number of SNV and indels
- 2) a TSV file with per sample counts for variants of type SNV, indel, novel, worldwide singletons, clinvar, gwascat and pharmgkb
- 3) a PDF file with the number of discernible variants in sample groups of interest.

Important note: input file must be previously annotated by https://github.com/laguilaror/nf-VEPextended

EXTERNAL LINK

https://github.com/laguilaror/nf-100GMX-variant-summarizer

DOI

dx.doi.org/10.17504/protocols.io.bkv6kw9e

PROTOCOL CITATION

Israel Aguilar Ordoñez 2020. nf-100GMX-variant-summarizer. **protocols.io** https://dx.doi.org/10.17504/protocols.io.bkv6kw9e

EXTERNAL LINK

https://github.com/laguilaror/nf-100GMX-variant-summarizer

LICENSE

This is an open access protocol distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Sep 05, 2020

LAST MODIFIED

Sep 21, 2020

PROTOCOL INTEGER ID

41630

GUIDELINES

Installation

 $Download\ nf-100GMX-variant-summarizer\ from\ Github\ repository:$

git clone https://github.com/Iaguilaror/nf-100GMX-variant-summarizer

protocols.io
1
09/21/2020

 $\textbf{Citation:} \ \ \text{Israel Aguilar Ordo} \\ \tilde{\mathbb{A}} \\ \text{\pmez (09/21/2020)}. \ \ \text{nf-100GMX-variant-summarizer.} \\ \\ \underline{\text{$https://dx.doi.org/10.17504/protocols.io.bkv6kw9e}} \\ \text{$0.0000} \\ \text{$1.0000} \\ \text{$1.0000}$

Compatible OS*:

- <u>Ubuntu 18.04.03 LTS</u>
- <u>Ubuntu 16.04 LTS</u>
- * nf-100GMX-variant-summarizer may run in other UNIX based OS and versions, but testing is required.

Software Requirements:



MATERIALS TEXT

Pipeline Inputs

A compressed vcf file with extension '.vcf.gz'; the VCF must be previously annotated with https://github.com/laguilaror/nf-VEPextended

Example line(s):

AC=1;AF=0.00641;AN=140;DP=853;ANN=T|intron_variant|MODIFIER|GATD3B|ENSG00000280071|Transcript|ENST00000624810.3|protein_coding||4/5|ENST00000624810.3:c.357+19546C>A||||||rs1373489291||-

*.tsv: A metadata file, relating every sample ID (as registered in the VCF file) and a sample group in column format.

Example line(s):

sample group SM-3MG5L Chinanteco SM-3MG5F Chocholteco SM-3MG46 Kanjobal

BEFORE STARTING

Test

To test nf-100GMX-variant-summarizer's execution using test data, run:

./runtest.sh

Your console should print the Nextflow log for the run, once every process has been submitted, the following message will appear:

=====

VCF summarizer: Basic pipeline TEST SUCCESSFUL

nf-100GMX-variant-summarizer results for test data should be in the following file:

nf-100GMX-variant-summarizer/test/results/VCFsummarizer-results

Usage

To run nf-100GMX-variant-summarizer go to the pipeline directory and execute:

nextflow run summarize-vcf.nf --vcffile <path to input 1> --metadata <path to input 2>
--nsamples <integer> --group_minaf <numeric> --outgroup_maxaf <numeric> [--output_dir
path to results]

For information about options and parameters, run:

nextflow run summarize-vcf.nf --help

Branch A

1 Project Counts



- a) Count samples and raw stats for all samples.
- b) Give the total counts data.

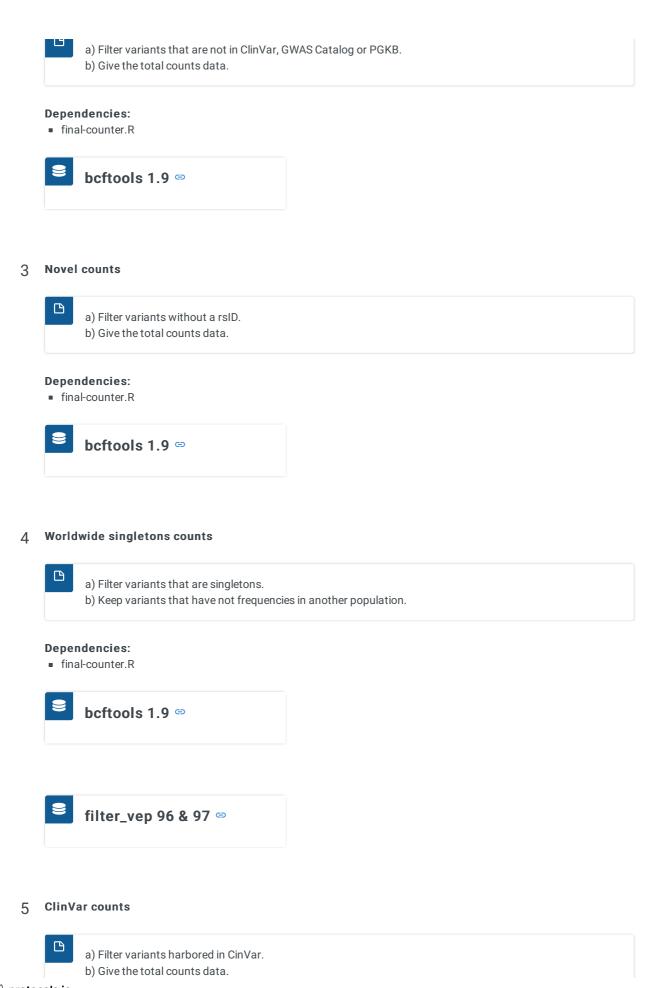
Branch B

2 No filter counts



09/21/2020

 $\textbf{Citation:} \ \ \text{Israel Aguilar Ordo} \\ \tilde{\mathbb{A}} \\ \text{\pmez (09/21/2020)}. \ \ \text{nf-100GMX-variant-summarizer.} \\ \\ \underline{\text{$https://dx.doi.org/10.17504/protocols.io.bkv6kw9e}} \\ \text{$0.0000} \\ \text{$1.0000} \\ \text{$1.0000}$



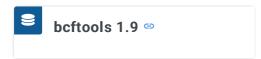
Dependencies: • final-counter.R filter_vep 96 & 97 © bcftools 1.9 ©

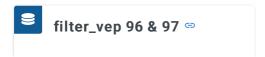
6 GWASCatalog counts



Dependencies:

• final-counter.R



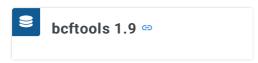


7 PGKB counts

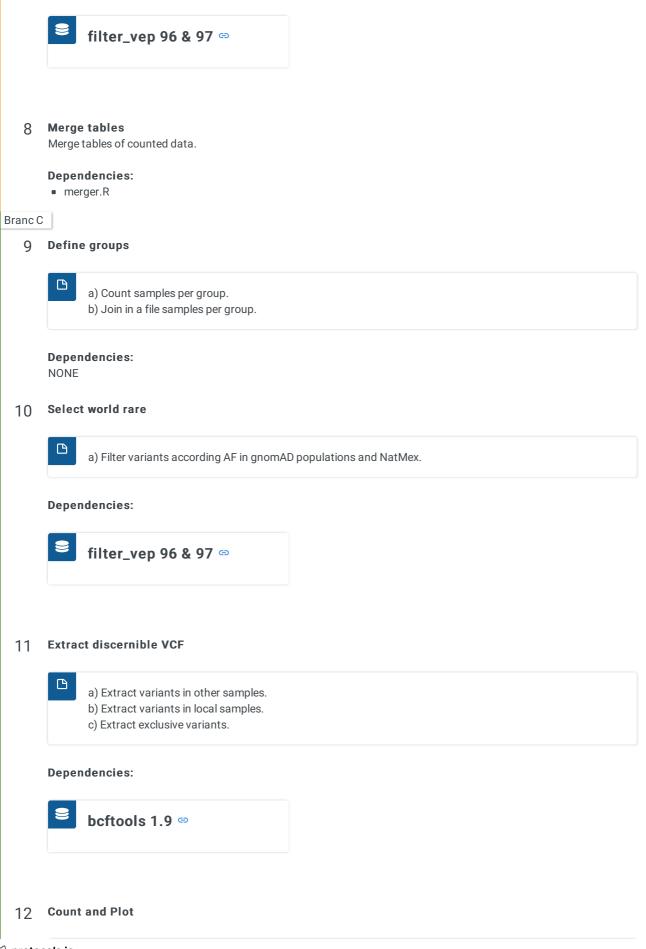


Dependencies:

• final-counter.R



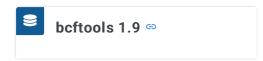
⋈ protocols.io 5 09/21/2020





- a) Count the number of discernible variants per group and type in a TSV file.
- b) Plot number of discernible variants per group and type.

Dependencies:



plotter.R