**Small tutorial of KGGSeq for annotation and prioritization of exome sequencing variants**

Miaoxin Li ( mxli@hku.hk)

**Reference**: <http://statgenpro.psychiatry.hku.hk/limx/kggseq/doc/UserManual.html>

|  |
| --- |
| **Input data:**   1. A Variant Call Format (VCF) file (a simulated data set)   examples/rare.disease.hg19.vcf   1. A linkage pedigree file:   examples/rare.disease.ped  **Purpose:** Identify sequence variant candidate that may cause Spinocerebellar ataxia  **Run the commands step by step to see what will happen**   1. Filter by genetic feature and inheritance model   kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --excel **--genotype-filter 3,5,6**  //when QC is imposed kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --excel --genotype-filter 3,5,6 **--seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8**   1. Annotate sequence variants by RefGenes:   kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --excel --genotype-filter 3,5,6 --seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8 **--db-gene refgene --gene-feature-in 0,1,2,3,4,5**   1. Filter sequence variants by Common variants   kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --excel --genotype-filter 3,5,6 --seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8 --db-gene refgene --gene-feature-in 0,1,2,3,4,5 **--db-filter hg19\_1kg201202,hg19\_ESP5400 --rare-allele-freq 0.01**   1. Prioritize sequence variants by disease-causing prediction   kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --excel --genotype-filter 3,5,6 --seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8 --db-gene refgene --gene-feature-in 0,1,2,3,4,5 --db-filter hg19\_1kg201202,hg19\_ESP5400 --rare-allele-freq 0.01 **--db-score dbnsfp --mendel-causing-predict all --filter-nondisease-variant**   1. Prioritize sequence variants by other genomic annotation   kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --excel --genotype-filter 3,5,6 --seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8 --db-gene refgene --gene-feature-in 0,1,2,3,4,5 --db-filter hg19\_1kg201202,hg19\_ESP5400 --rare-allele-freq 0.01 --db-score dbnsfp --mendel-causing-predict all **--filter-nondisease-variant --genome-annot**   1. Prioritize sequence variants by candidate genes with protein interaction information   kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --excel --genotype-filter 3,5,6 --seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8 --db-gene refgene --gene-feature-in 0,1,2,3,4,5 --db-filter hg19\_1kg201202,hg19\_ESP5400 --rare-allele-freq 0.01 --db-score dbnsfp --mendel-causing-predict all **--filter-nondisease-variant** --genome-annot **--candi-list HDAC2,HIC1 --ppi-annot string --ppi-depth 1**   1. Prioritize sequence variants by candidate genes with pathway information   kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --excel --genotype-filter 3,5,6 --seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8 --db-gene refgene --gene-feature-in 0,1,2,3,4,5 --db-filter hg19\_1kg201202,hg19\_ESP5400 --rare-allele-freq 0.01 --db-score dbnsfp --mendel-causing-predict all **--filter-nondisease-variant** --genome-annot --candi-list HDAC2,HIC1 --ppi-annot string --ppi-depth 1 **--pathway-annot cura**   1. Prioritize sequence variants by PubMed   kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --excel --genotype-filter 3,5,6 --seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8 --db-gene refgene --gene-feature-in 0,1,2,3,4,5 --db-filter hg19\_1kg201202,hg19\_ESP5400 --rare-allele-freq 0.01 --db-score dbnsfp --mendel-causing-predict all --filter-nondisease-variant --genome-annot --candi-list HDAC2,HIC1 --ppi-annot string --ppi-depth 1 --pathway-annot cura --pubmed-mining Spinocerebellar+ataxia |

**Others**

1. Output with plink binary files  
   kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8 --o-plink-bed
2. Output with ANNOVAR input files

kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8 --o-annovar

1. Output with VCF input files

kggseq --no-lib-check --no-resource-check --resource /opt/KGG/resources --vcf-file examples/rare.disease.hg19.vcf --ped-file examples/rare.disease.ped.txt --composite-subject-id --out test1 --seq-qual 50 --seq-mq 20 --gty-qual 20 --gty-dp 8 --o-vcf