
uga Documentation

Release 2014.01

Ryan Koesterer

November 28, 2014

CONTENTS

1	Modules	3
1.1	Analysis	3
1.2	Meta-Analysis	4
1.3	Annotation	5
1.4	Plots	6
2	Index and Search	7

Universal Genome Analyst (uga) is a pipeline for the distribution and management of large and small scale analysis of genetic data.

MODULES

1.1 Analysis

```
usage: uga analyze [-h] [--version] [-o] [-q QSUB] [--name NAME]
                  [-d DIRECTORY] [--cpus CPUS] [--mem MEM]
                  [-r REGION | --region-list REGION_LIST]
                  [-s | -n SPLIT_N | --chr CHR | --split-chr]
                  [-j JOB | --job-list JOB_LIST] --data DATA --out OUT
                  --samples SAMPLES --pheno PHENO --model MODEL --fid FID
                  --iid IID --method
                  {gee_gaussian,gee_binomial,glm_gaussian,glm_binomial,lme_gaussian,lme_b
                  [--focus FOCUS] [--sig SIG] [--sex SEX] [--male MALE]
                  [--female FEMALE] [--buffer BUFFER] [--miss MISS]
                  [--freq FREQ] [--rsq RSQ] [--hwe HWE] [--case CASE]
                  [--ctrl CTRL] [--nofail]
```

Options:

--version	show program's version number and exit
-o=False, --overwrite=False	overwrite existing output files
-q, --qsub	a group ID under which to submit jobs to the queue
--name	a job name (only used with -qsub; if not set, -out basename will be used
-d=/usr3/bustaff/koesterr/UGA/doc, --directory=/usr3/bustaff/koesterr/UGA/doc	an output directory path
--cpus=1	number of cpus (limited module availability)
--mem=3	amount of ram memory to request for queued job (in gigabytes)
-r, --region	a region specified in tabix format (ie. 1:10583-1010582).
--region-list	a filename for a list of tabix format regions
-s=False, --split=False	split region list entirely into separate jobs (requires -region-list)
-n, --split-n	split region list into n separate jobs (requires -region-list)
--chr	chromosome number (1-26)
--split-chr=False	split by chromosome (will generate up to 26 separate jobs depending on chromosome coverage)
-j, --job	run a particular job number (requires -split-n)
--job-list	a filename for a list of job numbers (requires -split-n)

--data	a genomic data file
--out	an output file name (basename only: do not include path)
--samples	a sample file (single column list of IDs in order of data)
--pheno	a tab delimited phenotype file
--model	a comma separated list of models in the format "pheno-type~age+factor(sex)+pc1+pc2+pc3+marker"
--fid	the column name with family ID
--iid	the column name with sample ID
--method	the analysis method Possible choices: gee_gaussian, gee_binomial, glm_gaussian, glm_binomial, lme_gaussian, lme_binomial, coxph, efftests
--focus	a comma separated list of variables for which stats will be output (default: marker)
--sig=5	significant digits to include in output (default: 5)
--sex	name of the column containing male/female status (requires MALE_CODE and female)
--male=1	the code for a male in sex (requires --sex and --female)
--female=2	the code for a male in sex (requires --sex and --female)
--buffer=100	a value for number of markers calculated at a time (WARNING: this argument will affect RAM memory usage; default: 100)
--miss	a threshold value for missingness
--freq	a threshold value for allele frequency
--rsq	a threshold value for r-squared (imputation quality)
--hwe	a threshold value for Hardy Weinberg p-value
--case=1	the code for a case in the dependent variable column (requires ctrl; binomial fxn family only; default: 1)
--ctrl=0	the code for a control in the dependent variable column (requires case; binomial fxn family only; default: 0)
--nofail=False	exclude filtered/failed analyses from results

1.2 Meta-Analysis

```
usage: uga meta [-h] [--version] [-o] [-q QSUB] [--name NAME] [-d DIRECTORY]
               [--cpus CPUS] [--mem MEM]
               [-r REGION | --region-list REGION_LIST]
               [-s | -n SPLIT_N | --chr CHR | --split-chr]
               [-j JOB | --job-list JOB_LIST] -c CFG [-v VARS]
```

Options:

--version	show program's version number and exit
-o=False, --overwrite=False	overwrite existing output files

-q, --qsub	a group ID under which to submit jobs to the queue
--name	a job name (only used with <code>-qsub</code> ; if not set, <code>-out</code> basename will be used
-d=/usr3/bustaff/koesterr/UGA/doc, --directory=/usr3/bustaff/koesterr/UGA/doc	an output directory path
--cpus=1	number of cpus (limited module availability)
--mem=3	amount of ram memory to request for queued job (in gigabytes)
-r, --region	a region specified in tabix format (ie. 1:10583-1010582).
--region-list	a filename for a list of tabix format regions
-s=False, --split=False	split region list entirely into separate jobs (requires <code>-region-list</code>)
-n, --split-n	split region list into n separate jobs (requires <code>-region-list</code>)
--chr	chromosome number (1-26)
--split-chr=False	split by chromosome (will generate up to 26 separate jobs depending on chromosome coverage)
-j, --job	run a particular job number (requires <code>-split-n</code>)
--job-list	a filename for a list of job numbers (requires <code>-split-n</code>)
-c, --cfg	a configuration file name
-v, --vars	a declaration of the form <code>A=B, C=D, E=F, ...</code> to replace <code>[A]</code> with <code>B</code> , <code>[C]</code> with <code>D</code> , <code>[E]</code> with <code>F</code> , ... in any line of the <code>cfg</code> file

1.3 Annotation

```
usage: uga annot [-h] [--version] [-o] [-q QSUB] [--name NAME] [-d DIRECTORY]
               [--cpus CPUS] [--mem MEM]
               [-r REGION | --region-list REGION_LIST]
               [-s | -n SPLIT_N | --chr CHR | --split-chr]
               [-j JOB | --job-list JOB_LIST]
```

Options:

--version	show program's version number and exit
-o=False, --overwrite=False	overwrite existing output files
-q, --qsub	a group ID under which to submit jobs to the queue
--name	a job name (only used with <code>-qsub</code> ; if not set, <code>-out</code> basename will be used
-d=/usr3/bustaff/koesterr/UGA/doc, --directory=/usr3/bustaff/koesterr/UGA/doc	an output directory path
--cpus=1	number of cpus (limited module availability)
--mem=3	amount of ram memory to request for queued job (in gigabytes)
-r, --region	a region specified in tabix format (ie. 1:10583-1010582).
--region-list	a filename for a list of tabix format regions
-s=False, --split=False	split region list entirely into separate jobs (requires <code>-region-list</code>)
-n, --split-n	split region list into n separate jobs (requires <code>-region-list</code>)

--chr	chromosome number (1-26)
--split-chr=False	split by chromosome (will generate up to 26 separate jobs depending on chromosome coverage)
-j, --job	run a particular job number (requires <code>--split-n</code>)
--job-list	a filename for a list of job numbers (requires <code>--split-n</code>)

1.4 Plots

```
usage: uga plot [-h] [--version] [-o] [-q QSUB] [--name NAME] [-d DIRECTORY]
               [--cpus CPUS] [--mem MEM]
               [-r REGION | --region-list REGION_LIST]
               [-s | -n SPLIT_N | --chr CHR | --split-chr]
               [-j JOB | --job-list JOB_LIST]
```

Options:

--version	show program's version number and exit
-o=False, --overwrite=False	overwrite existing output files
-q, --qsub	a group ID under which to submit jobs to the queue
--name	a job name (only used with <code>-qsub</code> ; if not set, <code>-out</code> basename will be used)
-d=/usr3/bustaff/koesterr/UGA/doc, --directory=/usr3/bustaff/koesterr/UGA/doc	an output directory path
--cpus=1	number of cpus (limited module availability)
--mem=3	amount of ram memory to request for queued job (in gigabytes)
-r, --region	a region specified in tabix format (ie. 1:10583-1010582).
--region-list	a filename for a list of tabix format regions
-s=False, --split=False	split region list entirely into separate jobs (requires <code>--region-list</code>)
-n, --split-n	split region list into n separate jobs (requires <code>--region-list</code>)
--chr	chromosome number (1-26)
--split-chr=False	split by chromosome (will generate up to 26 separate jobs depending on chromosome coverage)
-j, --job	run a particular job number (requires <code>--split-n</code>)
--job-list	a filename for a list of job numbers (requires <code>--split-n</code>)

INDEX AND SEARCH

- *genindex*
- *search*