QCall Manual

./QCALL options

-i data file| stdin

-snpcan: filter by SNPCan

-ct < low candidate threshold | 0.001>

-co <file output for SNP candidates>

-vcf <vcf snp lists

-sn <file of sample names> : output genotypes of samples given in the file

-pphet <prior probability of heterozygous | .01>

-ARG: <tree> <samples> with Margarita tree format

-farg: <tree> <samples> with fast Arg tree format

-pho: output files of phased haplotypes using ARG trees

-onego: genotyping/phasing with ARGs given as input. No new ARGs is built

Note: the current version supports only for ***onego*** option.

Options in detail

-i data file | stdin

SNP detection from BAM files

The plain text log likelihood in phred score from Samtool with the sample names attached at the end of each line. Input must be sorted with the <chr, pos> order. The best way to obtain data from Samtool is to extract the log likelihood from BAM files as:

```
(
./samtools view BAMfile_1 chr:pos1-pos2 | ./samtools pileup -g –st reference.fa.fai reference.fa -
| ./samtools glfview – | awk '{ printf "%s\t%s\n", $0, bamfile_1}';
./samtools view BAMfile_2 chr:pos1-pos2 | ./samtools pileup -g –st reference.fa.fai reference.fa -
| ./samtools glfview – | awk '{ printf "%s\t%s\n", $0, bamfile_2}';


….
./samtools view BAMfile_n chr:pos1-pos2 | ./samtools pileup -g –st reference.fa.fai reference.fa -
| ./samtools glfview – | awk '{ printf "%s\t%s\n", $0, bamfile_n}';

) | sort -k1,1n -k2,2n
```

The output looks like

chr	pos	ref	c1	c2	c3	AA	AC	AG	AT	CC	CT	CG	GG	GT	TT	Sample name
1	49999953	G	1	60	0	24	24	3	24	24	3	24	0	3	24	NA18961
1	49999953	G	2	60	0	53	53	6	53	53	6	53	0	6	53	NA18953
1	49999953	G	4	60	0	115	115	12	115	115	12	115	0	12	115	NA18970
1	49999954	T	1	60	0	21	21	21	3	21	21	3	21	3	0	NA18959
1	49999954	T	1	60	0	29	29	29	3	29	29	3	29	3	0	NA18964
1	49999954	T	1	60	0	31	31	31	3	31	31	3	31	3	0	NA18961
1	49999954	T	2	60	0	57	57	57	6	57	57	6	57	6	0	NA18953
1	49999954	T	4	60	0	112	112	112	12	112	112	12	112	12	0	NA18970
1	49999955	G	1	60	0	20	20	3	20	20	3	20	0	3	20	NA18959
1	49999955	G	1	60	0	26	26	3	26	26	3	26	0	3	26	NA18964
1	49999955	G	1	60	0	28	28	3	28	28	3	28	0	3	28	NA18961
1	49999955	G	2	60	0	47	47	6	47	47	6	47	0	6	47	NA18953
1	49999955	G	4	60	0	118	118	12	118	118	12	118	0	12	118	NA18970
1	49999956	A	1	60	0	0	3	3	3	16	16	16	16	16	16	NA18959
1	49999956	A	1	60	0	0	3	3	3	26	26	26	26	26	26	NA18964
1	49999956	A	1	60	0	0	3	3	3	34	34	34	34	34	34	NA18961
1	49999956	A	2	60	0	0	6	6	6	60	60	60	60	60	60	NA18953
1	49999956	A	4	60	0	0	12	12	12	116	116	116	116	116	116	NA18970
1	49999957	T	1	60	0	21	21	21	3	21	21	3	21	3	0	NA18959
1	49999957	T	1	60	0	25	25	25	3	25	25	3	25	3	0	NA18964

1	49999957	T	1	60	0	28	28	28	3	28	28	3	28	3	0	NA18961
1	49999957	T	1	60	0	29	29	29	3	29	29	3	29	3	0	NA18571
1	49999957	T	1	60	0	30	30	30	3	30	30	3	30	3	0	NA18547
1	49999957	T	3	60	0	96	96	96	9	96	96	9	96	9	0	NA18953
1	49999957	T	4	60	0	105	105	105	12	105	105	12	105	12	0	NA18970
1	49999958	A	1	60	0	0	3	3	3	24	24	24	24	24	24	NA18571
1	49999958	A	1	60	0	0	3	3	3	25	25	25	25	25	25	NA18547
1	49999958	A	1	60	0	0	3	3	3	26	26	26	26	26	26	NA18959
1	49999958	A	1	60	0	0	3	3	3	26	26	26	26	26	26	NA18964
1	49999958	A	1	60	0	0	3	3	3	30	30	30	30	30	30	NA18576
1	49999958	A	1	60	0	0	3	3	3	31	31	31	31	31	31	NA18563
1	49999958	A	1	60	0	0	3	3	3	31	31	31	31	31	31	NA18573
1	49999958	A	1	60	0	0	3	3	3	34	34	34	34	34	34	NA18961
1	49999958	A	3	60	0	0	9	9	9	81	81	81	81	81	81	NA18953

For indel:

We could use QCALL for indel by treating the log likelihood of hom/hom, hom/het and het/het indels as log likelihood of ref/ref, ref/alt, and alt/alt. For example, if we have log hom/hom, hom/het, and het/het: 15, 0, 50 at the site (1, 49999958) with reference A, then we could make the input data with alt = C and phred score of log 0 = 93:

chr	pos	ref				AA	AC	AG	AT	CC	CT	CG	GG	GT	TT	Sample names
1	49999958	A	1	60	0	15	0	93	93	50	93	93	93	93	93	NA18571

-snpcan

Filter sites by using non linkage analysis. It fasts and filter about 90 homozygous sites with threshold = 0.001.

-ct < low candidate threshold | 0.001>

-co <file output for SNP candidates>

Output file (VCF format) for SNP candidates using non linkage analysis.

-pphet <prior probability of heterozygous | .01>

-ARG: <tree> <samples> with Margarita tree format

We need trees built from HapMap3 by Margarita for Linkage Disequilibrium Analysis.

-pho: output files of phased haplotypes using ARG trees

-onego: genotyping/phasing with ARGs given as input. No new ARGs is built