

Supplementary Data for SNVSniffer

Yongchao Liu, Martin Loewer, Srinivas Aluru, Bertil Schmidt

1. Parameters for Alignment and Variant Calling

1.1 Sequence alignment

```
bwa aln -t 4 reference.fa normal_1.fa.gz >tmp1.sai
bwa aln -t 4 reference.fa normal_2.fq.gz >tmp2.sai
bwa sampe -r '@RG\tID:gatech\tSM:illumina' -a 700 reference.fa tmp1.sai tmp2.sai reads_1.fq.gz
reads_2.fq.gz | samtools view -bhS - >reads.bam
```

1.2 Generation of pileup files

```
samtools sort reads.bam reads_sorted
samtools index reads_sorted.bam
samtools mpileup -s -C 50 -f reference.fa reads_sorted.bam >reads.pileup
```

1.3 SNP Calling

1.3.1 Command Lines for the Venter Dataset

```
SNVSniffer  SNVSniffer snp in.pileup
SAMtools    samtools mpileup -C50 -u -f reference.fa in.bam | bcftools view -bvcg - > out.var.raw.bcf
            bcftools out.var.raw.bcf" | vcfutils.pl varFilter -D100 > out.var.raw.vcf"
GATK        java -jar GenomeAnalysisTK.jar -T HaplotypeCaller -R reference.fa -I in.bam --
            genotyping_mode DISCOVERY -stand_emit_conf 10 -stand_call_conf 30 -rf BadCigar -o
            out
FaSD        FaSD.sh -d coverage in.pileup -o out
```

1.3.2 Command Lines for Contaminated Venter Dataset

```
SNVSniffer  SNVSniffer snp in.pileup --relaxed 0
SAMtools    samtools mpileup -C50 -u -f reference.fa in.bam | bcftools view -bvcg - > out.var.raw.bcf
```

```

bcftools out.var.raw.bcf' | vcfutils.pl varFilter -D100 > out.var.raw.vcf
GATK      java -jar GenomeAnalysisTK.jar -T HaplotypeCaller -R reference.fa -I in.bam --
          genotyping_mode DISCOVERY -stand_emit_conf 10 -stand_call_conf 30 -rf BadCigar -o
          out
FaSD      FaSD.sh -d coverage in.pileup -o out

```

1.4 Somatic Variant Calling

1.4.1 Command Lines for Synthetic Tumors

```

SNVSniffer  SNVSniffer somatic normal_header.sam tumor_header.sam normal.pileup tumor.pileup
            -o out.vcf
VarScan2    java -Xmx4g -jar VarScan.jar somatic normal.pileup tumor.pileup out.vcf -output-vcf 1
            --strand-filter --tumor-purity $purity --min-coverage-normal 3 --min-coverage-tumor 3
SomaticSniper bam-somaticsniper -F vcf -f reference.fa tumor.sorted.bam normal.sorted.bam out.vcf
JointSNVMix2 jsm.py classify --min_base_qual 0 --min_map_qual 20 --model snvmix2 --
            somatic_threshold 0.05 --out_file out.txt reference.fa normal.sorted.bam
            tumor.sorted.bam
MuTect      java -Xmx16g -jar muTect.jar --analysis_type MuTect --reference_sequence
            reference.fa --input_file:normal normal.sorted.bam --input_file:tumor tumor.sorted.bam
            --vcf out.vcf

```

1.4.2 Command Lines for Real Tumors

```

SNVSniffer  SNVSniffer somatic normal_header.sam tumor_header.sam normal.pileup tumor.pileup
            -o out.vcf
VarScan2    java -Xmx4g -jar VarScan.jar somatic normal.pileup tumor.pileup out.vcf -output-vcf 1
            --strand-filter --min-coverage-normal 3 --min-coverage-tumor 3
SomaticSniper bam-somaticsniper -F vcf -f reference.fa tumor.sorted.bam normal.sorted.bam out.vcf
JointSNVMix2 jsm.py classify --min_base_qual 0 --min_map_qual 20 --model snvmix2 --
            somatic_threshold 0.05 --out_file out.txt reference.fa normal.sorted.bam
            tumor.sorted.bam
MuTect      java -Xmx16g -jar muTect.jar --analysis_type MuTect --reference_sequence
            reference.fa --input_file:normal normal.sorted.bam --input_file:tumor tumor.sorted.bam
            --vcf out.vcf

```

2.1 Evaluation Results on Real Tumors

2.1.1 TCGA-13-0885-01A-02W-0421-09

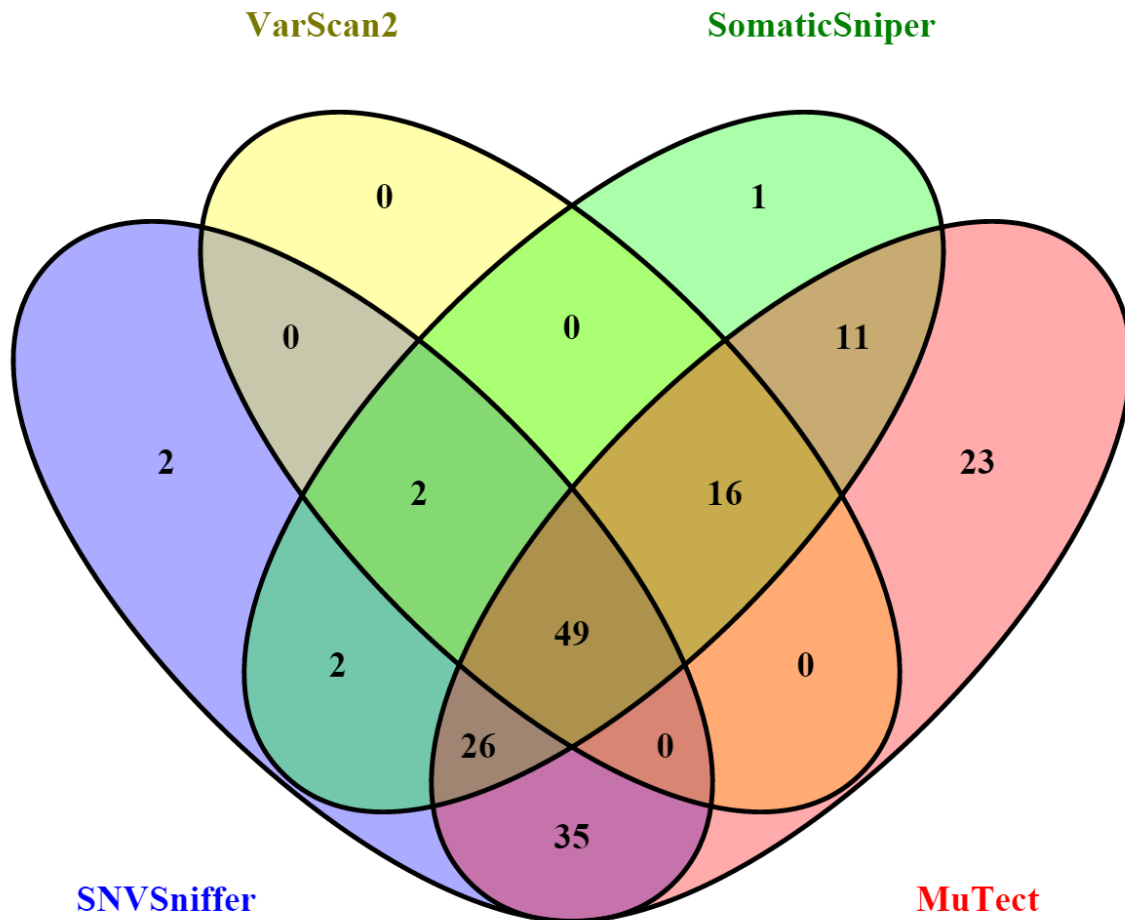


Figure S1. Somatic variant concordance on the sample TCGA-13-0885-01A-02W-0421-09

2.1.2 TCGA-13-1481-01A-01W-0549-09

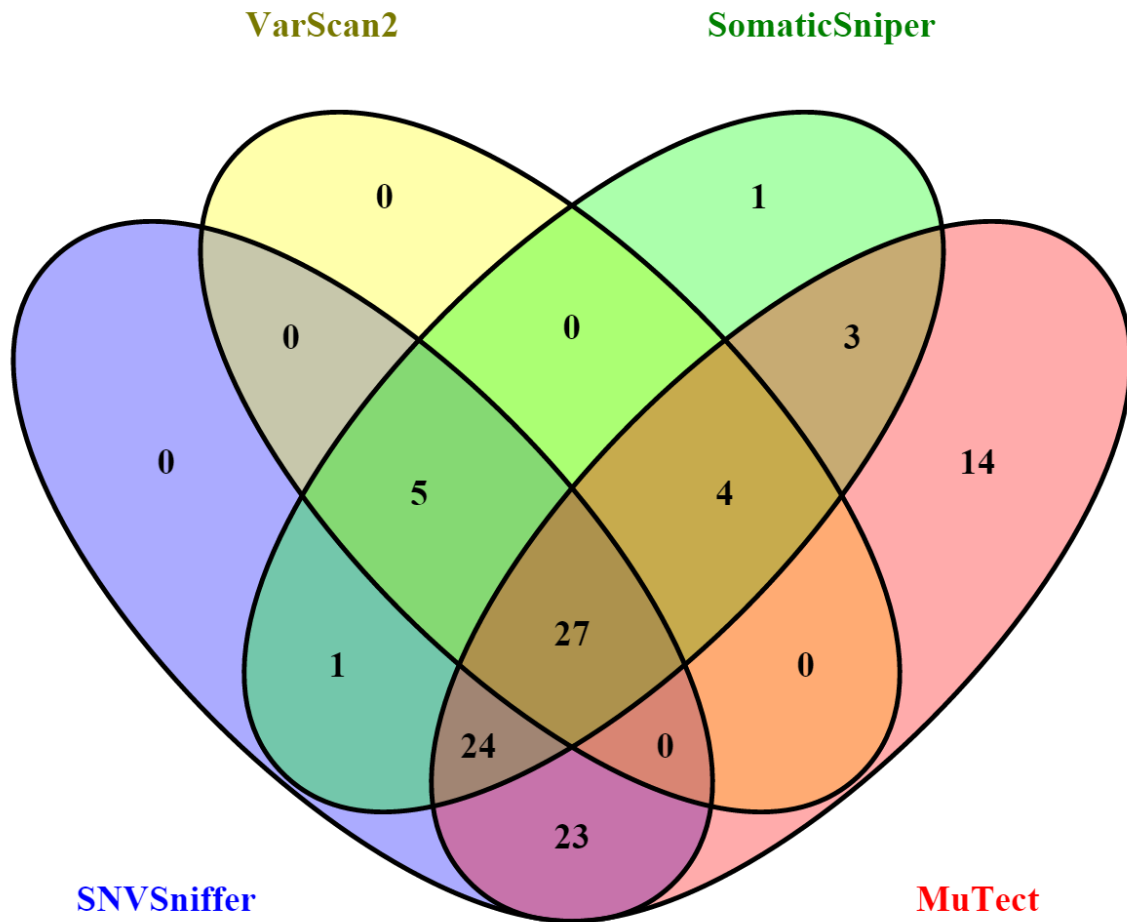


Figure S2. Somatic variant concordance on the sample TCGA-13-1481-01A-01W-0549-09

2.1.3 TCGA-13-1488-01A-01W-0549-09

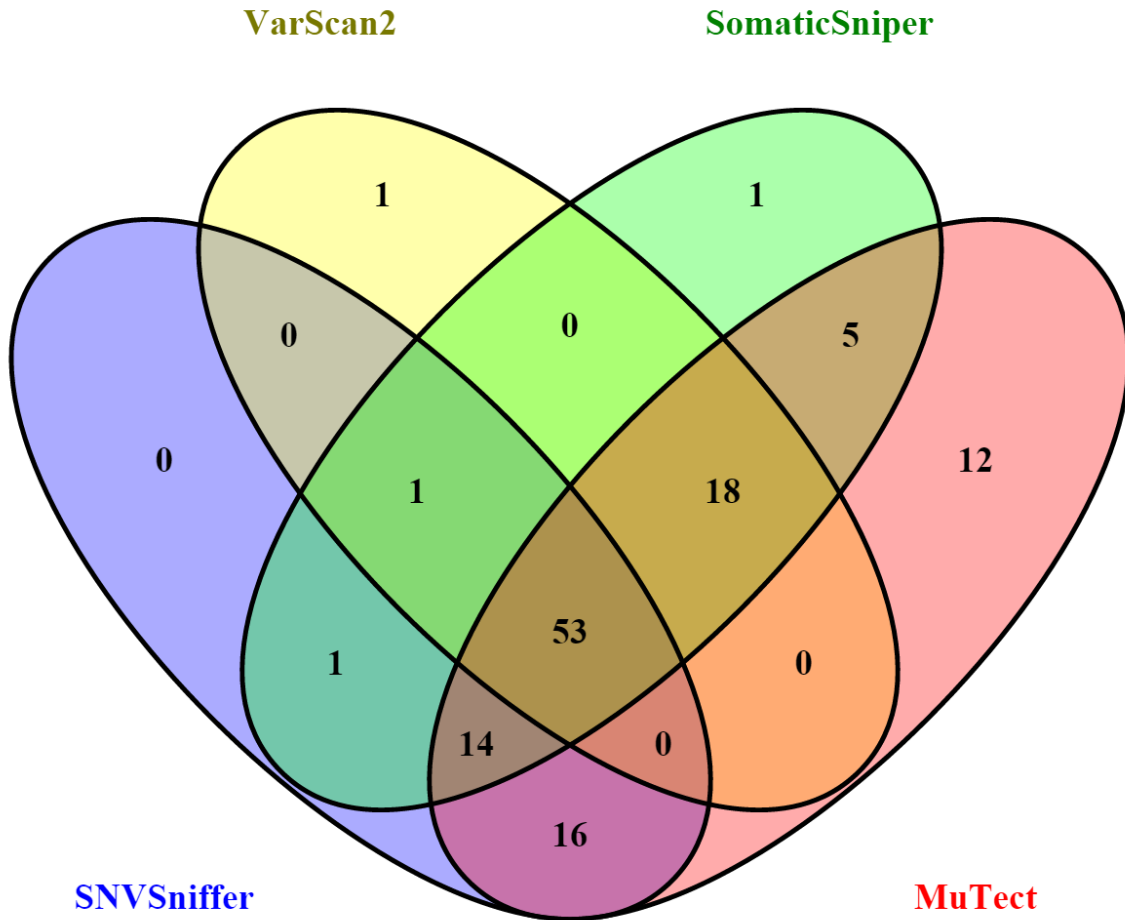


Figure S3. Somatic variant concordance on the sample TCGA-13-1488-01A-01W-0549-09

2.1.4 TCGA-24-1417-01A-01W-0549-09

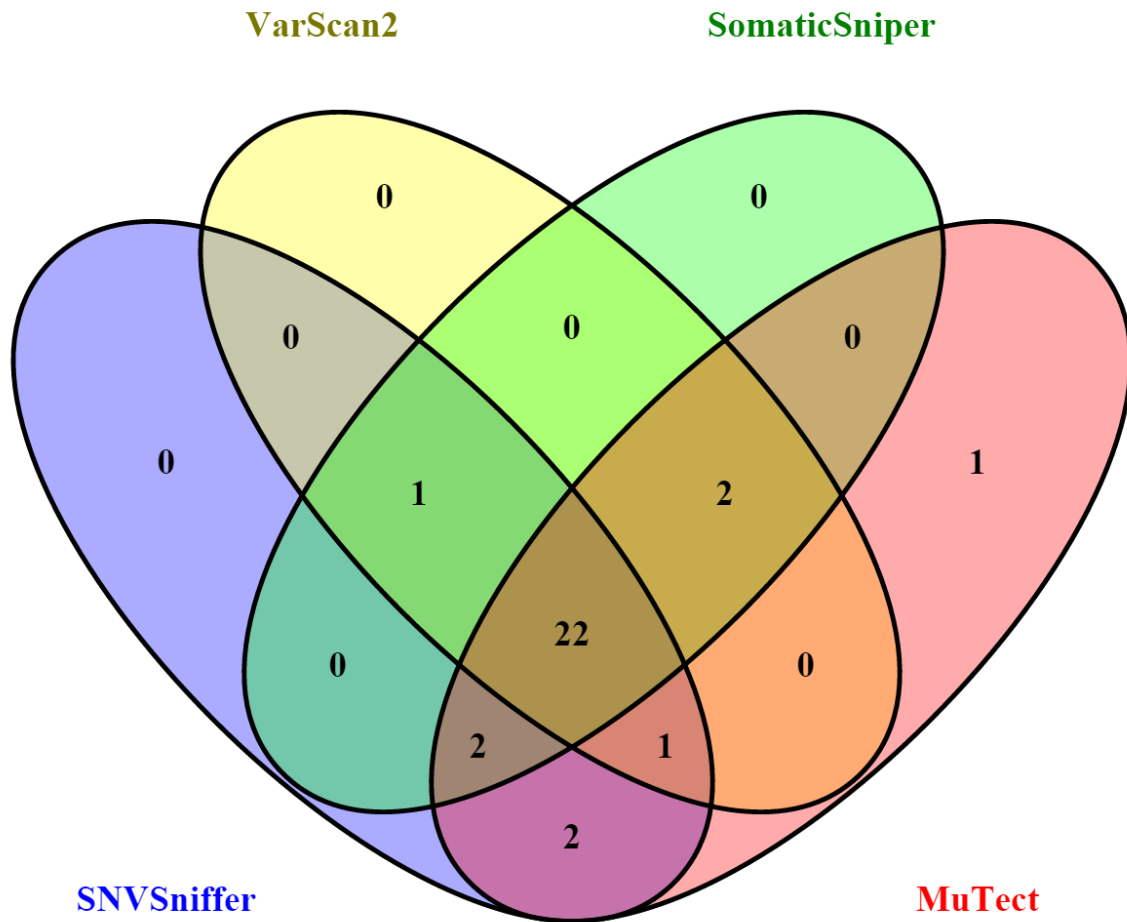


Figure S4. Somatic variant concordance on the sample TCGA-24-1417-01A-01W-0549-09

2.1.5 TCGA-24-1424-01A-01W-0549-09

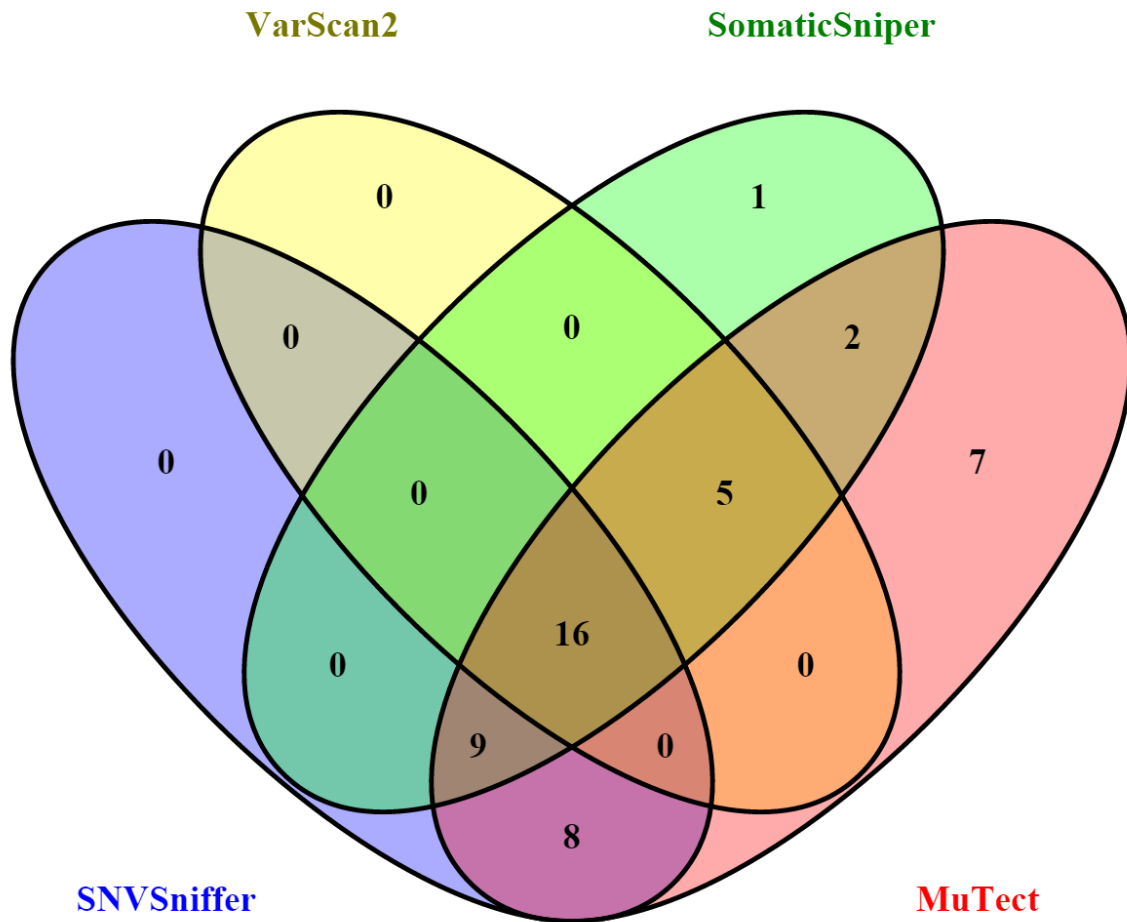


Figure S5. Somatic variant concordance on the sample TCGA-24-1424-01A-01W-0549-09