

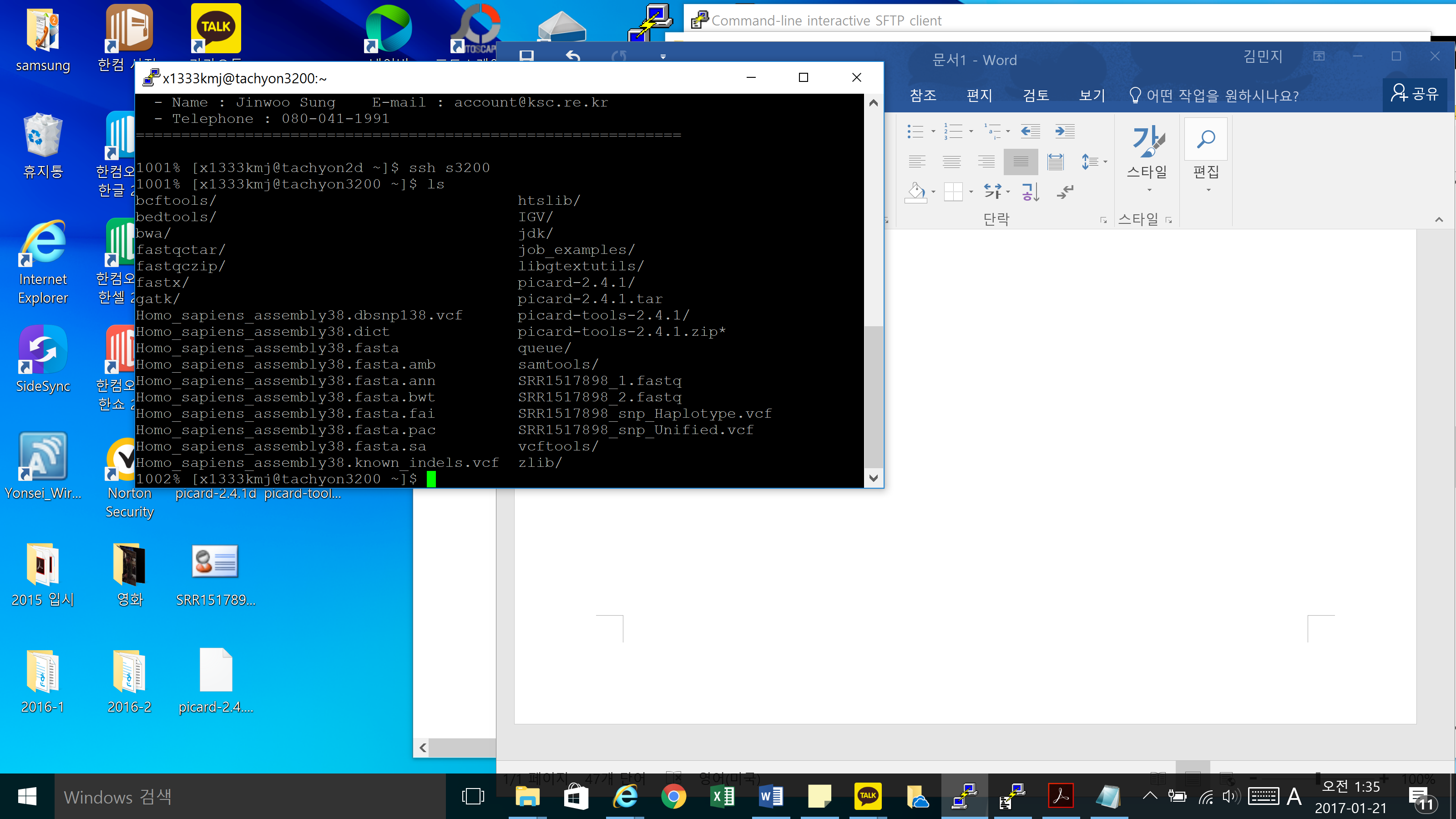
Bwa, Samtools 사용한 파일들 모두 일치

1111% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -jar ~/picard2/picard-2.4.1/picard.jar CreateSequenceDictionary R=Homo\_sapiens\_assembly38.fasta O=Homo\_sapiens\_assembly38.dict

-rw-r--r-- 1 x1333kmj pc1081 436974 Jan 21 00:54 Homo\_sapiens\_assembly38.dict

1116% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -jar ~/picard/picard-2.8.1/picard.jar CreateSequenceDictionary R=Homo\_sapiens\_assembly38.fasta O=Homo\_sapiens\_assembly38.dict

-rw-r--r-- 1 x1333kmj pc1081 436962 Jan 21 00:57 Homo\_sapiens\_assembly38.dict

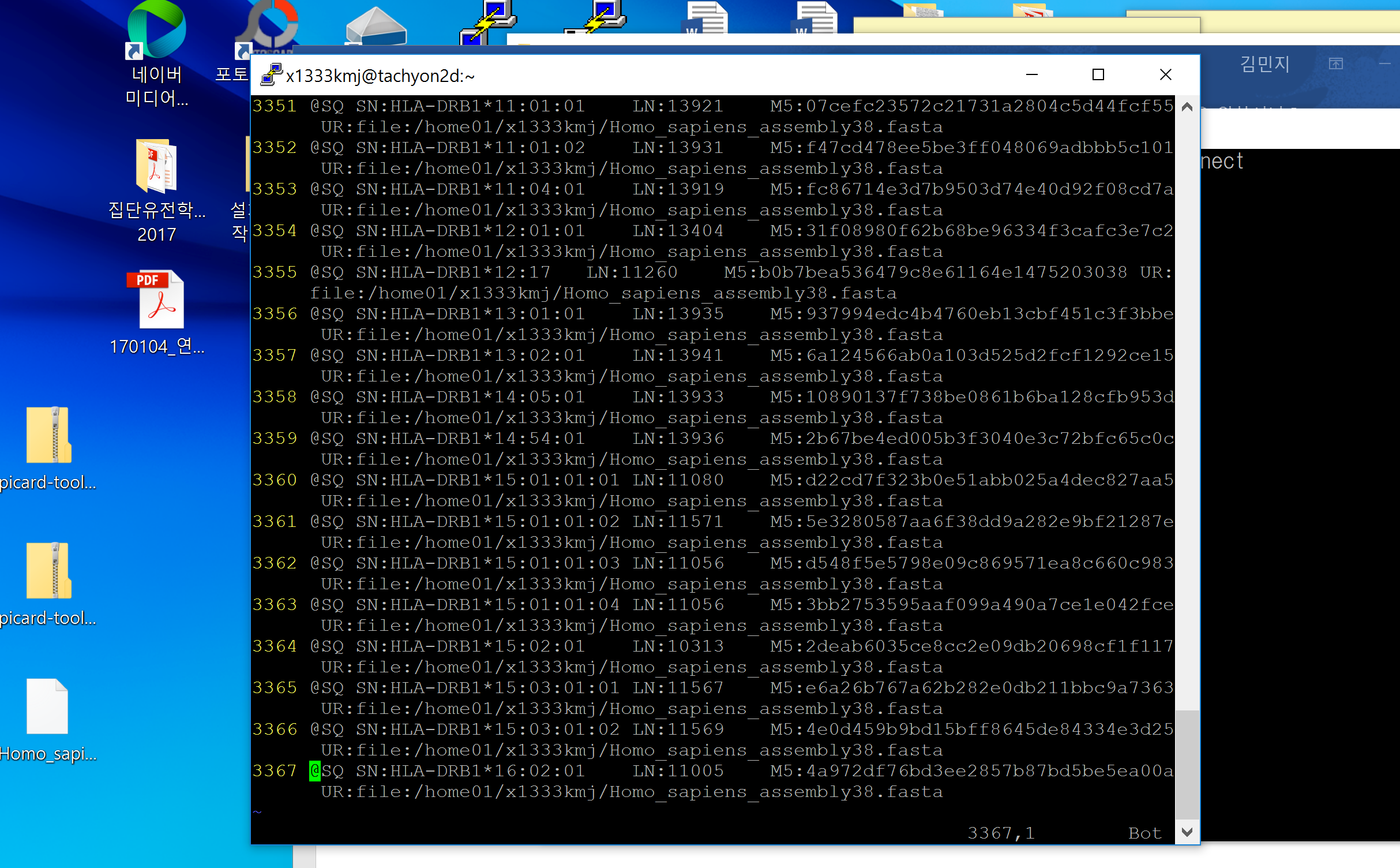


Picard 다시 install

1004% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -jar /home01/x1333kmj/picard-tools-2.4.1/picard.jar CreateSequenceDictionary R=Homo\_sapiens\_assembly38.fasta O=Homo\_sapiens\_assembly38.dict

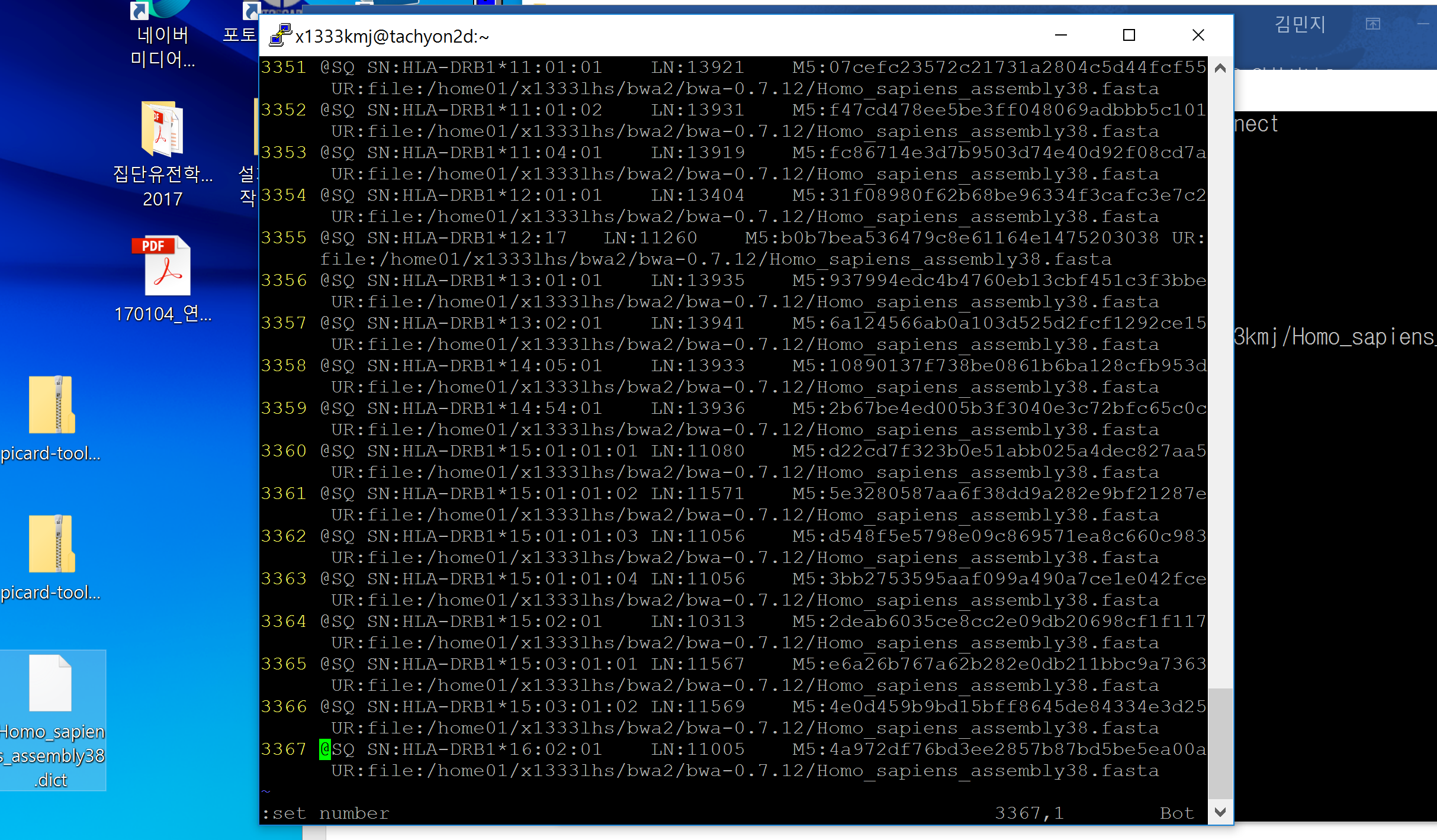
-rw-r--r-- 1 x1333kmj pc1081 436974 Jan 21 01:37 Homo\_sapiens\_assembly38.dict

여전히 용량 다름



내가 생성한 것 (one that I made)

이형성 선배님이 생성한 것 (made by the other student who got the final vcf file right)



Same line number

홍영기 선배님에 따르면 이형성 선배님의 .dict파일에는 경로가 추가적으로 붙어서 나타나는 현상임. 파일에는 문제 없는 것으로 보인다고 하심

1007% [x1333kmj@tachyon3200 ~]$ bwa/bwa-0.7.15/bwa mem -M -t 8 Homo\_sapiens\_assembly38.fasta SRR1517898\_1.fastq SRR1517898\_2.fastq > SRR1517898.sam

1011% [x1333kmj@tachyon3200 ~]$ samtools/samtools-1.3.1/samtools view -bhS SRR1517898.sam > SRR1517898.bam

1014% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -Xmx3g -Djava.io.tmpdir=/scratch2/x1333kmj/x1333kmj/tmp1 -jar ~/picard-tools-2.4.1/picard.jar AddOrReplaceReadGroups I=SRR1517898.bam O=SRR1517898\_addrg.bam SORT\_ORDER=coordinate RGID=SRR1517898ID RGLB=SRR1517898 RGPL=illumina RGPU=none RGSM=SRR1517898SM

1016% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -Xmx3g -Djava.io.tmpdir=/scratch2/x1333kmj/x1333kmj/tmp1 -jar ~/picard-tools-2.4.1/picard.jar FixMateInformation I=SRR1517898\_addrg.bam O=SRR1517898\_addrg\_fixm.bam SORT\_ORDER=coordinate

1018% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -Xmx3g -Djava.io.tmpdir=/scratch2/x1333kmj/x1333kmj/tmp1 -jar ~/picard-tools-2.4.1/picard.jar SortSam I=SRR1517898\_addrg\_fixm.bam O=SRR1517898\_addrg\_fixm\_sorted.bam SORT\_ORDER=coordinate

1021% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -Xmx3g -jar ~/picard-tools-2.4.1/picard.jar MarkDuplicates I=SRR1517898\_addrg\_fixm\_sorted.bam O=SRR1517898\_addrg\_fixm\_sorted\_md.bam METRICS\_FILE=SRR1517898.dub REMOVE\_DUPLICATES=true ASSUME\_SORTED=true

1023% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -Xmx3g -jar ~/picard-tools-2.4.1/picard.jar BuildBamIndex I=SRR1517898\_addrg\_fixm\_sorted\_md.bam

1025% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -Xmx3g -jar ~/gatk/GenomeAnalysisTK.jar -T RealignerTargetCreator -R Homo\_sapiens\_assembly38.fasta -I SRR1517898\_addrg\_fixm\_sorted\_md.bam --known Homo\_sapiens\_assembly38.known\_indels.vcf -o SRR1517898\_addrg\_fixm\_md\_IndelRealigner.intervals

1028% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -Xmx3g -Djava.io.tmpdir=/scratch2/x1333kmj/x1333kmj/tmp1 -jar ~/gatk/GenomeAnalysisTK.jar -T IndelRealigner -R Homo\_sapiens\_assembly38.fasta -I SRR1517898\_addrg\_fixm\_sorted\_md.bam -known Homo\_sapiens\_assembly38.known\_indels.vcf -targetIntervals SRR1517898\_addrg\_fixm\_md\_IndelRealigner.intervals -o SRR1517898\_addrg\_fixm\_sorted\_md\_realigned.bam

1030% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -jar ~/gatk/GenomeAnalysisTK.jar -T BaseRecalibrator -R Homo\_sapiens\_assembly38.fasta -I SRR1517898\_addrg\_fixm\_sorted\_md\_realigned.bam -knownSites Homo\_sapiens\_assembly38.dbsnp138.vcf -o SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.table

1032% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -jar ~/gatk/GenomeAnalysisTK.jar -T PrintReads -R Homo\_sapiens\_assembly38.fasta -I SRR1517898\_addrg\_fixm\_sorted\_md\_realigned.bam -BQSR SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.table -o SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.bam

1037% [x1333kmj@tachyon3200 ~]$ jdk/jdk1.8.0\_102/bin/java -jar ~/gatk/GenomeAnalysisTK.jar -T UnifiedGenotyper -R Homo\_sapiens\_assembly38.fasta -I SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.bam --dbsnp Homo\_sapiens\_assembly38.dbsnp138.vcf -glm BOTH -o SRR1517898\_snp\_UnifiedGenotyper.vcf -stand\_call\_conf 30.0

-rw-r--r-- 1 x1333kmj pc1081 7353566 Jan 21 09:37 SRR1517898\_snp\_UnifiedGenotyper.vcf

아직 용량이 적음

I’ve tried to put everything inside bwa folder (the other student did so, but I don’t think this is the problem, though)

Bwa/bwa-0.7.15 폴더에 넣어서 함

Dictionary 용량 증가, line 수 그대로 3367

1007% [x1333kmj@tachyon3200 bwa-0.7.15]$ bwa index -a bwtsw Homo\_sapiens\_assembly38.fasta

1008% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/samtools/samtools-1.3.1/samtools faidx Homo\_sapiens\_assembly38.fasta

1009% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -jar ~/picard-tools-2.4.1/picard.jar CreateSequenceDictionary R=Homo\_sapiens\_assembly38.fasta O=Homo\_sapiens\_assembly38.dict

1012% [x1333kmj@tachyon3200 bwa-0.7.15]$ bwa mem -M -t 8 Homo\_sapiens\_assembly38.fasta SRR1517898\_1.fastq SRR1517898\_2.fastq > SRR1517898.sam

1014% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/samtools/samtools-1.3.1/samtools view -bhS SRR1517898.sam > SRR1517898.bam

1015% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -jar -Xmx3g -Djava.io.tmpdir=/scratch2/x1333kmj/x1333kmj -jar ~/picard-tools-2.4.1/picard.jar AddOrReplaceReadGroups I=SRR1517898.bam O=SRR1517898\_addrg.bam SORT\_ORDER=coordinate RGID=SRR1517898ID RGLB=SRR1517898 RGPL=illumina RGPU=none RGSM=SRR1517898SM

1018% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -Xmx3g -Djava.io.tmpdir=/scratch2/x1333kmj/x1333kmj -jar ~/picard-tools-2.4.1/picard.jar FixMateInformation I=SRR1517898\_addrg.bam O=SRR1517898\_addrg\_fixm.bam SORT\_ORDER=coordinate

1020% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -Xmx3g -Djava.io.tmpdir=/scratch2/x1333kmj/x1333kmj -jar ~/picard-tools-2.4.1/picard.jar SortSam I=SRR1517898\_addrg\_fixm.bam O=SRR1517898\_addrg\_fixm\_sorted.bam SORT\_ORDER=coordinate

1022% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -Xmx3g -Djava.io.tmpdir=/scratch2/x1333kmj/x1333kmj -jar ~/picard-tools-2.4.1/picard.jar SortSam I=SRR1517898\_addrg\_fixm\_sorted.bam O=SRR1517898\_addrg\_fixm\_sorted\_md.bam METRICS\_FILE=SRR1517898.dub REMOVE\_DUPLICATES=true ASSUME\_SORTED=true

1022% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -Xmx3g -jar ~/picard-tools-2.4.1/picard.jar MarkDuplicates I=SRR1517898\_addrg\_fixm\_sorted.bam O=SRR1517898\_addrg\_fixm\_sorted\_md.bam METRICS\_FILE=SRR1517898.dub REMOVE\_DUPLICATES=true ASSUME\_SORTED=true

WARNING 2017-01-21 13:31:06 AbstractOpticalDuplicateFinderCommandLineProgramA field field parsed out of a read name was expected to contain an integer and did not. Read name: SRR1517898.2. Cause: String 'SRR1517898.2' did not start with a parsable number.

1023% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -Xmx3g -jar ~/picard-tools-2.4.1/picard.jar BuildBamIndex I=SRR1517898\_addrg\_fixm\_sorted\_md.bam

1024% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -Xmx3g -jar ~/gatk/GenomeAnalysisTK.jar -T RealignerTargetCreator -R Homo\_sapiens\_assembly38.fasta -I SRR1517898\_addrg\_fixm\_sorted\_md.bam --known ~/Homo\_sapiens\_assembly38.known\_indels.vcf -o SRR1517898\_addrg\_fixm\_md\_IndelRealigner.intervals

1026% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -Xmx3g -Djava.io.tmpdir=/scratch2/x1333kmj/x1333kmj -jar ~/gatk/GenomeAnalysisTK.jar -T IndelRealigner -R Homo\_sapiens\_assembly38.fasta -I SRR1517898\_addrg\_fixm\_sorted\_md.bam -known ~/Homo\_sapiens\_assembly38.known\_indels.vcf -targetIntervals SRR1517898\_addrg\_fixm\_md\_IndelRealigner.intervals -o SRR1517898\_addrg\_fixm\_sorted\_md\_realigned.bam

1028% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -jar ~/gatk/GenomeAnalysisTK.jar -T BaseRecalibrator -R Homo\_sapiens\_assembly38.fasta -I SRR1517898\_addrg\_fixm\_sorted\_md\_realigned.bam -knownSites ~/Homo\_sapiens\_assembly38.dbsnp138.vcf -o SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.table

1030% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -jar ~/gatk/GenomeAnalysisTK.jar -T PrintReads -R Homo\_sapiens\_assembly38.fasta -I SRR1517898\_addrg\_fixm\_sorted\_md\_realigned.bam -BQSR SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.table -o SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.bam

1031% [x1333kmj@tachyon3200 bwa-0.7.15]$ ~/jdk/jdk1.8.0\_102/bin/java -jar ~/gatk/GenomeAnalysisTK.jar -T UnifiedGenotyper -R Homo\_sapiens\_assembly38.fasta -I SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.bam --dbsnp ~/Homo\_sapiens\_assembly38.dbsnp138.vcf -glm BOTH -o SRR1517898\_snp1.vcf -stand\_call\_conf 30.0

-rw-r--r-- 1 x1333kmj pc1081 7353597 Jan 21 15:43 SRR1517898\_snp1.vcf

ALL THE FILE SIZES

-rw-r--r-- 1 x1333kmj pc1081 487464 Jan 21 13:12 Homo\_sapiens\_assembly38.dict

-rw-r--r-- 1 x1333kmj pc1081 3249912778 Jan 12 18:14 Homo\_sapiens\_assembly38.fasta

-rw-r--r-- 1 x1333kmj pc1081 20199 Jan 21 12:50 Homo\_sapiens\_assembly38.fasta.amb

-rw-r--r-- 1 x1333kmj pc1081 455474 Jan 21 12:50 Homo\_sapiens\_assembly38.fasta.ann

-rw-r--r-- 1 x1333kmj pc1081 3217347004 Jan 21 12:49 Homo\_sapiens\_assembly38.fasta.bwt

-rw-r--r-- 1 x1333kmj pc1081 160928 Jan 21 13:10 Homo\_sapiens\_assembly38.fasta.fai

-rw-r--r-- 1 x1333kmj pc1081 804336731 Jan 21 12:50 Homo\_sapiens\_assembly38.fasta.pac

-rw-r--r-- 1 x1333kmj pc1081 1608673512 Jan 21 13:07 Homo\_sapiens\_assembly38.fasta.sa

-rw------- 1 x1333kmj pc1081 490013400 Jan 18 13:05 SRR1517898\_1.fastq

-rw-r--r-- 1 x1333kmj pc1081 490013400 Jan 12 13:47 SRR1517898\_2.fastq

-rw-r--r-- 1 x1333kmj pc1081 350156364 Jan 21 13:22 SRR1517898\_addrg.bam

-rw-r--r-- 1 x1333kmj pc1081 353818965 Jan 21 13:26 SRR1517898\_addrg\_fixm.bam

-rw-r--r-- 1 x1333kmj pc1081 44372288 Jan 21 14:02 SRR1517898\_addrg\_fixm\_md\_IndelRealigner.intervals

-rw-r--r-- 1 x1333kmj pc1081 353818965 Jan 21 13:29 SRR1517898\_addrg\_fixm\_sorted.bam

-rw-r--r-- 1 x1333kmj pc1081 3994952 Jan 21 13:35 SRR1517898\_addrg\_fixm\_sorted\_md.bai

-rw-r--r-- 1 x1333kmj pc1081 349637822 Jan 21 13:33 SRR1517898\_addrg\_fixm\_sorted\_md.bam

-rw-r--r-- 1 x1333kmj pc1081 3994808 Jan 21 14:07 SRR1517898\_addrg\_fixm\_sorted\_md\_realigned.bai

-rw-r--r-- 1 x1333kmj pc1081 349688300 Jan 21 14:07 SRR1517898\_addrg\_fixm\_sorted\_md\_realigned.bam

-rw-r--r-- 1 x1333kmj pc1081 3996664 Jan 21 14:29 SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.bai

-rw-r--r-- 1 x1333kmj pc1081 648575946 Jan 21 14:29 SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.bam

-rw-r--r-- 1 x1333kmj pc1081 701757 Jan 21 14:19 SRR1517898\_addrg\_fixm\_sorted\_md\_realigned\_recal.table

-rw-r--r-- 1 x1333kmj pc1081 346264472 Jan 21 13:17 SRR1517898.bam

-rw-r--r-- 1 x1333kmj pc1081 2752 Jan 21 13:33 SRR1517898.dub

-rw------- 1 x1333kmj pc1081 1334187764 Jan 21 13:15 SRR1517898.sam

-rw-r--r-- 1 x1333kmj pc1081 7353597 Jan 21 15:43 SRR1517898\_snp1.vcf

-rw-r--r-- 1 x1333kmj pc1081 192853 Jan 21 15:43 SRR1517898\_snp1.vcf.idx