Supplemental Information



Supplemental Fig. S1: Filtration of SV calls using in-house databases in **A)** srGS and **B)** IrGS. dashed lines GRCh38, full lines T2T. Number of calls in log10 scale.

Supplemental Table S1. Number of SV calls per individual in T2T-CHM13 aligned genomes and *de novo* genomes.

Case ID	Cytogenetic	Genom	Affec	Cytoband	Cytoband B	A-B
	abberation	е	ted	А		matc
			chr			h
RD_P525	inv(5)(p13q23)	12321	659	128	39	1
P4855_501	inv(6)(p12q16.3)	12636	503	42	32	1
P4855_501 de	inv(6)(p12q16.3)	37127	1579	189	83	2
novo						
BH16643-1	inv(9)(q12q34.3)	19105	1888	1524	86	0
BH16643-1 de	inv(9)(q12q34.3)	41400	3091	2237	235	1
novo						
P4855_106*	inv(10)(q11q23)	20068	515	129	54	2

RD_P541	inv(12)(p11.23q13.3)	10449	482	105	56	1
RD_P541 de novo	inv(12)(p11.23q13.3)	37551	1524	488	130	2
RD_P549	inv(14)(q24q32)	12513	466	45	107	1
RD_P526	inv(18)(p11.23q21.1)	12942	451	133	50	1
RD_P542	inv(19)(p13.2q13.4)	13021	574	42	91	1
RD_P546	inv(19)(p13.2?q13.4)	13050	483	44	94	1

*srGS data

Α	RD_P525 Chr5: (+) 42125501	TTTTAATAATGCTTTATATATCCAAGTCTATCTGCAGTTATAAAAATGCTTCCTAAGAGGCATTTAAGAGTAA
	Chr5: (-) 127429118	
	Chr5: (+) 42125495	CTTTTAATAATGCTTTATATATCCAAGTCTATCTGCAGTTATAAAAATGCTTCCTAAGAGGCATTTAAGAGTA
	Chr5: (-) 127429109	ACACCTCACACGGCCGGGTACTCCAACAGACCTGCAGCGGGGCCCTGTCTGT
в	P4855_501	
	Chr6: (-) 51032755	GGATTCACAGCCGAATTCTACCAGAGGTACAAGGAGGAACTGGTACGATTCCTTCTGAAACT IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
	Chr6: (+) 94376921 Chr6: (-) 51032765	CCAAGTGATGACAGTGCAAAGCAGCTGTATAATTCTTCAGTTGCTCTTTTTCTGTTGCTAGA GACCAGATGGATTCACAGCCGAATTCTACCAGAGGTACAAGGAGGAACTGGTACGATTCCTT
C	Chr6: (+) 94376918	CATCCCTTCCAAGTGATGACAGTGCAAAGCAGGAGGTACAAGGAGGAACTGGTACGATTCCTT IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
U	F4000_100	
	Chr10: (+) 42197576 -42315905	TCCATTCCATTCCATTCCATTCCACTCGGGTT TCCATTCCAT
	Chr10: (-) 96022615	ΤΑΑΤΑCΑΑGTAATACTTTTTTAAGACACAAATGATATTTTATTAATTGTGATAATATATACATAACATAAAAT
	Chr10: (-) 42197576 - 42315905	TTGTGTTGAATCCATTCCATTCCATTCCACTCGGGTTGATTCCCGTCCTTTCCATT
	Chr10: (+) 96022600	TITIAATGTCTAGATAATACAAGTAATACTTTIIIIAAGACACAAATGATAIIIIATTAATTGTGATAATATATAC
D	RD_P541	
	Chr12: (+) 32945545	GGCCATTAGAGTTACATAGAAATACACTAGAGTCCACTGCTACTGGAGCTCTTGCCCTAGAGTGAAAACTGGG
	Chr12: (-) 58051150	GCAAGTCAGAATCTTTTTGCTGGTGAAGGATCTTGCCTCAATGTTGACAGCTAATGAGTGCATCAGGGTGGGG
	Chr12: (-) 32945540	GCATCCCCCACCCTGATGCACTCATTAGCTGTCAACATTGAGGCAAGATCCTTCACCAGCAAAAAGATTCTGA
	Chr12: (+) 58051144	GTTTTCACTCTAGGGCAAGAGCTCCAGTAGCAGTGGATTGAGGCAAGATCCTTCACCAGCAAAAAGATTCTGA GTTTTCACTCTAGGGCAAGAGCTCCAGTAGCAGTGGACTCTAGTGTATTTCTATGTAACTCTAATGGCCTTAA
Е	RD_P549	
	Chr14: (+) 63951601	TGGAGTGCAGTGCGCGTAATCTTGGCTCACTGCAAGCTCTGCCTCGCGTCCAGGCACAGAGAAGCTGTAAGAC
	Chr14: (-) 97962156	GGGCGCCAGGCCTGGGGACGCCCTGAAGATGCTTCCTCAGCTGGAGGACCCAGGCACAGAGAAGCTGTAAGAC
	Chr14: (+) 63951610	
	Chr14: (-) 97962152	CGCCAGGCCTGGGGACGCCCTGAAGATGCTTCCTCAGGGTTCATGCCATTCTCCTGCCTCAGCCCCCGAGTAG
F	RD_P526 Chr18: (+) 7340601	
	Chr18: (-) 47883888	ATCAGAICIACCIGAAGIIGAGIITAIGICCIGCCITAGCCAGCIACIAGCCGAGIGAICIGIAAIAIAICI ATTAATTATCAATTAGGAAGAAGTTACATAATATATTGGTTATTCTTTTTGTAACTGAAAGACTTTATACTA
	Chirlo: (+) 7340602	
_	Chr18: (-) 47883889	ATCAGATCTACCTGAAGTTGAGTTTATGTCCTGCCTTAGCCAGCTACTAGCCGAGTGATCTGTAATATATCT
G	RD_P542	Inserted bases: taaaacagaccttaaaacatta
	Ghr19: (+) 9982025	
	Chr19: (-) 58312124	ΑΤΑΤΑΤΑCΑΤΤΑΤΑΤΑΤΑΤΑΤΑCACATTATATATATATATACATTATATATACACATTATATATATATACACTT
	Chr19: (+) 9982025	CTGCACCTGGCCACTATTGATCTTAAAACAGACCCTAAGTGGACAATGCTCCATCCCAAAGGACCCAGGGAA
		ACACATTATATATATATATATATATATATACACATAAGTGGACAATGCTCCATCCCAAAGGACCCAGGGAA
	Chr19: (-) 58312461	ACACATTATATATATATACATTATATATATACACATATATATACATTATATATACACATATATATACACAT

Supplemental Fig. S2: Breakpoint sequences of inversions. **A)** Inversion 5 (RD_P525), **B)** Inversion 6 (P4855_501), **C)** Inversion 10 (P4855_106), where the one of the breakpoint sequences at 10:42197576-42315905 could not be fully pinpointed, **D)** inversion 12 (RD_P541), **E)** inversion 14 (RD_P549), **F)** inversion 18 (RD_P526) and **G)** inversion 19 (RD_P542) breakpoint sequences. Microhomology is indicated in green, unmatching

sequences in black and repeat sequences is underlined. The direction and positions are indicated at the left side of the figure.

		Mellin and a second	
	a desired as some and as the second		
		canada a sumal a composition and a state of the second second	
30.00	ana je		Pase
INV18 RD P526	jan jan	D INV19 RD P542	-
INV18 RD_P526		D INV19 RD_P542	
INV18 RD_P526		D INV19 RD_P542	
INV18 RD_P526		D INV19 RD_P542	
INV18 RD_P526		D INV19 RD_P542	

Supplemental Fig. S3: IGV images of resolved inversions (INV) with breakpoints in mappable regions A) INV5 (RD_P525), B) INV14 (RD_P549), C) INV18 (RD_P526) and D) INV19 (RD_P542).

A GRCh37



B GRCh38

C T2T



Supplemental Fig. S4: Integrated genomics viewer (IGV) images of the inversion 6 breakpoint regions in short-read (upper), linked-read (middle, only A) and long-read (lower) data in GRCh37 (A), GRCh38 (B), T2T (C), chimpanzee (D) and bonobo (E).



Supplemental Fig. S5: Comparison of the inversion breakpoint region on chromosome 6p and chromosome 10q across reference genomes. The reference sequence on chromosome 6 affected by an inversion compared in A) GRCh38-T2T, B) Chimpanzee to T2T and C) Bonobo to T2T. The reference sequence on chromosome 10 affected by an inversion compared in D) Chimpanzee to T2T and E) Bonobo to T2T.

Α	chr6:35,869,257-112,1 p25.1 p24.1	121,925 p22.3 p22.2	p21.32	p21.1 p12.5	s p12.1 q1	1.1 q	13 q14.1	q14.3	q16.1 q	16.3 q21	q22.1	q22.32	q23.2	q24.1	q243	q25.2	q26 q27
	40 mb	1	50 mb 		68 mb 	1	70 mb 	- 76 mb —	80 mb 	1		10 mb	1	100	mb	I	110 mb
INV6.bed			_									_					
Aorta_STL002_Leung_2015-raw_ .bed.txt																	
Cortex_DLPFC_Schmitt2016-raw_' s.bed.txt																	
Hippocampus_Schmitt2016-raw_7, .bed.txt			-		-			-									
LNCAP_raw-merged_TADa_prosta ed.txt													_	-			
SJCRH30_raw-morged_TADs_mus ed.bt				4	-												
SKNMC_raw-merged_TADs_brain. txt												11					
Rofseq Genes	PK1 GLO1 TREMI	SUPT3H	MMUT MIR206	HCRTR2	KHDRBS2	EYS XR_94	2661.2 RIM	151 COL12A1	IRAK1BP1	IBTK TBX18	CGA M	DN1 EPH	H	9 PNKY	GRIK2	HACE1 S	OBP GPR6

	2-0	4.4						
	i i i i i i i i i i i i i i i i i i i							
			-					
			Concession in the local division in succession in the local division in the local divisi		and the second s			
the			Research & Man Discourse	No. 1996 In cost II	The supervised in the local division of the			COLUMN TWO IS NOT THE OWNER.
TITLANA					TAXABLE PARTY IN COMP.			
AC_TITHE								
TETANO								
	11	A LYZLARI EPC-RE CREMER AND/PLAN	LINE DI BRANCH BUILDER DE LE LE COMPANY	1.221 PM-221 TM/ME-222 NEAP3-221 PM-220			SOURCE AND A REAL PROPERTY A REAL PROPERTY A REAL PROPERTY AND A REAL PROPERTY AND A R	AD1 PHEP13E1 TOF715222 PALP-201
								-
	09/12/17.595.556-73.595.390 01112/011210/01121/01122/01111/00/27/11122	9122 PITZI PITZI PITZI	\$2 \$1111 \$1111	1117 (111 (112 (113 (11	211 - 2112 - 21131 - 4	aaa atai aa am	anz ann ann ar	
	0/12/7.144.536-73.594.390 	9121 <u>9121 9121 911</u>	412 - 1 1111 - 1111	1112 6061 4072 4083 405 64 mb	an an an .	ana ana an an	อา อาก องการหก	- 201 - 201 - 201
	04/2/7/36536-7536238 21537 21537 21537 2152 2153 2177 2122	9122 9123 9121 913	87 <u>8111 8111</u>		971 9717 9131 9	สม า อา - อ บ		aw 903 901 900
	91-12 (11.19) (11.19) 11.11 11	944 8121 9121 813	42 413.11 413.1		9713 9712 9713 9	ແມ່ວຍ ວິຍ	ານ ອກ ຈະດີຈະດີ	7153 (243) (243)
	001217,043.08.03,088.00 01127,0133.02.012,0133.02.01 0100 0100 0100 0100 0100 0100 010	922 9127 9127 913	41 <u>1111 111</u>	1117 (14) (172 (14) (11)	975 - 913 - 9131 - 9 Mai	ana <mark>ata ar an</mark>	สม อบ อนเร ลก 	903 900 900 1
17D.Root		912 TUT 111	67 <u>1111</u> 11	1112 112 112 112 113 11 1112 112 112 112 113 113 113 113 113 113	905 atta atta a	, ,	an an an an an	402 4301 400 ",**
T2TLR.sed	0412 773508 738589 2012 1313 1312 1313 1313 1313 1313 1313	на <u>на</u> на <u>на на н</u>		1177 - 1181 - 1177 (1191 - 1177 - 1197 - 1177 - 1197 - 1177 - 119	975 972 9131 - ***	1		
TTLR.sed	0012 7 14038 7.06.00 <u>01337 13337 0137 0137 0137 0137 0137 013</u>	900 802 ⁶ 9107 807	er ann ann ann ann ann ann ann ann ann an		975 orz oros o			201 203 203
. TTTLR.seef Roed Roed	947 94 94 94 94 94 94 94 94 94 94 94 94 94	900 800 900 900 **	44 1111 111		911 01.7 01.0 0		an an am ² an	- 100 - 100 - 100 - 100
TyTh.Rood Nord R3rd	0012 77 55 38 73 56 39 		,	1077 (101 (107 (10 (10) 107 (10 (10 (10 (10) 107 (10 (10 (10 (10 (10 (10 (10 (10 (10 (10	971 917 918 9	111 2111 21 211	en en en en en	203 uus 203
, T2TLR.sed R.Sed R.Sed FAD_T2T.Sed TAD_T2T.Sed		100 Inter Tor Tor	er en en		973 913 913 913 1	123 013 07 03	<i>φιτ φ</i> ια ^π φια ^π	200 ⁷ 2007 2009
LT2TL:Rood Ribed Ribed (ADL T2Tand LT2TTand		900 900 900 900 900 ***					ea - an em ² em ²	200 ³ 2007 200 ","
,1271.Rood Libed R.land ;1273.and								

Supplemental Fig. S6: Topologically Associated Domains and A) Inversion 6, B) Inversion

10 and C) Inversion 12.



Supplemental Fig. S7: Inversion affecting chromosome 12 in case RD_P541. A) The inversion breakpoint region on 12q in GRCh38, GRCh37, chimpanzee and bonobo aligned to T2T. B) Breakpoint regions in GRCh37. C) Breakpoint regions in GRCh38. D) Breakpoint regions in chimpanzee.



Supplemental Fig. S8: Optical genome mapping of chromosome 12, RD_P541.



Supplemental Fig. S9: Optical Genome mapping of the chromosome 9, BH16643-1.

Pinpoints of the inversion can be observed as unaligned regions.



Reduced Expression of EHMT1 in Patient

Supplemental Fig. S10: Reduced expression of EHMT1 in BH16643-1.



Supplemental Fig. S11: DRR analysis. Bar plot of all DRRs in GRCh38. Orange indicating all compared genomes, blue bonobo, purple chimpanzee, green GRCh37 and brown T2T.



Supplemental Fig. S12: Distribution of DRRs in 100 Swedish individuals. A) DRRs between T2T-GRCh38 and their presence in Swedish individuals. Blue indicating absent (<8X) and red present (>8X and <100X). B) Violin plot of coverage across the respective DRRs in 100 Swedish individuals. C) The distribution of population frequencies of the detected GRCh38-GRCh37 (green) and T2T-GRCh38 (orange) DRRs.



Supplemental Fig. S13: Bar plots A) DRRs in Chimpanzee-T2T **B)** DRRs between Bonobo-T2T. Blue indicating absent in Swedish individuals and red indicate present. **C)** Population frequency of DRRs of Chimp-T2T **D)** DRRs between Bonobo-T2T.



Supplemental Fig. S14: DRRs of GRCh38-GRCh37 and their presence in Swedish individuals. Blue indicating absent and red present.

Supplemental Table S2: Median, minimum and maximum length of DRRs (Mbp).

Median Min Max	
----------------	--

	GRCh38	20	10	2270
Template	T2T	20	10	930
	Chimpanzee	10	10	450
	Bonobo	10	10	300

Supplemental Table S3: Coverage thresholds determining genomics copies.

Coverage	Copies
8-24x	1
25-38	2
39-45	3
46-100	>3



Supplemental Fig. S15: Distribution of mapping quality of reads aligned to DRRs in A) GRCh38-GRCh37 and B) T2T-GRCh38.

Supplemental Document 1: Detailed clinical description of BH16643-1

BH16643-1 is a 4-year-old girl with hypotonia and global developmental delay. Pregnancy was complicated by polyhydramnios. She was born at 39 weeks gestational age (birth weight 3317 grams, birth length 49.5 cm, and FOC 33 cm). She required phototherapy for hyperbilirubinemia and otherwise received routine care in the nursery. As to her development, she gained head control at 7 months, rolled over at 9 months, sat up unsupported at 12 months, and learned to sit up independently when she was 2.5 years old. At age 4 years, she can crawl and bear weight on feet with support, but she is not cruising or walking. She reached to grab an object at age 7 months and developed pincer grasp at age 16 months. She is nonverbal. She shows self-harmful behaviors when frustrated or irritated. There is no history of developmental regression or seizures. The medical history is otherwise significant for feeding difficulties, initially presenting with difficulty latching to breast and poor suck, followed by dysphagia and persistent gastroesophageal reflux. She has chronic constipation. She has a congenital heart defect, including a small atrial septal defect that closed, and mild pulmonary valve stenosis. She has bilateral optic nerve atrophy and intermittent exotropia, requiring corrective lenses. Her history is also notable for recurrent respiratory infections, associated with frequent nasal congestion and ear infections requiring repeated pressure equalizing tube placement. Family ethnicity is mixed Caucasian, Native American and Cajun, and the family history was non-contributory. On physical examination, her growth was within the normal limits, and she was noted to have dysmorphic craniofacial features, including midface retrusion, everted lower lip and prognathism. The neurological exam demonstrated hypotonia and normal muscle strength. The patient had an abnormal brain magnetic resonance imaging (MRI) at age 2 years that showed hypoplastic corpus callosum and pontocerebellar hypoplasia. Spinal MRI revealed closed sacral dysraphism (spina bifida occulta) at S2-S3 that was felt to be an incidental finding. Electroencephalography (EEG) was normal. Metabolic labs, including ammonia, lactate, plasma amino acids, urine organic acids, acylcarnitine analysis, urine amino acids, very long chain fatty acids, 7-dehydrocholesterol, and congenital disorders of glycosylation biochemical screening panel did not show significant abnormalities. Genetic testing included a chromosomal microarray analysis (CMA), trio exome sequencing and mitochondrial DNA sequencing, Prader Willi methylation study, SMN1/SMN2 dosage analysis and DMPK repeat expansion analysis, all with normal results. Through clinical genome sequencing conducted in the UDN the patient was found to be a carrier for Friedreich ataxia, with a heterozygous FXN repeat expansion (approximately 1050 repeats) that was inherited from her mother. Chromosome analysis in blood and skin fibroblasts identified a heterozygous, paracentric inversion on the long arm of chromosome 9 [46,XX,inv(9)(q12q34.3)]. Parental chromosome analysis was normal, suggesting the inversion arose *de novo* in the patient.

Supplemental Code:

Align sr

bwa mem -p -t 16 <ref> <fastq>

Align and SV calling of linked-read longranger wgs –id <id> --reference <ref> --fastq <fastq> --vcmode freebayes

Denovo analysis

YAK yak count -k31 -b37 -t16 -o \$2.yak <fasta>

Hifiasm hifiasm -o \$2 -t 16 \$fastq

GFA to FA awk '/^S/{print ">"\$2;print \$3}' hap.gfa > \$2.fa

align FA
minimap2 -R "@RG\tID:\$2\tSM:\$2" -a -t 16 --MD -x map-pb -asm5 -Y -y \$ref \$2.fa |
samtools view -Sbh - | samtools sort -m 4G -@16 - > \$2.bam
samtools index \$2.bam

SVIM svim-asm diploid \$PWD <hap1> <hap2> \$ref --sample \$3

Quast python quast <hap1> <hap2> -o \$3.quast -t 16 -r \$ref

Reference genome anlaysis

minimap2 -cx asm5 template.fa query.fa > aln.paf

Find DRRs

Create BAM file ref-to-ref /ref1=ref ref2=asm

minimap2 -ax asm5 <ref1> <ref2> | samtools view -Sbh - > tmp.ref1.ref2 samtools sort -m 4G -@1 tmp.ref1.ref2 > ref1_ref2.bam samtools index ref1_ref2.bam

Run TIDDIT coverage module to find gaps

tiddit=container.sif singularity exec --bind /dataset \$tiddit tiddit --cov --bam ref1_ref2.bam o ref1_ref2.cov -z 10000

Remove known gap regions in template

bedtools intersect -v -a ref1_ref2.cov -b gapsRef1.bed > ref1_ref2.novelgap.bed

Collect all gap regions : coverage of 0 across the bin
grep -P '\t0\.0' ref1_ref2.novelgap.bed >> ref1_ref2.novelgap.bed

Count DRRs

import sys import statistics

find Mbp of DRR sequence that have low-medium-high coverage AND find amount average coverage across gap regions AND

```
identify variable regions
INPUT: cov file from TIDDIT
....
mychr = list(range(0,23))
mychr = [str(i) for i in mychr]
mychr.append('X')
mychr.append('Y')
genome='T2T'
#genome ='GRCh38'
#genome='GRCh37'
#genome='Chimpanzee'
#genome='Bonobo'
def collect_cov(file): # collects info from file
       covdict = \{\}
       bindict = \{\}
       for line in open(file):
               tabline = line.rstrip('\n').split()
               chr=tabline[0]
               #if chr not in mychr:
               #
                       continue
               start = tabline[1]
               end = tabline[2]
               cov = float(tabline[3])
               if chr not in covdict:
                       covdict[chr] = \{\}
                       bindict[chr]={}
               if start not in covdict[chr]:
                       covdict[chr][start] = []
                       bindict[chr][start] = end
               covdict[chr][start].append(cov)
               lst=[chr, start, end, str(cov)]
               #print('\t'.join(lst))
       return covdict, bindict
def count_presence(covdict, bindict):
       for chr in covdict:
               for s in covdict[chr]:
```

average = sum(covdict[chr][s])/len(covdict[chr][s])
covlist = covdict[chr][s]

#count presence
lowcov = len([i for i in covlist if i <= 8])
highcov = len([i for i in covlist if i >8 and i <100])</pre>

```
if highcov >= 5:
status = 'Common'
```

else:

status = 'Absent'

lst = [chr, s, bindict[chr][s], str(average), genome, status]
print('\t'.join(lst))

return

def find_amount(file):
 amount=[]
 for line in open(file):
 am = int(line.rstrip('\n'))
 amount.append(am)
 mean = sum(amount)/len(amount)
 median = statistics.median(amount)

```
meanMBP=mean*10000/1000000
medianMBP=median*10000/1000000
```

```
print('median', medianMBP)
print('mean', meanMBP)
```

```
find_amount(sys.argv[1]) #find amount of DRRs
all_cov= collect_cov(sys.argv[1]) #coverage across all regions
count_presence(all_cov[0], all_cov[1])
```

Multimapping reads

import sys import pysam

```
reference_filename='/sw/data/uppnex/ToolBox/hg38bundle/Homo_sapiens_assembly38.fast
a' )
bamfile = pysam.AlignmentFile(sys.argv[2], "rc", reference_filename='reference.fna')
regions=sys.argv[1]
```

```
def myregions(file):
    reg= {}
```

```
for line in open(file):

chr=line.split(':')[0]

if chr == 'chrMT':

continue

start=int(line.split(':')[-1].split('-')[0])

end=int(line.rstrip('\n').split(':')[-1].split('-')[1])

if chr not in reg:

reg[chr]={}
```

```
if start not in reg[chr]:
reg[chr][start]=end
```

return reg

```
def findMM(bam, regdict):
       mm = \{\}
       flagmm={}
       myqual=[256,2048,2057, 2056, 2129, 2209, 2193]
       for chr in regdict:
              for start in regdict[chr]:
                     end=regdict[chr][start]
                     for read in bamfile.fetch(chr, start, end, until_eof=True):
                             flag= int(read.flag)
                             #print(flag)
                             name=read.query_name
                             if flag in myqual:
                                    if name not in flagmm:
                                            flagmm[name]=0
                                    flagmm[name]+=1
                             if name not in mm:
                                    mm[name]=0
                             mm[name]+=1
       return mm, flagmm
```

regionsdict=myregions(regions) multimapping= findMM(bamfile, regionsdict)

total=0 for r in multimapping[1]:

```
if multimapping[1][r] > 1:
total += 1
```

```
perc1=len(multimapping[1])/len(multimapping[0])*100
print(str(bamfile), len(multimapping[0]), len(multimapping[1]), perc1, total )
```

```
# Mapping quality
import sys
import pysam
bamfile = pysam.AlignmentFile(sys.argv[2], "rc",
reference_filename='/sw/data/uppnex/ToolBox/hg38bundle/Homo_sapiens_assembly38.fast
a')
#bamfile = pysam.AlignmentFile(sys.argv[2], "rc",
reference_filename='/proj/sens2017106/nobackup/kristine/reference_stuff/GCF_009914755.
1_T2T-CHM13v2.0_genomic.fna')
regions=sys.argv[1]
def myregions(file):
       reg = \{\}
       for line in open(file):
               chr=line.split(':')[0]
               if chr == 'chrMT':
                       continue
               start=int(line.split(':')[-1].split('-')[0])
               end=int(line.rstrip('\n').split(':')[-1].split('-')[1])
               if chr not in reg:
                       reg[chr]={}
               if start not in reg[chr]:
                       reg[chr][start]=end
       return reg
def findMQ(bam, regdict):
       mq=[]
       for chr in regdict:
               for start in regdict[chr]:
                       end=regdict[chr][start]
                      for read in bamfile.fetch(chr, start, end, until_eof=True):
                               mapq= str(read.mapping_quality)
                               mq.append(mapq)
                               print('\t'.join([str(chr), str(start), str(end), str(mapq)]))
       return mg
```

regionsdict=myregions(regions) mappingquality= findMQ(bamfile, regionsdict)