Analysis of Breast-Predict Project

ChIP-Seq Experiment

Introduction

A ChIP-Seq experiment targeting a specific transcription factor (TF) was performed on a human breast cancer cell line from the Irish Cancer Society's Breast-Predict Project. The resultant data has been analysed here to provide genome-wide TF binding locations, to offer potential regulatory roles of the TF, and to perform *de novo* TF motif discovery and comparison.

Although this analysis was performed using a premade script with automated file naming, the following will follow the steps taken to provide these results by explaining all relevant commands one at a time using what would be the generated file names in the command line. The actual script used is provided for comparison. It should also be noted that this script and thus all commands shown were run on the NUIG campus high-powered computing cluster to facilitate reasonable computing times.

Quality Control

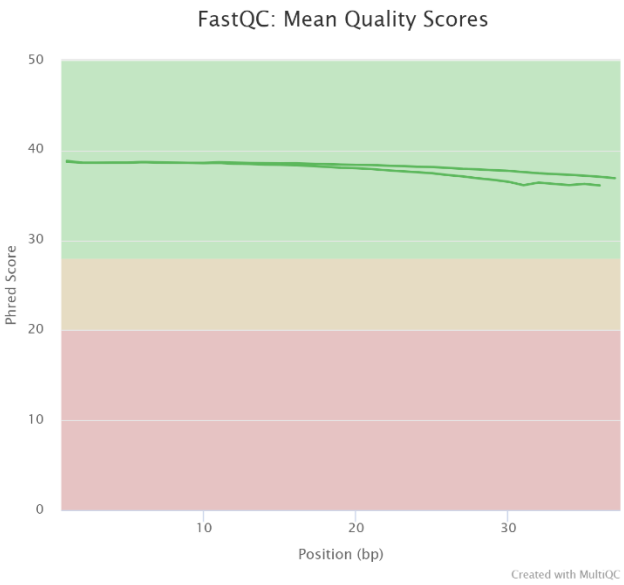
Two fastq sequence files were provided for analysis: chip.fastq, which contains the TF-targeted immunoprecipitated sequences; and input.fastq, which contains the experimental control sequences that were not immunoprecipitated. FastQC, followed by MultiQC for visualization, were used to first examine the quality of the sequence data:

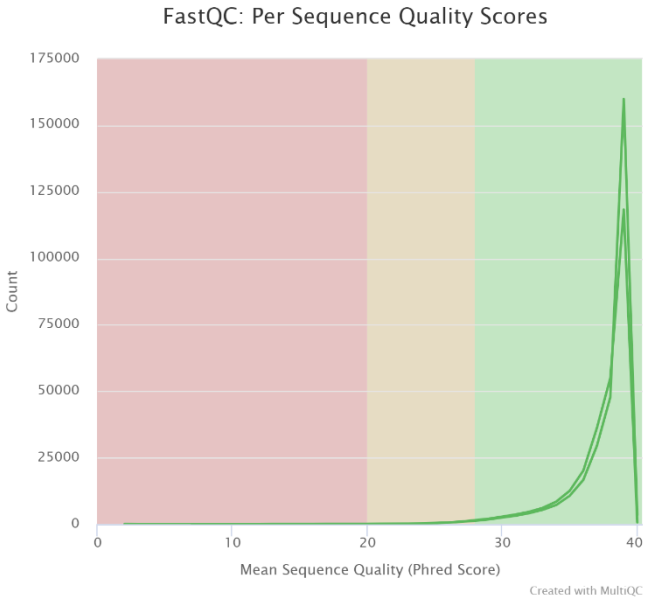
```
fastqc chip.fastq

fastqc input.fastq

multiqc .
```

Summary data are shown below:





Both sets of sequences show high quality. While mean quality drops towards the ends of the sequences, this is a normal feature of high-throughput sequencing. Furthermore, both sets of data show very few amounts of individual low-quality reads overall. Because of the high quality of reads, both data sets will be used as-is for the remainder of the analysis.

Alignment

To begin looking at binding site locations, both the ChIP and control sequences were first aligned to human reference genome hg19 using Bowtie2. For this analysis, only chromosome 21 was used for mapping, as the data provided correspond to a subset of the dataset that map to chromosome 21.

First, an index (named "ref_index" in the command) of the reference genome (chr21.fa) is built:

```
bowtie2-build chr21.fa ref_index
```

- bowtie2-build: specifies for bowtie2 to build a reference genome index
- chr21.fa: the name of the file containing the reference genome
- ref_index: the prefix name that will be given to the index files generated

Then, the alignment is run. The following is the code and parameters used to align the chip.fastq sequences:

```
bowtie2 -x ref_index -U chip.fastq -S chip.sam
```

- -x: specifies that ref_index is the prefix name of the genome index files to use
- -U: specifies that the file chip.fastq for alignment contains unpaired reads
- -S: specifies that the output should be written to a SAM (sequence alignment map) file with the name chip.sam

This produces a SAM file with alignment information for each read sequence. The same procedure was followed to produce a SAM file for the control sequences in input.fastq as well. An excerpt of 4 sequences from the chip.sam file is shown here, which besides sequence and alignment information also contains per-base alignment confidence scores:

```
@HD   VN:1.0     SO:unsorted
@SQ   SN:chr21   LN:48129895
@PG   ID:bowtie2 PN:bowtie2 VN:2.1.0
SRR540192.1580   0     chr21 20649661   42    37M   *     0     0
      CATCTTGGCCTCTGTGCAGCATTCCTTTCTCCATGGT
      IIIIIIHIIIIIIHIIIIIIIIIIIIIIIHIIHIHID AS:i:0     XN:i:0     XM:i:0
      XO:i:0     XG:i:0     NM:i:0     MD:Z:37    YT:Z:UU
SRR540192.1752   0     chr21 44763347   42    37M   *     0     0
      GCTCCCAGAAACCCAGGGCCACTGGCAGCTTCAGGGA
      GGGGGGGBG@GGGGB@>D<GGGF@<?<?9??;(?:2( AS:i:0     XN:i:0     XM:i:0
      XO:i:0     XG:i:0     NM:i:0     MD:Z:37    YT:Z:UU
SRR540192.1788   0     chr21 38025990   40    37M   *     0     0
      ATGGGCTTCCTCCGGCTTTCAGCCACCTGCGCCCTGC
      GG@G>G@E3<B=B;B<E>EDEAAAB:B.:=>A?;A8D AS:i:-5    XN:i:0     XM:i:1
      XO:i:0     XG:i:0     NM:i:1     MD:Z:26G10 YT:Z:UU
SRR540192.2271   0     chr21 41711175   42    37M   *     0     0
      TGATCATCTGGCTGATGCGGTGACTGCCACCCTTGAG
      IIGIIIIIIIIIIIIDIIIGIIIHGHHHIIIIHIHD AS:i:0     XN:i:0     XM:i:0
      XO:i:0     XG:i:0     NM:i:0     MD:Z:37    YT:Z:UU
```

File Post-Processing

The two SAM files produced were then processed with Samtools to:

- create faster, smaller, binary versions of the SAM files (BAM files)
- remove potential PCR duplicate reads
- sort each aligned sequence in the file by chromosome coordinate
- create an index of the BAM file
- create a summary of the mapping results

This was achieved for chip.sam with the following commands:

```
samtools view -Sb chip.sam > chip.bam
```

- outputs the data contained in the specified file
- -Sb: uses both the -S and -b options, which specify a SAM input file and produce a BAI index file, respectively
- chip.sam > chip.bam: the SAM file to view and the new file to write its contents to in binary format

```
samtools rmdup chip.bam chip.rmdup.bam
```

- removes duplicate alignments from chip.bam and outputs them to a new file chip.rmdup.bam

```
samtools sort chip.rmdup.bam chip.rmdup.sorted
```

- orders the contents of chip.rmdup.bam and writes them to a new bam file with the prefix chip.rmdup.sorted

```
samtools index chip.rmdup.sorted.bam
```

- creates a BAI index file of chip.rmdup.sorted.bam for quicker content access

```
samtools flagstat chip.rmdup.sorted.bam > chip_mappingstats.txt
```

- creates a text file containing summary data of the alignment

The same procedure was followed for input.sam. Mapping summary data are shown below. Note the high percentage of mapped reads, lack of QC-failed reads, and lack of duplicate reads.

From chipping_mappingstats.txt:

```
295896 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 duplicates
291344 + 0 mapped (98.46%:nan%)
```

From input_mappingstats.txt:

```
275043 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 duplicates
270767 + 0 mapped (98.45%:nan%)
```

Peak Calling

After the ChIP-Seq reads were aligned, peaks in the alignment were located and scored using MACS2. The control data in input.rmdup.sorted.bam served to control for noise in chip.rmdup.sorted.bam, which helped to identify significantly enriched alignment locations in the genome.

The following command was executed to run the MACS2 software:

```
macs2 callpeak -t chip.rmdup.sorted.bam -c input.rmdup.sorted.bam
     -f BAM -g hs -n macs_out --call-summits -B
```

- -t and -c: specify the treatment (ChIP) and control (input) files to use
- -f: specifies that BAM file format is being used
- -g: specifies that hs (homo sapiens) is the species whose genome is being used, so that statistical calculations include an accurate genome size
- -n: specifies that macs_out will be the prefix name for the output files produced
- --call-summits: specifies that subpeaks within peaks should try to be deconvolved to provide peak scores and positions for each subpeak
- -B: directs MACS2 to store additional information in bedGraph files for use with a genome browser

Following this peak calling, the MACS2 output file macs_out_peaks.xls provides information on alignment pile-ups, which hypothetically correspond to TF binding locations. From this file, the chromosome number, start position, and stop position were extracted to create a BED file for further use.

The command line language AWK and Unix utility sed were used for this purpose:

```
awk '!/^#|^$/ {print $1"\t"$2"\t"$3}' macs_out_peaks.xls | sed
'1d' > peaks.bed
```

- The AWK command above searches (/ /) for lines in the macs_out_peaks.xls file that do NOT (!) begin with (^) a hash sign (#) or (|) newline character ($)
    - This selects lines that do not begin with a comment and are not blank, removing the header region of this MACS2 output file
    - All that remains at this point are the data columns and their column names
- AWK is then commanded ({ }) to print columns 1 through 3 ($1, $2, $3) of each line separated by tab characters (“\t”)
    - This ensures that the file is tab separated and only contains the chromosome, start, and stop columns
- This output is then piped to sed, which deletes (d) the first (1) line of the text
    - This removes the line containing the column headings, leaving only the data in a BED format
- The output is then ultimately written to a file named peaks.bed

The complete list of peak locations, indicating possible TF binding sites, is provided at the end of this report.

Motif Search

To begin searching for motifs, the sequences under the peak regions were first obtained using peaks.bed as the coordinates to extract the sequences from a reference using bedtools' getfasta function:

```
bedtools getfasta -fi chr21.fa -bed peaks.bed -fo peaks.fasta
```

- -fi: specifies that chr21.fa is the reference genome from which to extract sequences
- -bed: specifies that peaks.bed is the BED file containing the sequence coordinates
- -fo: specifies that the extracted sequence should be written to peaks.fasta

An example sequence written to peaks.fasta is shown below:

```
>chr21:9478966-9479334
agtgcaagagtgaacacacactttgtttcggctttaaagaacccagcacgaagccagtctgcatggc
ctagagacatattttgctggataatgattacttgtttttctttttgtgtttgttgtatttgccggtt
tgcttagttcctgacatacaagaaaatcactgtcaaaacattagcttaacatttgtcaaggaaacaa
aaagacttcggtgaccacaccttataaagcaaacagttttgtaaatcactttggaaatttcagtaaa
aaaaaaaatccttaacaatataataagtaaagaaaatttaaaaccccaaaacattactgtgtttggg
ggggggggggttctgatttacagagtaaccaca
```

Once these sequences were obtained, a *de novo* motif search was performed across all peak sequences contained in peaks.fasta using MEME:

```
meme peaks.fasta -dna -mod zoops -minw 6 -maxw 26 -nmotifs 5
   -maxsize 750000 -o meme_out
```

- -dna: specifies that this is a DNA motif search using "ACGT"
- -mod zoops: specifies that Zero Or One (motif) Per Sequence should be searched for, rather than only one or multiple motifs per sequence
- -minw and -maxw: specifies that MEME should search for motifs of length 6 to 26 bases long
- -nmotifs: specifies that the search should stop once 5 motifs have been found
- -maxsize: sets the maximum allowable number of characters in the data to 750,000
    - Used because the default value of 100,000 is much too small for this ChIP dataset of ~650,000 characters
- -o: specifies the name of the directory to create that will hold the MEME output files

The 5 motif logos discovered by MEME are shown on the following page in order of descending E-value, which indicates the motifs' probability of random occurrence. While all 5 motifs show low E-values, motif 5 appears to have low conservation at each base overall, likely indicating less significance and confidence in the motif.

Motif 1

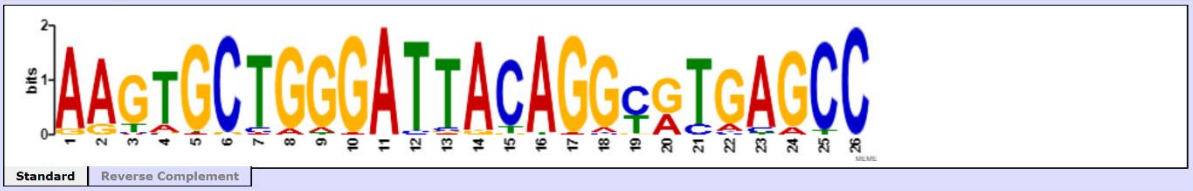
Summary	
E-value	9.5e-319
Width	26
Sites	52
show more	

Sequence Logo

Motif 2

Summary	
E-value	1.3e-253
Width	26
Sites	43
show more	

Sequence Logo



Motif 3

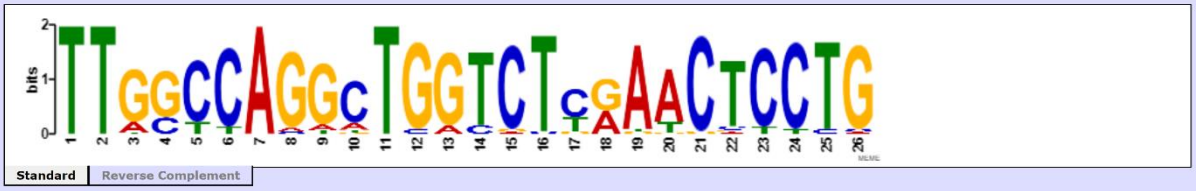
Summary	
E-value	2.2e-238
Width	26
Sites	43
show more	

Sequence Logo

Motif 4

Summary	
E-value	1.7e-215
Width	26
Sites	38
show more	

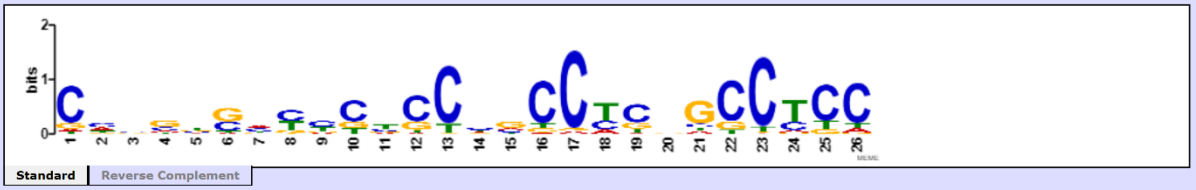
Sequence Logo



Motif 5

Summary	
E-value	3.4e-189
Width	26
Sites	252
show more	

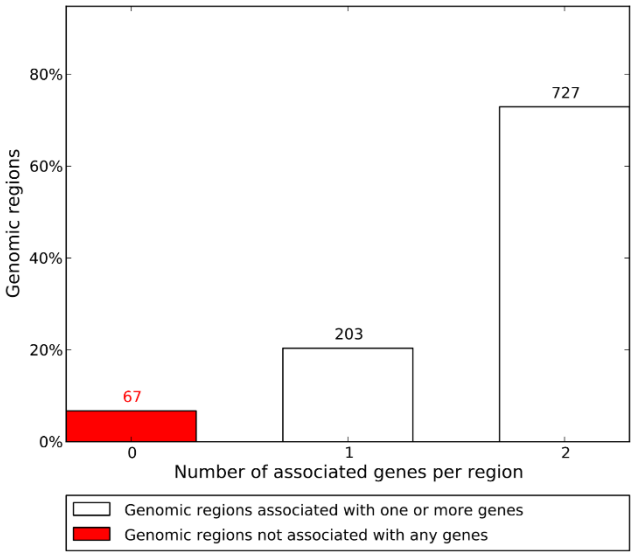
Sequence Logo



Peak Annotations

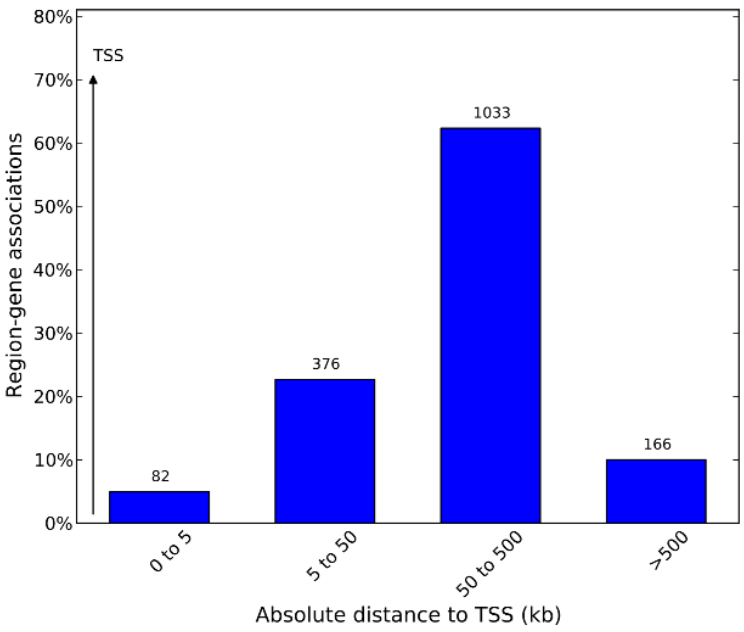
To explore the actual meaning and possible significance of the software outputs generated, the online GREAT tool was used to compile known associations between the 997 regions highlighted by this ChIP-Seq (now stored in peaks.bed) and annotated gene in online databases. In this way, cis-regulatory information can be gained by observing relationships between TF binding locations and the genes they may regulate.

The GREAT tool returns a summary of discovered associations across the human genome, also noting the distances from the submitted ChIP peak regions and the resultant associated annotated transcription start sites (TSS).



The graph at left shows the number of peak regions that have zero, one, or two found gene associations. Note that over 70% of peaks found in the ChIP-Seq experiment show two associated genes, while only ~6% of peaks found no associations. These sequences without gene associations may be indicative of non-specific TF binding, non-specific antibody binding, or may be novel associations that warrant further investigation.

The graph below shows the numbers of found gene associations grouped by distance from the peak region. Over 60% of found associations occurred from 50 to 500 kb away from the peak regions, while roughly 5% and 10% were found under 5 kb away and over 500 kb away, respectively. This shows that the bulk of gene associations were found to be within the expected range for a cis-regulatory element. In other words, the targeted transcription factor in this ChIP-Seq experiment appears to have high potential for gene regulatory function.



Motif Comparison

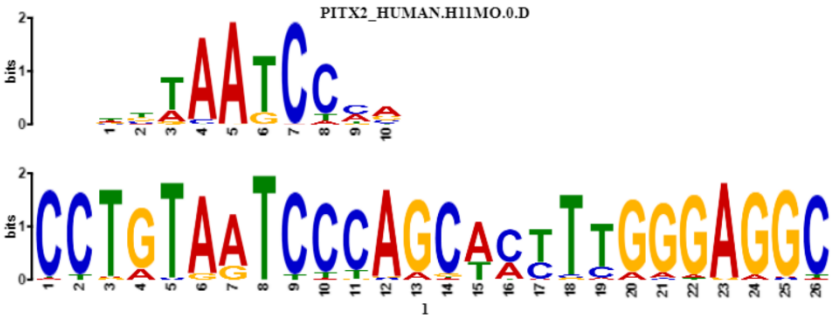
Finally, the online MEME-Suite tool Tomtom was used to compare the MEME motif results for this experiment with the HOCOMOCOv11 human and mouse motif databases. The general results for motifs 1 through 5 are shown below:

Database	ID	Matches	List
meme	1	10	PITX2_HUMAN.H11MO.0.D, PITX2_MOUSE.H11MO.0.D, IKZF1_HUMAN.H11MO.0.C, IKZF1_MOUSE.H11MO.0.C, ZN250_HUMAN.H11MO.0.C, TEAD4_MOUSE.H11MO.0.A, TEAD2_HUMAN.H11MO.0.D, TEAD2_MOUSE.H11MO.0.C,
meme	2	6	PITX2_HUMAN.H11MO.0.D, PITX2_MOUSE.H11MO.0.D, TBX5_HUMAN.H11MO.0.D, TBX5_MOUSE.H11MO.0.D, GFI1B_MOUSE.H11MO.0.A, TGIF2_HUMAN.H11MO.0.D
meme	3	47	RXRA_MOUSE.H11MO.0.A, RARA_MOUSE.H11MO.0.A, RXRG_MOUSE.H11MO.0.B, RARB_HUMAN.H11MO.0.D, RARB_MOUSE.H11MO.0.D, RXRA_MOUSE.H11MO.1.A, ERR2_MOUSE.H11MO.0.A, RARA_HUMAN.H11MO.0.A,
meme	4	24	ERR2_MOUSE.H11MO.0.A, GLI3_MOUSE.H11MO.0.D, GLI1_HUMAN.H11MO.0.D, GLI1_MOUSE.H11MO.0.C, RARA_MOUSE.H11MO.0.A, RXRA_MOUSE.H11MO.0.A, RXRG_MOUSE.H11MO.0.B, ZKSC3_HUMAN.H11MO.0.D,
meme	5	114	ZN770_HUMAN.H11MO.0.C, SP1_HUMAN.H11MO.0.A, ZN770_HUMAN.H11MO.1.C, ZFX_HUMAN.H11MO.1.A, MAZ_HUMAN.H11MO.0.A, MAZ_MOUSE.H11MO.0.A, SALL1_MOUSE.H11MO.0.D, PATZ1_HUMAN.H11MO.0.C,

From the results above, the PITX2 transcription factor shows matching motifs with both motif 1 in the reverse complement and with motif 2 in the forward direction, shown below. Without any information about the chosen antibody, its specificity, or the experimental conditions, this at least lends some credence to the idea that PITX2 could have been bound to this motif in the ChIP-Seq experiment.

Name	PITX2_HUMAN.H11MO.0.D
Database	HOCOMOCOv11 full HUMAN mono meme format
***p*-value**	1.21e-04
***E*-value**	1.57e-01
***q*-value**	1.57e-01
Overlap	10
Offset	-2
Orientation	Reverse Complement

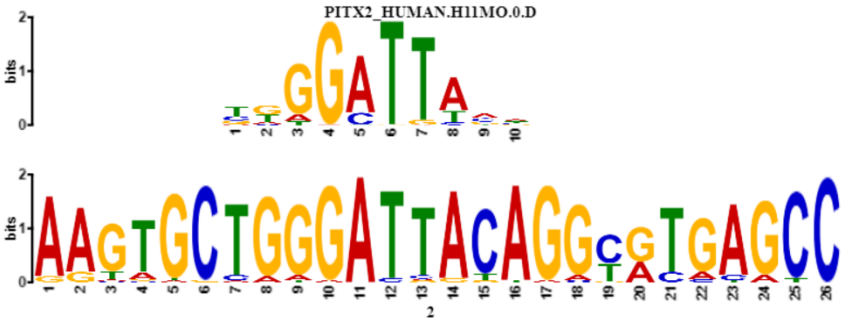
Show logo download options



1

Name	PITX2_HUMAN.H11MO.0.D
Database	HOCOMOCOv11 full HUMAN mono meme format
***p*-value**	1.03e-04
***E*-value**	1.34e-01
***q*-value**	1.34e-01
Overlap	10
Offset	-6
Orientation	Normal

Show logo download options



2

PITX2 may have implications for breast cancer, as it is involved in the production of prolactin, a hormone that is in turn involved in mammalian milk production and therefore breast function. However, PITX2 function occurs predominantly in the pituitary gland, and so any gene connection with actual breast cancer cell lines may not be causal, though it may warrant further investigation to identify connections with hormonal breast cancers.

Script Used

```
#!/bin/bash

# Job name:
#$ -N ChIPper

# The job should be placed into the queue 'all.q'
#$ -q all.q

# Running in the current directory
#$ -cwd

# Export some necessary environment variables
#$ -v PATH
#$ -v LD_LIBRARY_PATH
#$ -v PYTHONPATH
#$ -S /bin/bash

#Commands----------------------------

#Command-line arguments: qsub script.sh referencegenome chip_fastqfile input_fastqfile


#FastQC
for f in *.fastq;
do
            fastqc $f;
done


#MultiQC
multiqc .;


#Make Indexes from Reference Genome
bowtie2-build $1 ref_index
touch IndexingDone.txt

#Alignment and Post-Processing
for file in *.fastq;
do
      sam=${file//.fastq/.sam}
      bam=${file//.fastq/.bam}
      dups_out=${bam//.bam/.rmdup.bam}
      sorted=${dups_out//.rmdup.bam/.rmdup.sorted.bam}
      stats=${file//.fastq/_mappingstats.txt}


      bowtie2 -x ref_index -U $file -S $sam
      samtools view -Sb $sam > $bam
      samtools rmdup $bam $dups_out
      samtools sort $dups_out ${sorted//.bam}
      samtools index $sorted
      samtools flagstat $sorted > $stats

      touch ${file//.fastq/IsDone.txt}
done
```

```
#ChIP Peak Calling
macs2 callpeak -t ${2//.fastq/.rmdup.sorted.bam} -c ${3//.fastq/.rmdup.sorted.bam} -f BAM
-g hs -n macs_out --call-summits -B
touch MacsIsDone.txt

#MACS XLS Trimmer
awk '!/^#|^$/ {print $1"\t"$2"\t"$3}' macs_out_peaks.xls | sed '1d' > peaks.bed

#Extract ChIP Sequences
bedtools getfasta -fi $1 -bed peaks.bed -fo peaks.fasta
touch GetFastaIsDone.txt

#Motif Analysis
meme peaks.fasta -dna -mod zoops -minw 6 -maxw 26 -nmotifs 5 -maxsize 750000 -o meme_out
touch MemeIsDone.txt
```

Complete List of Called Peaks

Chrom.	Start	Stop						
chr21	9478966	9479334	chr21	16575731	16577694	chr21	18639918	18640461
chr21	9488140	9488479	chr21	16575731	16577694	chr21	18766256	18766644
chr21	9825338	9827051	chr21	16580008	16580812	chr21	18873626	18874207
chr21	9882928	9883227	chr21	16581191	16583209	chr21	18879800	18880330
chr21	10576105	10576614	chr21	16581191	16583209	chr21	18895076	18895811
chr21	10924251	10924815	chr21	16581191	16583209	chr21	18899458	18899880
chr21	11026478	11027107	chr21	16583796	16584411	chr21	18909528	18910222
chr21	11144289	11144781	chr21	16595649	16595983	chr21	18914568	18914972
chr21	14724142	14724644	chr21	16614146	16614735	chr21	18922015	18922674
chr21	14765889	14766285	chr21	16617123	16617693	chr21	18922886	18923615
chr21	14892260	14892828	chr21	16628157	16628800	chr21	19080671	19081191
chr21	14897414	14897810	chr21	16644462	16644914	chr21	19104561	19104944
chr21	14994757	14995125	chr21	16663099	16663599	chr21	19118285	19118790
chr21	15000171	15000534	chr21	16665462	16666001	chr21	19131500	19132563
chr21	15056140	15056531	chr21	16714586	16715837	chr21	19131500	19132563
chr21	15057037	15057801	chr21	16743092	16743787	chr21	20088933	20089484
chr21	15061144	15061820	chr21	16744927	16745826	chr21	20089949	20090568
chr21	15077114	15077870	chr21	16773584	16774204	chr21	20128490	20128829
chr21	15229316	15229800	chr21	16779056	16779831	chr21	20144014	20144899
chr21	15229943	15230480	chr21	16805250	16805828	chr21	20722489	20723313
chr21	15249176	15249828	chr21	16816381	16817076	chr21	20855099	20855558
chr21	15268764	15269356	chr21	16841350	16841940	chr21	20859581	20859897
chr21	15342255	15342582	chr21	16854666	16855157	chr21	20898090	20898389
chr21	15359675	15360343	chr21	16862408	16862712	chr21	20899963	20901179
chr21	15426954	15427494	chr21	16904172	16904576	chr21	20899963	20901179
chr21	15430644	15431447	chr21	16905872	16906429	chr21	20901229	20901918
chr21	15599037	15599773	chr21	16919580	16920706	chr21	20913687	20914433
chr21	15635425	15635927	chr21	16922056	16922582	chr21	20916800	20917099
chr21	15639394	15639864	chr21	16958985	16959896	chr21	20938661	20940018
chr21	15642778	15643481	chr21	16965215	16965923	chr21	20945614	20945964
chr21	15646038	15646486	chr21	16990291	16990822	chr21	20965867	20966508
chr21	15676269	15676951	chr21	17000774	17001788	chr21	20975382	20975836
chr21	15678285	15679069	chr21	17000774	17001788	chr21	20990359	20990950
chr21	15682861	15683319	chr21	17011123	17011863	chr21	20992973	20993634
chr21	15688582	15689393	chr21	17078957	17079426	chr21	20994109	20994936
chr21	15697334	15697658	chr21	17098332	17098655	chr21	21073914	21074354
chr21	15699832	15700191	chr21	17102651	17102950	chr21	21084815	21085436
chr21	15722522	15723085	chr21	17103585	17103934	chr21	21103636	21103935
chr21	15982557	15982918	chr21	17161368	17161925	chr21	21109259	21109617
chr21	16072075	16073377	chr21	17212875	17213248	chr21	21154706	21155400
chr21	16072075	16073377	chr21	17220980	17221359	chr21	21196616	21196937
chr21	16084239	16084960	chr21	17223959	17224283	chr21	21236669	21236996
chr21	16130110	16130515	chr21	17364434	17365039	chr21	21294104	21294771
chr21	16140360	16140758	chr21	17431052	17431659	chr21	21350196	21350881
chr21	16143702	16144283	chr21	17452016	17452746	chr21	21358457	21359003
chr21	16151407	16151759	chr21	17481125	17481424	chr21	21504256	21504594
chr21	16202491	16203499	chr21	17485760	17486068	chr21	21579805	21580146
chr21	16211958	16212325	chr21	17506994	17507748	chr21	21652939	21653433
chr21	16235472	16236167	chr21	17517867	17518352	chr21	21711152	21711462
chr21	16242795	16243620	chr21	17531333	17531841	chr21	21832291	21832676
chr21	16246845	16247155	chr21	17551908	17552596	chr21	21879583	21880122
chr21	16249577	16250638	chr21	17726530	17727234	chr21	21920541	21921129
chr21	16251333	16252184	chr21	17747011	17747513	chr21	21924090	21924652
chr21	16312983	16313505	chr21	17770601	17770928	chr21	21941632	21942062
chr21	16315738	16316335	chr21	17781785	17782087	chr21	21963536	21963949
chr21	16328356	16328744	chr21	17791378	17791941	chr21	21994503	21995168
chr21	16351520	16352034	chr21	17792292	17793030	chr21	22005296	22005724
chr21	16353552	16354019	chr21	17795550	17796087	chr21	22015206	22015787
chr21	16375524	16376254	chr21	17800737	17801294	chr21	22143090	22143389
chr21	16377461	16378230	chr21	17803414	17803990	chr21	22184418	22184871
chr21	16431222	16431696	chr21	17848562	17849187	chr21	22220455	22220949
chr21	16437709	16438271	chr21	17894810	17895246	chr21	22229202	22229621
chr21	16493768	16495146	chr21	18167417	18167752	chr21	22274639	22275575
chr21	16493768	16495146	chr21	18305602	18306143	chr21	22368867	22369806
chr21	16504937	16505621	chr21	18319106	18319716	chr21	22379776	22380360
chr21	16536903	16537405	chr21	18355688	18356236	chr21	22400253	22402082
chr21	16545038	16546169	chr21	18423326	18423885	chr21	22400253	22402082
chr21	16555061	16555461	chr21	18461348	18461834	chr21	22402459	22403144
chr21	16563132	16563670	chr21	18478180	18478768	chr21	22404797	22405148
chr21	16564451	16565024	chr21	18490711	18491392	chr21	22434854	22435890
chr21	16571541	16573414	chr21	18492594	18493173	chr21	22434854	22435890
chr21	16571541	16573414	chr21	18500486	18501145	chr21	22436016	22436708
chr21	16571541	16573414	chr21	18636252	18636769	chr21	22436980	22437393
chr21	16575731	16577694	chr21	18638706	18639224	chr21	22448981	22449535

chr21	22500365	22501028	chr21	25643973	25644363	chr21	32710234	32710770
chr21	22510427	22510726	chr21	26142541	26143067	chr21	32723634	32724349
chr21	22520276	22520867	chr21	26745347	26745893	chr21	32725179	32725507
chr21	22564809	22565469	chr21	26924997	26925629	chr21	32797604	32797977
chr21	22582702	22583249	chr21	26946155	26946856	chr21	32825640	32826074
chr21	22604934	22605358	chr21	27038693	27039238	chr21	32843635	32843967
chr21	22608144	22608787	chr21	27072747	27073559	chr21	32864695	32865289
chr21	22610261	22610854	chr21	27076173	27076486	chr21	32872687	32873030
chr21	22610917	22611756	chr21	27105292	27105764	chr21	32877290	32877994
chr21	22617868	22618305	chr21	27250571	27251077	chr21	32879713	32880268
chr21	22619665	22620305	chr21	27307252	27307638	chr21	32889805	32890365
chr21	22621145	22621569	chr21	27335508	27335998	chr21	32895690	32896248
chr21	22627829	22628498	chr21	27342795	27343362	chr21	32898189	32898868
chr21	22632461	22632959	chr21	27395011	27395499	chr21	32899579	32900109
chr21	22641075	22641557	chr21	27461239	27461765	chr21	32901539	32902261
chr21	22669390	22670011	chr21	27478933	27479445	chr21	32903164	32903962
chr21	22687456	22688155	chr21	27488576	27489200	chr21	32917203	32917745
chr21	22689662	22690216	chr21	27498115	27498522	chr21	32929729	32930858
chr21	22695443	22695979	chr21	27510694	27511455	chr21	32929729	32930858
chr21	22708142	22709199	chr21	27512321	27512954	chr21	32995465	32995814
chr21	22708142	22709199	chr21	27520487	27520792	chr21	33002337	33002716
chr21	22712855	22713393	chr21	27539169	27539593	chr21	33062091	33062978
chr21	22729991	22730685	chr21	27726227	27726569	chr21	33093006	33093306
chr21	22749437	22749914	chr21	27761028	27761687	chr21	33163235	33164002
chr21	22813937	22814313	chr21	27761964	27762962	chr21	33247251	33247582
chr21	22881195	22881810	chr21	27761964	27762962	chr21	33262838	33263315
chr21	23055513	23056030	chr21	27960119	27960914	chr21	33276446	33277092
chr21	23070096	23070524	chr21	28184485	28184848	chr21	33344848	33345498
chr21	23112541	23112840	chr21	28252477	28253026	chr21	33355324	33355915
chr21	23114923	23115494	chr21	28413150	28413449	chr21	33358842	33360534
chr21	23179289	23180305	chr21	29485953	29486294	chr21	33358842	33360534
chr21	23261106	23262197	chr21	29580102	29581353	chr21	33358842	33360534
chr21	23262760	23263249	chr21	29580102	29581353	chr21	33393990	33394467
chr21	23265021	23265363	chr21	29792094	29792552	chr21	33443144	33443477
chr21	23269439	23270415	chr21	29803159	29803686	chr21	33468808	33469292
chr21	23289813	23290132	chr21	29911231	29912039	chr21	33471968	33472278
chr21	23293579	23294993	chr21	30014046	30014529	chr21	33503490	33504091
chr21	23293579	23294993	chr21	30018444	30018811	chr21	33504234	33504760
chr21	23340890	23341612	chr21	30047663	30048091	chr21	33688023	33688372
chr21	23343407	23344211	chr21	30069039	30069978	chr21	33711863	33712502
chr21	23479209	23479939	chr21	30071505	30072402	chr21	33720889	33721454
chr21	23487979	23488479	chr21	30110108	30110470	chr21	33722466	33723137
chr21	23489904	23490263	chr21	30124318	30124868	chr21	33740572	33741196
chr21	23617320	23617813	chr21	30149692	30150208	chr21	33754304	33754805
chr21	23683852	23684479	chr21	30195190	30195641	chr21	33765508	33766196
chr21	23733909	23734426	chr21	30204228	30204836	chr21	33827890	33828374
chr21	23735678	23736168	chr21	30205402	30206038	chr21	33867204	33868223
chr21	23771887	23772905	chr21	30229973	30230629	chr21	33878733	33879333
chr21	23776041	23776484	chr21	30231032	30232383	chr21	33897888	33898187
chr21	23836824	23837506	chr21	30234261	30234823	chr21	33904605	33904953
chr21	23862289	23862921	chr21	30334625	30334946	chr21	34241514	34241855
chr21	23896335	23896672	chr21	30509423	30509997	chr21	34254820	34255119
chr21	23907587	23908418	chr21	30547647	30548528	chr21	34383686	34384202
chr21	23907587	23908418	chr21	30554191	30554556	chr21	34555577	34556241
chr21	23908724	23909450	chr21	30566674	30567014	chr21	34599569	34600137
chr21	23943007	23943577	chr21	30568657	30569150	chr21	34641299	34641844
chr21	23971827	23972142	chr21	30739456	30740053	chr21	34699880	34700406
chr21	24005349	24005648	chr21	30750746	30751514	chr21	34791841	34792247
chr21	24014280	24014579	chr21	30754238	30754803	chr21	34960924	34961421
chr21	24096726	24097615	chr21	30758306	30759594	chr21	35029290	35030441
chr21	24199416	24199843	chr21	30758306	30759594	chr21	35029290	35030441
chr21	24227884	24228315	chr21	30763002	30763478	chr21	35174107	35174463
chr21	24284795	24285319	chr21	30767879	30768403	chr21	35226294	35227428
chr21	24372216	24372746	chr21	30768711	30769213	chr21	35226294	35227428
chr21	24376255	24376819	chr21	30788474	30789049	chr21	35232161	35232555
chr21	24388599	24389327	chr21	30881356	30882287	chr21	35296859	35297483
chr21	24406377	24406877	chr21	31190834	31191320	chr21	35320018	35320327
chr21	24496163	24496575	chr21	31518869	31519168	chr21	35426435	35427083
chr21	24725360	24725812	chr21	31519420	31519719	chr21	35454837	35455349
chr21	24775275	24775880	chr21	31600046	31600417	chr21	35459118	35459756
chr21	24852392	24853680	chr21	32141943	32142503	chr21	35558280	35558966
chr21	24859300	24859662	chr21	32145187	32145663	chr21	35563048	35563468
chr21	25038248	25038755	chr21	32535022	32535471	chr21	35651374	35652028
chr21	25340918	25341436	chr21	32608212	32609027	chr21	35752700	35753159
chr21	25414916	25415455	chr21	32609405	32609854	chr21	35756824	35757645
chr21	25418395	25418705	chr21	32654542	32655105	chr21	35787957	35788468
chr21	25532792	25533384	chr21	32706993	32707508	chr21	35858292	35858751

chr21	35884417	35884858	chr21	37538883	37539394	chr21	39764262	39764915
chr21	35894477	35895096	chr21	37560182	37561174	chr21	39830243	39830737
chr21	35906749	35907274	chr21	37565397	37566090	chr21	39831450	39831993
chr21	35908255	35908585	chr21	37566319	37566870	chr21	39842742	39843041
chr21	35910988	35911371	chr21	37569395	37569942	chr21	39856956	39857290
chr21	35937524	35937847	chr21	37570132	37570706	chr21	39904872	39905274
chr21	35941614	35942127	chr21	37571203	37571788	chr21	39932945	39933598
chr21	36021069	36021461	chr21	37574452	37575054	chr21	39963597	39964099
chr21	36062601	36063308	chr21	37575845	37576173	chr21	39965082	39965723
chr21	36067626	36068282	chr21	37580490	37581063	chr21	39968231	39968544
chr21	36070956	36071255	chr21	37582437	37582847	chr21	39984599	39985196
chr21	36072897	36073355	chr21	37585150	37585631	chr21	40007613	40007920
chr21	36085806	36086298	chr21	37606300	37606700	chr21	40008446	40008745
chr21	36095075	36095640	chr21	37607237	37607939	chr21	40033022	40033669
chr21	36100567	36101512	chr21	37618328	37618970	chr21	40040340	40040843
chr21	36105910	36106294	chr21	37626402	37626701	chr21	40159352	40160053
chr21	36106810	36107601	chr21	37670185	37670495	chr21	40169407	40169824
chr21	36109129	36109915	chr21	37705465	37706040	chr21	40186879	40187205
chr21	36116012	36116683	chr21	37719531	37720011	chr21	40191538	40191996
chr21	36134087	36134686	chr21	37774236	37775103	chr21	40211668	40212073
chr21	36136375	36136899	chr21	37802183	37802765	chr21	40260711	40261027
chr21	36147041	36147483	chr21	37834346	37834899	chr21	40267189	40267640
chr21	36167967	36168383	chr21	37850405	37850892	chr21	40270202	40270638
chr21	36184533	36185235	chr21	37852422	37852751	chr21	40278836	40279473
chr21	36208804	36209367	chr21	37853960	37854309	chr21	40282238	40283085
chr21	36218004	36218487	chr21	37873522	37874002	chr21	40283367	40283666
chr21	36218687	36219220	chr21	37882275	37882852	chr21	40285232	40285648
chr21	36236246	36236739	chr21	37918289	37918945	chr21	40286431	40287088
chr21	36237621	36237926	chr21	37942053	37942533	chr21	40437727	40438302
chr21	36248002	36248422	chr21	37959735	37960545	chr21	40443685	40444077
chr21	36249619	36250218	chr21	37968416	37969191	chr21	40458934	40459253
chr21	36250665	36251349	chr21	37979844	37980143	chr21	40466909	40467447
chr21	36293594	36293905	chr21	37982440	37982998	chr21	40497915	40498240
chr21	36306561	36306957	chr21	37985137	37986343	chr21	40583184	40583664
chr21	36416394	36416837	chr21	38012619	38013077	chr21	40596512	40597178
chr21	36423618	36424094	chr21	38019923	38020473	chr21	40686434	40687018
chr21	36435927	36436242	chr21	38033809	38034307	chr21	40688048	40688530
chr21	36589615	36590428	chr21	38119658	38120036	chr21	40689089	40689733
chr21	36604815	36605530	chr21	38219170	38219811	chr21	40693202	40694453
chr21	36606017	36606352	chr21	38427580	38428187	chr21	40694799	40695228
chr21	36615448	36615930	chr21	38443952	38444816	chr21	40705377	40705972
chr21	36617783	36618333	chr21	38455158	38455486	chr21	40817090	40818084
chr21	36621690	36622243	chr21	38548934	38549373	chr21	40818877	40819353
chr21	36628840	36629338	chr21	38630294	38630789	chr21	40824496	40824818
chr21	36633954	36635359	chr21	38639564	38640682	chr21	40857038	40857337
chr21	36637927	36638242	chr21	38643242	38643665	chr21	40864966	40865373
chr21	36645584	36646086	chr21	38710104	38710403	chr21	40895434	40896390
chr21	36648078	36648377	chr21	38711203	38711891	chr21	40899492	40900411
chr21	36699517	36699828	chr21	38713760	38714059	chr21	40913839	40914247
chr21	36716182	36716645	chr21	38738223	38738524	chr21	40934559	40935191
chr21	36733491	36733859	chr21	38818703	38819314	chr21	40951543	40952129
chr21	36739777	36740169	chr21	38830752	38831495	chr21	40957266	40957773
chr21	36743863	36744384	chr21	38908223	38908654	chr21	40959433	40960235
chr21	36746950	36747531	chr21	38936479	38937868	chr21	40965177	40965899
chr21	36880676	36881000	chr21	38995874	38996334	chr21	40966425	40967050
chr21	36937491	36938161	chr21	39018604	39019117	chr21	40985684	40986095
chr21	36975711	36976151	chr21	39031375	39031743	chr21	40988268	40988801
chr21	36986720	36987337	chr21	39075834	39076794	chr21	41000255	41000879
chr21	37009776	37010075	chr21	39082849	39083209	chr21	41009671	41010097
chr21	37049587	37050372	chr21	39137513	39137889	chr21	41026376	41026910
chr21	37083124	37083429	chr21	39140177	39140844	chr21	41037488	41037864
chr21	37087552	37088004	chr21	39241746	39242140	chr21	41038621	41039464
chr21	37210645	37211196	chr21	39356014	39356539	chr21	41041237	41041592
chr21	37246350	37247181	chr21	39358866	39359740	chr21	41041860	41042165
chr21	37316165	37316464	chr21	39495091	39495880	chr21	41048085	41048427
chr21	37324363	37324939	chr21	39501668	39502162	chr21	41053683	41054187
chr21	37325762	37326612	chr21	39504100	39504501	chr21	41058497	41059050
chr21	37330662	37331274	chr21	39511672	39512888	chr21	41073975	41074481
chr21	37347160	37347632	chr21	39520243	39520555	chr21	41085078	41085384
chr21	37359686	37360330	chr21	39589076	39589575	chr21	41095269	41095825
chr21	37445332	37446050	chr21	39606649	39607146	chr21	41109507	41110147
chr21	37498480	37499071	chr21	39631544	39632023	chr21	41114264	41114888
chr21	37513773	37514227	chr21	39633724	39634255	chr21	41179100	41180094
chr21	37522627	37523139	chr21	39703258	39703557	chr21	41179100	41180094
chr21	37527787	37528355	chr21	39714661	39715281	chr21	41223911	41224734
chr21	37534871	37535213	chr21	39715415	39716189	chr21	41357057	41357676
chr21	37536309	37537140	chr21	39747137	39747590	chr21	41386473	41387082

chr21	41419253	41419591	chr21	43052125	43052587	chr21	44753139	44753690
chr21	41426027	41426794	chr21	43077000	43077561	chr21	44762667	44763777
chr21	41442365	41443496	chr21	43098245	43099024	chr21	44764024	44764885
chr21	41446493	41447027	chr21	43101217	43101929	chr21	44799600	44800204
chr21	41462098	41462693	chr21	43102507	43103145	chr21	44816557	44817176
chr21	41513641	41514112	chr21	43104557	43105067	chr21	44817896	44818726
chr21	41587873	41588192	chr21	43109168	43109574	chr21	44829185	44829828
chr21	41681954	41682743	chr21	43119714	43120369	chr21	44832160	44832741
chr21	41685199	41685890	chr21	43124262	43124891	chr21	44848783	44849296
chr21	41686037	41686872	chr21	43135933	43136393	chr21	44849865	44850344
chr21	41687033	41687362	chr21	43147445	43147954	chr21	44850641	44851367
chr21	41687492	41689096	chr21	43187156	43187653	chr21	44883281	44883795
chr21	41687492	41689096	chr21	43203441	43204148	chr21	44889163	44889512
chr21	41690247	41691174	chr21	43225094	43225652	chr21	44897669	44898342
chr21	41692536	41693540	chr21	43241429	43241851	chr21	44905914	44906507
chr21	41694609	41694983	chr21	43346482	43346906	chr21	44908454	44909065
chr21	41696101	41697368	chr21	43373835	43374350	chr21	44914014	44914401
chr21	41706052	41706407	chr21	43376135	43376741	chr21	44921716	44922088
chr21	41707545	41708086	chr21	43378135	43378632	chr21	44922252	44922789
chr21	41709332	41709823	chr21	43379304	43379902	chr21	44923068	44923424
chr21	41710846	41711542	chr21	43385012	43385481	chr21	44933057	44933722
chr21	41711832	41712828	chr21	43386225	43386856	chr21	44936221	44937243
chr21	41722866	41723427	chr21	43464966	43465398	chr21	44936221	44937243
chr21	41739671	41740256	chr21	43476310	43477854	chr21	44937354	44937883
chr21	41743658	41744004	chr21	43476310	43477854	chr21	44939593	44940277
chr21	41745596	41746507	chr21	43479933	43480657	chr21	44998386	44999339
chr21	41752042	41752742	chr21	43505897	43506802	chr21	44999878	45000332
chr21	41754257	41755270	chr21	43526371	43526740	chr21	45005791	45006368
chr21	41758877	41759447	chr21	43539984	43540862	chr21	45006908	45007500
chr21	41760001	41760633	chr21	43588112	43588770	chr21	45029505	45030169
chr21	41779624	41779946	chr21	43613377	43613954	chr21	45047268	45047686
chr21	41815934	41816507	chr21	43682765	43683446	chr21	45073852	45074209
chr21	41817069	41817811	chr21	43710962	43711305	chr21	45087328	45087822
chr21	41817959	41818525	chr21	43735217	43736085	chr21	45104721	45105199
chr21	41824990	41825569	chr21	43739479	43740141	chr21	45131947	45132918
chr21	41879942	41880430	chr21	43749431	43749766	chr21	45139570	45140053
chr21	41978234	41978768	chr21	43749893	43750453	chr21	45153967	45154460
chr21	41983447	41983804	chr21	43751017	43751563	chr21	45187925	45188294
chr21	42013750	42014301	chr21	43757993	43758601	chr21	45196020	45196661
chr21	42014847	42015388	chr21	43770961	43771593	chr21	45234265	45234775
chr21	42021975	42022612	chr21	43783255	43784260	chr21	45236532	45237151
chr21	42030016	42030638	chr21	43784505	43785121	chr21	45299645	45300494
chr21	42032325	42033120	chr21	43786267	43787344	chr21	45349224	45350017
chr21	42212373	42212784	chr21	43794171	43794723	chr21	45374887	45375295
chr21	42344773	42345072	chr21	43795357	43797527	chr21	45397168	45397639
chr21	42348154	42348949	chr21	43795357	43797527	chr21	45399596	45399944
chr21	42350401	42350952	chr21	43804187	43804703	chr21	45430188	45430764
chr21	42396016	42396678	chr21	43809541	43810049	chr21	45432530	45432845
chr21	42400910	42401249	chr21	43810740	43811770	chr21	45449580	45450114
chr21	42401440	42401860	chr21	43810740	43811770	chr21	45482943	45483489
chr21	42514869	42515407	chr21	43812746	43813275	chr21	45574314	45574842
chr21	42520720	42521559	chr21	43835683	43836095	chr21	45595790	45596456
chr21	42537042	42537686	chr21	43933516	43933966	chr21	45604915	45605315
chr21	42551897	42552406	chr21	43936118	43936851	chr21	45615726	45616920
chr21	42591325	42591808	chr21	43938316	43938823	chr21	45615726	45616920
chr21	42632627	42633144	chr21	43943258	43943717	chr21	45647326	45647977
chr21	42688275	42688845	chr21	43945164	43945700	chr21	45660371	45661262
chr21	42693933	42694956	chr21	43958577	43959469	chr21	45663614	45664343
chr21	42795602	42796073	chr21	43958577	43959469	chr21	45669003	45669533
chr21	42803531	42803988	chr21	43970971	43971364	chr21	45714629	45715093
chr21	42805425	42805959	chr21	44002141	44002633	chr21	45761932	45762492
chr21	42850265	42851011	chr21	44084132	44084699	chr21	45789890	45790811
chr21	42859863	42860599	chr21	44142701	44143168	chr21	45795730	45796069
chr21	42870163	42870660	chr21	44231043	44231408	chr21	45797914	45798459
chr21	42872368	42872805	chr21	44339805	44340400	chr21	45861691	45862056
chr21	42872978	42873413	chr21	44375601	44376292	chr21	46009856	46010219
chr21	42879864	42880176	chr21	44422643	44422947	chr21	46124065	46124671
chr21	42892412	42893041	chr21	44423051	44423590	chr21	46148578	46148926
chr21	42893335	42893932	chr21	44425687	44426232	chr21	46150813	46151270
chr21	42902147	42902748	chr21	44426535	44427552	chr21	46156431	46157013
chr21	42908192	42908827	chr21	44432727	44433083	chr21	46167656	46168313
chr21	42923062	42923493	chr21	44542281	44542832	chr21	46172778	46173337
chr21	42937363	42937780	chr21	44595082	44595410	chr21	46184480	46185152
chr21	42955079	42955794	chr21	44606087	44606451	chr21	46193146	46194177
chr21	43021714	43022505	chr21	44630353	44630956	chr21	46199645	46200018
chr21	43025492	43026265	chr21	44707568	44708691	chr21	46272838	46273253
chr21	43050396	43050795	chr21	44726875	44727390	chr21	46278543	46279090

chr21	46289719	46290883	chr21	46742670	46743023	chr21	47129721	47130241
chr21	46291710	46292644	chr21	46743728	46744027	chr21	47167725	47168173
chr21	46293702	46294276	chr21	46746912	46747705	chr21	47175941	47176418
chr21	46323039	46323694	chr21	46749412	46750816	chr21	47287853	47288940
chr21	46325509	46326174	chr21	46772587	46773230	chr21	47326122	47326469
chr21	46331395	46332325	chr21	46783955	46784263	chr21	47330326	47330840
chr21	46331395	46332325	chr21	46798264	46799426	chr21	47334690	47335626
chr21	46336701	46337161	chr21	46806108	46807276	chr21	47335981	47336553
chr21	46337885	46338421	chr21	46816479	46816866	chr21	47403845	47404347
chr21	46347314	46349053	chr21	46823035	46823666	chr21	47504756	47505604
chr21	46347314	46349053	chr21	46835825	46836800	chr21	47538046	47538378
chr21	46409744	46410456	chr21	46841310	46842095	chr21	47554116	47554670
chr21	46416372	46417016	chr21	46847119	46848206	chr21	47585049	47585405
chr21	46448207	46448737	chr21	46850854	46851371	chr21	47608145	47609004
chr21	46474292	46474933	chr21	46852811	46853329	chr21	47620660	47621755
chr21	46489859	46490405	chr21	46861719	46862071	chr21	47638203	47638904
chr21	46493024	46493888	chr21	46867621	46868152	chr21	47639718	47640149
chr21	46522472	46522970	chr21	46868845	46869231	chr21	47645998	47646457
chr21	46525690	46526531	chr21	46873957	46874464	chr21	47689211	47689819
chr21	46561415	46562073	chr21	46890855	46891401	chr21	47709581	47710177
chr21	46568031	46568571	chr21	46901839	46902327	chr21	47712247	47714831
chr21	46570638	46571277	chr21	46933144	46933873	chr21	47712247	47714831
chr21	46611399	46611899	chr21	46935562	46936030	chr21	47712247	47714831
chr21	46616752	46617067	chr21	46953662	46954298	chr21	47955799	47956158
chr21	46655843	46656253	chr21	47032109	47032516	chr21	48054324	48054680
chr21	46679974	46680427	chr21	47037042	47037526	chr21	48083144	48083446
chr21	46737467	46737778	chr21	47088044	47088569			
chr21	46740047	46740600	chr21	47125217	47125775			