iGenomX Sample Demultiplexing Tool

This documented describes how to use a command line tool for sample demultiplexing iGenomX data and is intended for individuals familiar with using a command line bioinformatics tool.

Table of Contents

Example Command Line

Below shows an example command line for a paired end 2x150bp experiment:

```
java -Xmx8G -jar fgbio.jar DemuxFastqs \
  --inputs example_S1_L001_R1_001.fastq.gz example_S1_L001_R2_001.fastq.gz \
  --metadata SampleSheet.csv \
  --read-structures 8B8M134T 8M142T \
  --output /path/to/output/directory \
  --metrics example.sample_barcode_metrics.txt
```

A BAM per sample will be written to the `path/to/output/directory` directory, including a BAM for reads that were not assigned a sample.
The `--output-fastqs true` option can be outputted to output FASTQs instead of BAMs.

Prerequisites

Java 8

Java 8 is required to run the command line tools.

Fgbio

The `fgbio` tool set can be downloaded from https://github.com/fulcrumgenomics/fgbio. Version 0.2.0 or higher is required; at the time of this writing, building from source is required.

To install from source, see Building Fgbio.

Please note that `git`, `java8`, the java JDK, `r-base`, `sbt`, and `scala` are required to build from source. The executable JAR can be found in `fgbio/target/scala-2.11/fgbio-*.jar` after building.

Fgbio is an open source tool developed by Fulcrum Genomics. Please visit the Fgbio project page to obtain support and report any bugs.

Running the Tool

The `DemuxFastqs` tool will be used within the `fgbio` set of tools. Please see the Prerequisites section for obtaining the tool set. To view the available commands, use `java -jar fgbio.jar --help`. To view the all options for the `DemuxFastqs` tool, use `java -jar fgbio.jar DemuxFastqs --help`.

Inputs

Input FASTQs

The input FASTQ(s) should be provided with the `--inputs` option. For paired end reads, the FASTQ corresponding to read one and read two should be specified in sequence: `--inputs <read_one>.fastq.gz <read_two>.fastq.gz`. Compressed FASTQs (gzipped: `fastq.gz`) are supported.

Read Structure

The read structure describes the logical structure or contents of the reads. A read structure per input FASTQ (i.e. input read) should be given using the `--read-structures` option. The read structure is the sequence of `(number, character)` segments (pairs) for a single read. The `character` describes the type of segment and the `number` describes the number of stretch bases of that type. For example, an 80bp read may a read structure of `10B8M2S60T` which describes four segments:

1. 10 bases describing a sample barcode (`10B`).
2. 8 bases describing a molecular barcode (`8M`).
3. 2 bases to be skipped (`2S`).
4. 60 bases of template (`60T`).

The allowable segment types are:

Segment Type	Character	Description
Sample barcode	`B`	The sample barcode bases I.e the section of the reads that identify to which sample or library the read belongs
Molecular barcode	`M`	The molecular barcode bases I.e the section of the reads that identify to which source molecule the read belongs. Frequently referred to as unique molecular identifiers (UMIs) or molecular identifiers (MIDs).
Template	`T`	The template bases. Traditionally “read 1” and “read 1” bases from the sample being sequenced.
Bases to skip	`S`	The bases to skip (omit from output). For example, monotemplate introduced during library preparation and so are not useful in analysis.

If more than one read contains a sample barcode, the tool extracts the sample barcodes from all reads and concatenates them together. The concatenated barcode is then matched against the barcode sequences in the Sample Metadata. The sample barcode bases for each read are concatenated, delimited by `-` , and stored in the `BC` SAM tag. Similarly, the molecular barcode bases for each read are concatenated, delimited by `-` , and stored in the `RX` SAM tag by default controlled by a command-line option.

Any bases described as skipped (`S`) are removed, and the remaining template bases (`T`) are used as the read’s bases (`SEQ` field in the SAM file).

In the example above (`10B8M2S60T`), the read structure applied to the following read:
`AAAAAAAAAACCCCCCCCGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT`

would have `AAAAAAAAAA` sample barcode bases stored in the `BC` SAM tag, `CCCCCCCC` molecular

barcode bases stored in the `RX` tag, and `TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT` template bases stored in the SEQ field. The two skipped bases (`GG`) would be not be outputted.

Sample Metadata

Metadata about the sample should be given in either a Sample Sheet or a Metadata CSV using the `--metadata` option respectively. The metadata includes for each sample: the sample identifier, sample name, library identifier, and expected sample barcode.

Sample Sheet

The Illumina Experiment Manager sample sheet can be used to describe the sample metadata with the `--metadata` option. The sample section (`[Data]`) of the sample sheet should contain information related to each sample with the following columns:

Column Name (case insensitive)	Description	Required
`Sample_ID`	The identifier of the sample unique within the sample sheet.	Yes
`Sample_Name`	The name of the sample unique within the sample sheet.	Yes
`Sample_Barcode`	The sample barcode bases (no delimiters).	Yes
`Library_ID`	The identifier for the library.	No
`Sample_Project`	The name of the project associated with the sample	No
`Description`	The free-text description of the sample	No

Additional columns may be specified but will not be used.

The expected sample barcode should be placed in the 'Sample*Barcode' column. If the sample barcode is present across multiple reads (ex. dual-index, or inline in both reads of a pair), then the expected barcode bases from each read should be concatenated with no delimiter and placed in the 'Sample*Barcode' column. The concatenation should be in the same order as the order of the reads' FASTQs and read structures given to this tool.

An Example Sample Sheet is shown in the Appendix.

Metadata CSV

In lieu of a Sample Sheet, a CSV file with just the sample section (`[Data]`) described in the Sample Sheet can be given.

An Example Metadata CSV is shown in the Appendix.

Outputs

The output directory should be specified with the `--output` option. The output directory will contain one BAM file per sample in the sample sheet, plus a BAM for reads that could not be assigned to a sample given the criteria.

The BAM file name for a sample will be the concatenation of sample id, sample name, and sample barcode bases (from the sample sheet), delimited by "-" (i.e. `<sample_id>-<sample_name>-<sample_barcode>.bam`). The BAM's read group will have sample id, sample name, and library id corresponding to the similarly named values in the sample sheet. The library id will be the sample id if not found, and the platform unit will be the sample name concatenated with the sample barcode bases delimited by a "." (i.e. `<sample_id>.<sample_barcode>.bam`). Additional command line options are available to specify additional metadata in the BAM's read group. The name for the unmatched sample is `unmatched` by default, but can be specified using the `--unmatched` option.

Alternatively, gzipped FASTQs can be written using the "–output-fastqs=true" option instead of BAMs. For paired end data, the output will have the suffix "*R1.fastq.gz" and "*R2.fastq.gz" for read one and read two respectively. The sample barcode and molecular barcodes (concatenated) will be appended to the read name and delimited by a colon.

A metrics file will also be output providing analogous information to the metric described here: SampleBarcodeMetric. Use the `--metrics` option to specify the path to where the metrics file should be written. An [Example Sample Barcode Metrics)[#example-sample-barcode-metrics] file is show in the Appendix.

Advanced Options

For a read to be assigned to a sample two criteria must be met:

1. The read's sample barcode must match the sample's barcode with `<= the maximum allowed number of mismatches` .
2. The read must not match any other sample's barcode with <= `mismatches to best barcode + minimum mismatch delta` .

For example, with `--max-mismatches=1` and `--min-delta=2`:

- If a sample is matched with a perfect match (no mismatches) all other samples' barcodes must be at least 2 mismatches away from the barcode sequence.
- If a sample is matched with a single mismatch, all other samples' barcodes must be at least 3 mismatches away from the barcode sequence.

Additional options are shown in the usage by specifying the `--help` option.

Appendix

Additional Recommended Best Practices

Adapter Marking

Care must be taken when adapter trimming or adapter marking. If the insert size (length of fragment being sequenced) is less than the number of bases sequence, both the inline sample barcode bases for the opposite read pair and the adapter sequenced may be present at the end of the given read. Alignment tools like `bwa` will soft-clip bases at the ends of reads that do not map well. Post-alignment tools like may be used to further ensure that the 3' end of any read does not extend past the 5' end of the opposite read (`MergeBamAlignment`: Picard) or to ensure that there are no overlapping mapped bases (`ClipOverlappingReads`: fgbio).

Specifying the adapter sequences to match, for example with Picard tool `MarkIlluminaAdapters` to include leading masked bases (`N`s) may be appropriate. Furthermore, using an alignment tool that can soft-clip the ends of reads (ex. `bwa`) can also be used.

PCR Duplicate Marking

It is recommended to use the molecular barcodes for each read when marking PCR duplicates, for example with the Picard tool `MarkDuplicates` by specifying the `BARCODE_TAG=RX` command line option. This allow duplicate marking to better discriminate true PCR duplicates versus those that mapped to the same location by chance, which can occur frequently when sequencing to deep coverage.

Converting BAMs to FASTQ

In some cases it is necessary to convert a BAM to FASTQ for downstream processing. The Picard tool `SamToFastq` is recommended for this conversion. Since the sample barcode AND molecular barcodes are stored in SAM tags, they will not be present in the converted FASTQs using this tool. This, if FASTQs are required for mapping (ex. `bwa`), it is recommended to use the Picard tool `MergeBamAlignment` to post-

process the mapped BAM and thus restore the various read-level metadata (ex. SAM tags and read groups) stored in the original demultiplexed BAM.

Alternatively, the “–output-fastqs=true” option can be used to write gzipped FASTQs instead of BAMs, but with the loss of sample metadata that can be stored in a read group in a BAM’s header.

Memory Usage

It is recommended to run the tool with at least 8GB of memory, which can be specified when running the tool: `java -Xmx8g -jar fgbio.jar ...`.

Example Sample Sheet

Below is an minimal example sample sheet for demultiplexing 96 samples.

```
[Header],,,,,,,
IEMFileVersion,4,,,,,,
Investigator Name,Joe,,,,,,
Experiment Name,EXPID,,,,,,
Date,1/1/00,,,,,,
Workflow,GenerateFASTQ,,,,,,
Application,FASTQ Only,,,,,,
Assay,Assay Name,,,,,,
Description,The Description,,,,,,
Chemistry,Amplicon,,,,,,
,,,,,,,
[Reads],,,,,,,
151,,,,,,,
151,,,,,,,
,,,,,,,
[Settings],,,,,,,
ReverseComplement,0,,,,,,
Adapter,AGATCGGAAGAGCACACGTCTGAACTCCAGTCA,,,,,,
AdapterRead2,AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT,,,,,,
,,,,,,,
[Data],,,,,,,
Sample_ID,Sample_Name,Library_Id,Sample_Plate,Sample_Well,index,Sample_Project,Descri
ption
ID-1,1,,,,AACCAAGG,,
ID-2,2,,,,AACGTTGC,,
ID-3,3,,,,AAGGTAGC,,
ID-4,4,,,,ACACGTGT,,
ID-5,5,,,,ACCAAGGA,,
```

```
ID-6,6,,,,ACCTGTTC,,
ID-7,7,,,,ACGTAGGA,,
ID-8,8,,,,ACTCGTGA,,
ID-9,9,,,,AGACAGTG,,
ID-10,10,,,,AGAGGTGT,,
ID-11,11,,,,AGCTAGGA,,
ID-12,12,,,,AGTCAGTC,,
ID-13,13,,,,AGTGGTCT,,
ID-14,14,,,,AGTGGTGA,,
ID-15,15,,,,ATCCTACC,,
ID-16,16,,,,ATCCTAGG,,
ID-17,17,,,,ATGCGCAT,,
ID-18,18,,,,ATGCGCTA,,
ID-19,19,,,,ATTAGCGC,,
ID-20,20,,,,ATTAGGCC,,
ID-21,21,,,,CAACGTTG,,
ID-22,22,,,,CAAGCTAC,,
ID-23,23,,,,CAAGCTTG,,
ID-24,24,,,,CACATGAC,,
ID-25,25,,,,CACATGTG,,
ID-26,26,,,,CAGACTCA,,
ID-27,27,,,,CAGACTGT,,
ID-28,28,,,,CAGTTGAC,,
ID-29,29,,,,CAGTTGTG,,
ID-30,30,,,,CATGCTAG,,
ID-31,31,,,,CATGCTTC,,
ID-32,32,,,,CCATATCC,,
ID-33,33,,,,CCATATGG,,
ID-34,34,,,,CCTACCAT,,
ID-35,35,,,,CCTACCTA,,
ID-36,36,,,,CGAACCAT,,
ID-37,37,,,,CGAACCTA,,
ID-38,38,,,,CGATTACG,,
ID-39,39,,,,CGATTAGC,,
ID-40,40,,,,CGTAGCAT,,
ID-41,41,,,,CGTAGCTA,,
ID-42,42,,,,CTACCAAG,,
ID-43,43,,,,CTACCATC,,
ID-44,44,,,,CTAGTCCT,,
ID-45,45,,,,CTAGTCGA,,
ID-46,46,,,,CTCTCACA,,
ID-47,47,,,,CTCTCAGT,,
ID-48,48,,,,CTGATCAG,,
ID-49,49,,,,CTGATCTC,,
ID-50,50,,,,CTTCCAAC,,
```

```
ID-51,51,,,,CTTCCATG,,
ID-52,52,,,,CTTGTCCA,,
ID-53,53,,,,CTTGTCGT,,
ID-54,54,,,,GAAGCAAC,,
ID-55,55,,,,GAAGCATG,,
ID-56,56,,,,GACATCAC,,
ID-57,57,,,,GACATCTG,,
ID-58,58,,,,GAGACACA,,
ID-59,59,,,,GAGACAGT,,
ID-60,60,,,,GAGTTCAC,,
ID-61,61,,,,GAGTTCTG,,
ID-62,62,,,,GATGCAAG,,
ID-63,63,,,,GATGCATC,,
ID-64,64,,,,GCATAACC,,
ID-65,65,,,,GCATAAGG,,
ID-66,66,,,,GCGCTATA,,
ID-67,67,,,,GCGCTTAA,,
ID-68,68,,,,GCTTCCAT,,
ID-69,69,,,,GCTTCCTA,,
ID-70,70,,,,GGATCCAT,,
ID-71,71,,,,GGATCCTA,,
ID-72,72,,,,GGTACGAA,,
ID-73,73,,,,GGTACGTT,,
ID-74,74,,,,GTACAGCT,,
ID-75,75,,,,GTACAGGA,,
ID-76,76,,,,GTAGGTAG,,
ID-77,77,,,,GTCTAGAC,,
ID-78,78,,,,GTGAGTCT,,
ID-79,79,,,,GTTCAGCA,,
ID-80,80,,,,GTTGGTAC,,
ID-81,81,,,,TACGAACC,,
ID-82,82,,,,TAGCTACG,,
ID-83,83,,,,TATTGCGG,,
ID-84,84,,,,TCAGCTCT,,
ID-85,85,,,,TCCATGCT,,
ID-86,86,,,,TCGACTAG,,
ID-87,87,,,,TCGTTGCT,,
ID-88,88,,,,TCTGCTCA,,
ID-89,89,,,,TGACTGAC,,
ID-90,90,,,,TGCACTAG,,
ID-91,91,,,,TGCTTGCT,,
ID-92,92,,,,TGGTCTAG,,
ID-93,93,,,,TGTCTGAG,,
ID-94,94,,,,TTATCCGC,,
ID-95,95,,,,TTCGGCAA,,
```

```
ID-96,96,,,,TTGGCCAA,,
```

Example Metadata CSV

Below is an minimal example metadta CSV file for demultiplexing 96 samples.

```
Sample_ID,Sample_Name,Library_Id,Sample_Plate,Sample_Well,index,Sample_Project,Descri
ption
ID-1,1,,,,AACCAAGG,,
ID-2,2,,,,AACGTTGC,,
ID-3,3,,,,AAGGTAGC,,
ID-4,4,,,,ACACGTGT,,
ID-5,5,,,,ACCAAGGA,,
ID-6,6,,,,ACCTGTTC,,
ID-7,7,,,,ACGTAGGA,,
ID-8,8,,,,ACTCGTGA,,
ID-9,9,,,,AGACAGTG,,
ID-10,10,,,,AGAGGTGT,,
ID-11,11,,,,AGCTAGGA,,
ID-12,12,,,,AGTCAGTC,,
ID-13,13,,,,AGTGGTCT,,
ID-14,14,,,,AGTGGTGA,,
ID-15,15,,,,ATCCTACC,,
ID-16,16,,,,ATCCTAGG,,
ID-17,17,,,,ATGCGCAT,,
ID-18,18,,,,ATGCGCTA,,
ID-19,19,,,,ATTAGCGC,,
ID-20,20,,,,ATTAGGCC,,
ID-21,21,,,,CAACGTTG,,
ID-22,22,,,,CAAGCTAC,,
ID-23,23,,,,CAAGCTTG,,
ID-24,24,,,,CACATGAC,,
ID-25,25,,,,CACATGTG,,
ID-26,26,,,,CAGACTCA,,
ID-27,27,,,,CAGACTGT,,
ID-28,28,,,,CAGTTGAC,,
ID-29,29,,,,CAGTTGTG,,
ID-30,30,,,,CATGCTAG,,
ID-31,31,,,,CATGCTTC,,
ID-32,32,,,,CCATATCC,,
ID-33,33,,,,CCATATGG,,
ID-34,34,,,,CCTACCAT,,
ID-35,35,,,,CCTACCTA,,
ID-36,36,,,,CGAACCAT,,
```

```
ID-37,37,,,,CGAACCTA,,
ID-38,38,,,,CGATTACG,,
ID-39,39,,,,CGATTAGC,,
ID-40,40,,,,CGTAGCAT,,
ID-41,41,,,,CGTAGCTA,,
ID-42,42,,,,CTACCAAG,,
ID-43,43,,,,CTACCATC,,
ID-44,44,,,,CTAGTCCT,,
ID-45,45,,,,CTAGTCGA,,
ID-46,46,,,,CTCTCACA,,
ID-47,47,,,,CTCTCAGT,,
ID-48,48,,,,CTGATCAG,,
ID-49,49,,,,CTGATCTC,,
ID-50,50,,,,CTTCCAAC,,
ID-51,51,,,,CTTCCATG,,
ID-52,52,,,,CTTGTCCA,,
ID-53,53,,,,CTTGTCGT,,
ID-54,54,,,,GAAGCAAC,,
ID-55,55,,,,GAAGCATG,,
ID-56,56,,,,GACATCAC,,
ID-57,57,,,,GACATCTG,,
ID-58,58,,,,GAGACACA,,
ID-59,59,,,,GAGACAGT,,
ID-60,60,,,,GAGTTCAC,,
ID-61,61,,,,GAGTTCTG,,
ID-62,62,,,,GATGCAAG,,
ID-63,63,,,,GATGCATC,,
ID-64,64,,,,GCATAACC,,
ID-65,65,,,,GCATAAGG,,
ID-66,66,,,,GCGCTATA,,
ID-67,67,,,,GCGCTTAA,,
ID-68,68,,,,GCTTCCAT,,
ID-69,69,,,,GCTTCCTA,,
ID-70,70,,,,GGATCCAT,,
ID-71,71,,,,GGATCCTA,,
ID-72,72,,,,GGTACGAA,,
ID-73,73,,,,GGTACGTT,,
ID-74,74,,,,GTACAGCT,,
ID-75,75,,,,GTACAGGA,,
ID-76,76,,,,GTAGGTAG,,
ID-77,77,,,,GTCTAGAC,,
ID-78,78,,,,GTGAGTCT,,
ID-79,79,,,,GTTCAGCA,,
ID-80,80,,,,GTTGGTAC,,
ID-81,81,,,,TACGAACC,,
```

```
ID-82,82,,,,TAGCTACG,,
ID-83,83,,,,TATTGCGG,,
ID-84,84,,,,TCAGCTCT,,
ID-85,85,,,,TCCATGCT,,
ID-86,86,,,,TCGACTAG,,
ID-87,87,,,,TCGTTGCT,,
ID-88,88,,,,TCTGCTCA,,
ID-89,89,,,,TGACTGAC,,
ID-90,90,,,,TGCACTAG,,
ID-91,91,,,,TGCTTGCT,,
ID-92,92,,,,TGGTCTAG,,
ID-93,93,,,,TGTCTGAG,,
ID-94,94,,,,TTATCCGC,,
ID-95,95,,,,TTCGGCAA,,
ID-96,96,,,,TTGGCCAA,,
```

Example Sample Barcode Metrics

Below is shown an example SampleBarcodeMetric metrics file.

```
barcode_name  library_name  barcode   reads   pf_reads  perfect_matches  pf_perfect_m
atches  one_mismatch_matches  pf_one_mismatch_matches  pct_matches  ratio_this_barcod
e_to_best_barcode_pct  pf_pct_matches  pf_ratio_this_barcode_to_best_barcode_pct  pf_
normalized_matches
1             1             AACCAAGG  288463  288463    283853           283853
        4610                  4610                     0.011726     0.290292
                      0.011726        0.290292                                  1.1
37428
2             2             AACGTTGC  211274  211274    208563           208563
        2711                  2711                     0.008588     0.212613
                      0.008588        0.212613                                  0.8
33067
3             3             AAGGTAGC  317510  317510    314589           314589
        2921                  2921                     0.012907     0.319523
                      0.012907        0.319523                                  1.2
51962
4             4             ACACGTGT  115454  115454    114650           114650
        804                   804                      0.004693     0.116186
                      0.004693        0.116186                                  0.4
55242
5             5             ACCAAGGA  14084   14084     13170            13170
        914                   914                      0.000573     0.014173
                      0.000573        0.014173                                  0.0
55534
```

```
6               6               ACCTGTTC  11772   11772     11586           11586
        186                     186                       0.000479     0.011847
                        0.000479        0.011847                                  0.0
46418
7               7               ACGTAGGA  9182    9182      9107            9107
        75                      75                        0.000373     0.00924
                        0.000373        0.00924                                   0.0
36205
8               8               ACTCGTGA  9446    9446      9390            9390
        56                      56                        0.000384     0.009506
                        0.000384        0.009506                                  0.0
37246
9               9               AGACAGTG  709179  709179    702664          702664
        6515                    6515                      0.028828     0.713675
                        0.028828        0.713675                                  2.7
96338
10              10              AGAGGTGT  596421  596421    593342          593342
        3079                    3079                      0.024245     0.600202
                        0.024245        0.600202                                  2.3
51726
11              11              AGCTAGGA  351888  351888    346499          346499
        5389                    5389                      0.014304     0.354119
                        0.014304        0.354119                                  1.3
87517
12              12              AGTCAGTC  93779   93779     92942           92942
        837                     837                       0.003812     0.094374
                        0.003812        0.094374                                  0.3
69777
13              13              AGTGGTCT  368605  368605    365969          365969
        2636                    2636                      0.014984     0.370942
                        0.014984        0.370942                                  1.4
53433
14              14              AGTGGTGA  187000  187000    185816          185816
        1184                    1184                      0.007602     0.188186
                        0.007602        0.188186                                  0.7
37353
15              15              ATCCTACC  592621  592621    588125          588125
        4496                    4496                      0.02409      0.596378
                        0.02409         0.596378                                  2.3
36742
16              16              ATCCTAGG  158134  158134    156775          156775
        1359                    1359                      0.006428     0.159137
                        0.006428        0.159137                                  0.6
23532
17              17              ATGCGCAT  405648  405648    401566          401566
```

```
        4082                    4082                        0.01649      0.40822
                        0.01649         0.40822                                     1.5
99496
18              18              ATGCGCTA  329727  329727    326387           326387
        3340                    3340                        0.013403     0.331817
                        0.013403        0.331817                                    1.3
00134
19              19              ATTAGCGC  993700  993700    986891           986891
        6809                    6809                        0.040394     1
                        0.040394        1                                           3.9
18222
20              20              ATTAGGCC  21223   21223     20057            20057
        1166                    1166                        0.000863     0.021358
                        0.000863        0.021358                                    0.0
83684
21              21              CAACGTTG  379670  379670    376922           376922
        2748                    2748                        0.015434     0.382077
                        0.015434        0.382077                                    1.4
97063
22              22              CAAGCTAC  373005  373005    371689           371689
        1316                    1316                        0.015163     0.37537
                        0.015163        0.37537                                     1.4
70782
23              23              CAAGCTTG  10444   10444     10425            10425
        19                      19                          0.000425     0.01051
                        0.000425        0.01051                                     0.0
41181
24              24              CACATGAC  338900  338900    338429           338429
        471                     471                         0.013776     0.341049
                        0.013776        0.341049                                    1.3
36304
25              25              CACATGTG  13034   13034     13003            13003
        31                      31                          0.00053      0.013117
                        0.00053         0.013117                                    0.0
51394
26              26              CAGACTCA  250271  250271    249225           249225
        1046                    1046                        0.010174     0.251858
                        0.010174        0.251858                                    0.9
86834
27              27              CAGACTGT  7349    7349      7314             7314
        35                      35                          0.000299     0.007396
                        0.000299        0.007396                                    0.0
28978
28              28              CAGTTGAC  155718  155718    155431           155431
        287                     287                         0.00633      0.156705
```

```
                    0.00633         0.156705                                0.6
14006
29          29          CAGTTGTG  552715  552715    551447          551447
       1268                1268                    0.022468    0.556219
                    0.022468        0.556219                                2.1
7939
30          30          CATGCTAG  15770   15770     15635           15635
       135                 135                     0.000641    0.01587
                    0.000641        0.01587                                 0.0
62182
31          31          CATGCTTC  356315  356315    355219          355219
       1096                1096                    0.014484    0.358574
                    0.014484        0.358574                                1.4
04973
32          32          CCATATCC  15914   15914     15855           15855
       59                  59                      0.000647    0.016015
                    0.000647        0.016015                                0.0
6275
33          33          CCATATGG  395536  395536    393799          393799
       1737                1737                    0.016079    0.398044
                    0.016079        0.398044                                1.5
59623
34          34          CCTACCAT  10567   10567     9925            9925
       642                 642                     0.00043     0.010634
                    0.00043         0.010634                                0.0
41666
35          35          CCTACCTA  281861  281861    281378          281378
       483                 483                     0.011458    0.283648
                    0.011458        0.283648                                1.1
11396
36          36          CGAACCAT  14062   14062     13879           13879
       183                 183                     0.000572    0.014151
                    0.000572        0.014151                                0.0
55447
37          37          CGAACCTA  262284  262284    261907          261907
       377                 377                     0.010662    0.263947
                    0.010662        0.263947                                1.0
34202
38          38          CGATTACG  619066  619066    614034          614034
       5032                5032                    0.025165    0.622991
                    0.025165        0.622991                                2.4
41016
39          39          CGATTAGC  365763  365763    363225          363225
       2538                2538                    0.014868    0.368082
                    0.014868        0.368082                                1.4
```

```
42227
40              40              CGTAGCAT  479541  479541    477891          477891
        1650                    1650                      0.019493    0.482581
                        0.019493        0.482581                                  1.8
90861
41              41              CGTAGCTA  210351  210351    209413          209413
        938                     938                       0.008551    0.211685
                        0.008551        0.211685                                  0.8
29427
42              42              CTACCAAG  538000  538000    534061          534061
        3939                    3939                      0.02187     0.541411
                        0.02187         0.541411                                  2.1
21368
43              43              CTACCATC  239773  239773    238159          238159
        1614                    1614                      0.009747    0.241293
                        0.009747        0.241293                                  0.9
4544
44              44              CTAGTCCT  283812  283812    282068          282068
        1744                    1744                      0.011537    0.285611
                        0.011537        0.285611                                  1.1
19089
45              45              CTAGTCGA  10248   10248     10156           10156
        92                      92                        0.000417    0.010313
                        0.000417        0.010313                                  0.0
40409
46              46              CTCTCACA  247248  247248    245467          245467
        1781                    1781                      0.010051    0.248816
                        0.010051        0.248816                                  0.9
74915
47              47              CTCTCAGT  9679    9679      9500            9500
        179                     179                       0.000393    0.00974
                        0.000393        0.00974                                   0.0
38165
48              48              CTGATCAG  137402  137402    136057          136057
        1345                    1345                      0.005585    0.138273
                        0.005585        0.138273                                  0.5
41785
49              49              CTGATCTC  10283   10283     10193           10193
        90                      90                        0.000418    0.010348
                        0.000418        0.010348                                  0.0
40547
50              50              CTTCCAAC  302701  302701    300833          300833
        1868                    1868                      0.012305    0.30462
                        0.012305        0.30462                                   1.1
93569
```

```
51          51          CTTCCATG  9826    9826     9566          9566
        260                 260                  0.000399    0.009888
                  0.000399       0.009888                              0.0
38745
52          52          CTTGTCCA  235932  235932   233301        233301
        2631                2631                 0.009591    0.237428
                  0.009591       0.237428                              0.9
30295
53          53          CTTGTCGT  212940  212940   211585        211585
        1355                1355                 0.008656    0.21429
                  0.008656       0.21429                               0.8
39636
54          54          GAAGCAAC  17216   17216    17146         17146
        70                  70                   0.0007      0.017325
                  0.0007         0.017325                              0.0
67884
55          55          GAAGCATG  725654  725654   722869        722869
        2785                2785                 0.029498    0.730255
                  0.029498       0.730255                              2.8
613
56          56          GACATCAC  17280   17280    17192         17192
        88                  88                   0.000702    0.01739
                  0.000702       0.01739                               0.0
68136
57          57          GACATCTG  599177  599177   598211        598211
        966                 966                  0.024357    0.602976
                  0.024357       0.602976                              2.3
62593
58          58          GAGACACA  16878   16878    16783         16783
        95                  95                   0.000686    0.016985
                  0.000686       0.016985                              0.0
66551
59          59          GAGACAGT  588025  588025   585325        585325
        2700                2700                 0.023903    0.591753
                  0.023903       0.591753                              2.3
1862
60          60          GAGTTCAC  16337   16337    16237         16237
        100                 100                  0.000664    0.016441
                  0.000664       0.016441                              0.0
64418
61          61          GAGTTCTG  176093  176093   175633        175633
        460                 460                  0.007158    0.177209
                  0.007158       0.177209                              0.6
94346
62          62          GATGCAAG  741904  741904   738970        738970
```

```
        2934                    2934                        0.030158     0.746608
                        0.030158        0.746608                                    2.9
25374
63              63              GATGCATC  138194  138194    137521           137521
        673                     673                         0.005618     0.13907
                        0.005618        0.13907                                     0.5
44908
64              64              GCATAACC  708525  708525    706272           706272
        2253                    2253                        0.028802     0.713017
                        0.028802        0.713017                                    2.7
93759
65              65              GCATAAGG  354180  354180    352832           352832
        1348                    1348                        0.014397     0.356425
                        0.014397        0.356425                                    1.3
96554
66              66              GCGCTATA  475894  475894    470844           470844
        5050                    5050                        0.019345     0.478911
                        0.019345        0.478911                                    1.8
7648
67              67              GCGCTTAA  453613  453613    448494           448494
        5119                    5119                        0.018439     0.456489
                        0.018439        0.456489                                    1.7
88625
68              68              GCTTCCAT  377381  377381    376226           376226
        1155                    1155                        0.015341     0.379774
                        0.015341        0.379774                                    1.4
88037
69              69              GCTTCCTA  9025    9025      8851             8851
        174                     174                         0.000367     0.009082
                        0.000367        0.009082                                    0.0
35586
70              70              GGATCCAT  129093  129093    128657           128657
        436                     436                         0.005248     0.129911
                        0.005248        0.129911                                    0.5
09022
71              71              GGATCCTA  7676    7676      7573             7573
        103                     103                         0.000312     0.007725
                        0.000312        0.007725                                    0.0
30267
72              72              GGTACGAA  124991  124991    123288           123288
        1703                    1703                        0.005081     0.125783
                        0.005081        0.125783                                    0.4
92847
73              73              GGTACGTT  8126    8126      8029             8029
        97                      97                          0.00033      0.008178
```

```
                    0.00033         0.008178                                0.0
32041
74          74          GTACAGCT  211000  211000    209308          209308
       1692                1692                  0.008577    0.212338
                    0.008577        0.212338                                0.8
31986
75          75          GTACAGGA  14573   14573     14419           14419
       154                 154                   0.000592    0.014665
                    0.000592        0.014665                                0.0
57462
76          76          GTAGGTAG  99705   99705     98679           98679
       1026                1026                  0.004053    0.100337
                    0.004053        0.100337                                0.3
93143
77          77          GTCTAGAC  10939   10939     10816           10816
       123                 123                   0.000445    0.011008
                    0.000445        0.011008                                0.0
43133
78          78          GTGAGTCT  9840    9840      9678            9678
       162                 162                   0.0004      0.009902
                    0.0004          0.009902                                0.0
388
79          79          GTTCAGCA  16317   16317     16164           16164
       153                 153                   0.000663    0.01642
                    0.000663        0.01642                                 0.0
64339
80          80          GTTGGTAC  9392    9392      9278            9278
       114                 114                   0.000382    0.009452
                    0.000382        0.009452                                0.0
37033
81          81          TACGAACC  640623  640623    634304          634304
       6319                6319                  0.026041    0.644685
                    0.026041        0.644685                                2.5
26017
82          82          TAGCTACG  509337  509337    504551          504551
       4786                4786                  0.020705    0.512566
                    0.020705        0.512566                                2.0
08348
83          83          TATTGCGG  636050  636050    629482          629482
       6568                6568                  0.025856    0.640083
                    0.025856        0.640083                                2.5
07985
84          84          TCAGCTCT  286300  286300    283936          283936
       2364                2364                  0.011638    0.288115
                    0.011638        0.288115                                1.1
```

```
28899
85              85              TCCATGCT  236240  236240    234395          234395
        1845                    1845                      0.009603    0.237738
                        0.009603        0.237738                                  0.9
31509
86              86              TCGACTAG  138847  138847    138133          138133
        714                     714                       0.005644    0.139727
                        0.005644        0.139727                                  0.5
47483
87              87              TCGTTGCT  153178  153178    151958          151958
        1220                    1220                      0.006227    0.154149
                        0.006227        0.154149                                  0.6
03991
88              88              TCTGCTCA  109759  109759    108954          108954
        805                     805                       0.004462    0.110455
                        0.004462        0.110455                                  0.4
32787
89              89              TGACTGAC  13248   13248     13154           13154
        94                      94                        0.000539    0.013332
                        0.000539        0.013332                                  0.0
52238
90              90              TGCACTAG  12677   12677     12612           12612
        65                      65                        0.000515    0.012757
                        0.000515        0.012757                                  0.0
49986
91              91              TGCTTGCT  11311   11311     11217           11217
        94                      94                        0.00046     0.011383
                        0.00046         0.011383                                  0.0
446
92              92              TGGTCTAG  13347   13347     13225           13225
        122                     122                       0.000543    0.013432
                        0.000543        0.013432                                  0.0
52628
93              93              TGTCTGAG  429981  429981    427101          427101
        2880                    2880                      0.017479    0.432707
                        0.017479        0.432707                                  1.6
95442
94              94              TTATCCGC  845035  845035    838036          838036
        6999                    6999                      0.034351    0.850392
                        0.034351        0.850392                                  3.3
32026
95              95              TTCGGCAA  491200  491200    487103          487103
        4097                    4097                      0.019967    0.494314
                        0.019967        0.494314                                  1.9
36833
```

```
96            96            TTGGCCAA  359032  359032    354958           354958
        4074                    4074                    0.014595     0.361308
                      0.014595        0.361308                                  1.4
15686
unmatched     unmatched     NNNNNNNN  525880  525880    0                0
        0                       0                       0.021377     0.529214
                      0.021377        0.529214                                  2.0
73578
```