ChIA-PET Tool package output illustration

ChIA-PET Tool package output illustration

This document explains all the output files, including those used in the visualization report. Figure 1 shows the structure of the DNA constructs generated by ChIA-PET protocol, which is labeled with the terms used in this document. Figure 2 shows the result files along the ChIA-PET data processing steps.

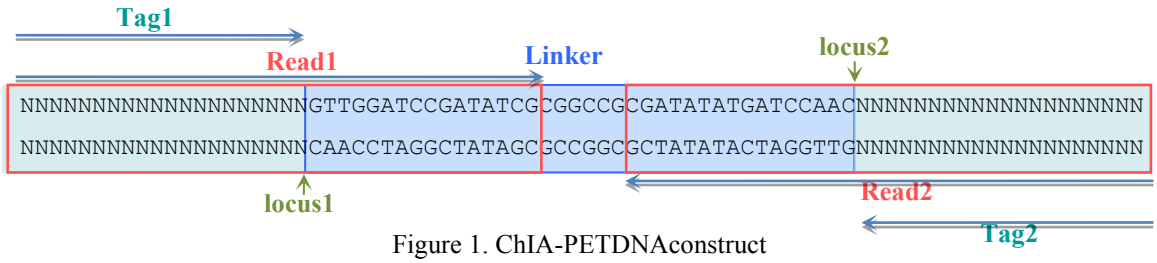

Figure 1. ChIA-PETDNAconstruct

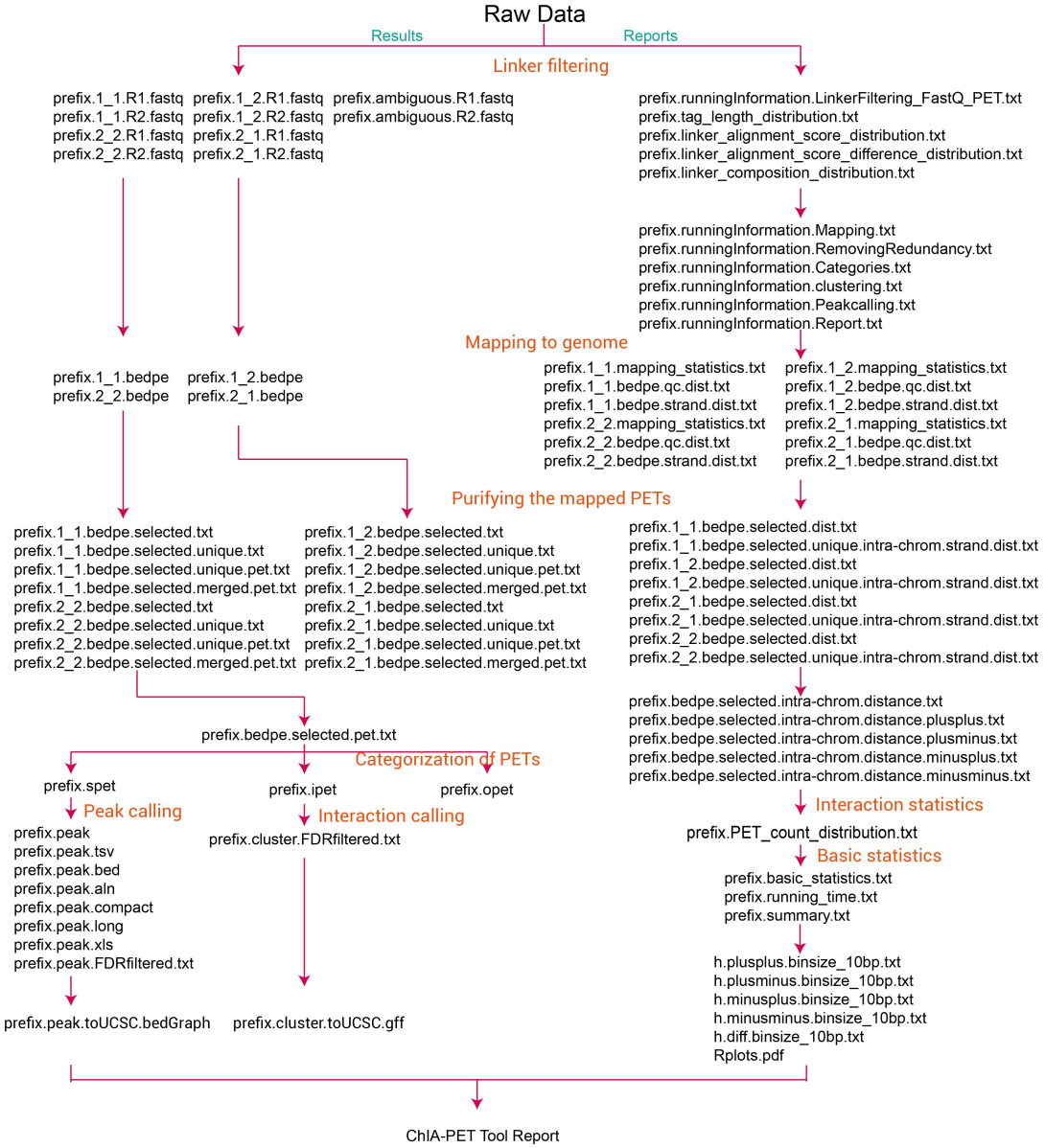

Figure 2. Result files along the ChIA-PET data processing steps

In this file, ChIA-PET data associated with RNA polymerase II (RNAPII) from human breast cancer cell line MCF7 is used for illustration, and "PREFIX" is used as the prefix of the output files. Table 1 contains the file list generated by ChIA-PET Tool, and Table 2 contains the file list used in the visualization report. The files in Table 1 are grouped by the ChIA-PET data processing steps. The formats of the files are elaborated below.

Table 1. Files in the output directory

Processing step	File name	Description
Linker Filtering	PREFIX.1_1.R1.fastq	Fastq file of read1, both reads aligned to linker A
	PREFIX.1_1.R2.fastq	Fastq file of read2, both reads aligned to linker A
	PREFIX.1_2.R1.fastq	Fastq file of read1, read1 and read2 aligned to linker A and linker B respectively
	PREFIX.1_2.R2.fastq	Fastq file of read2, read1 and read2 aligned to linker A and linker B respectively
	PREFIX.2_1.R1.fastq	Fastq file of read1, read1 and read2 aligned to linker B and linker A respectively
	PREFIX.2_1.R2.fastq	Fastq file of read2, read1 and read2 aligned to linker B and linker A respectively
	PREFIX.2_2.R1.fastq	Fastq file of read1, both reads aligned to linker B
	PREFIX.2_2.R2.fastq	Fastq file of read2, both reads aligned to linker B
	PREFIX.ambiguous.R1.fastq	Fastq file of read1, with ambiguous linker information
	PREFIX.ambiguous.R2.fastq	Fastq file of read2, with ambiguous linker information
Mapping to Genome	PREFIX.1_1.bedpe	A bedpe format file of unique mapping PETs converted by BEDtools from mapped BAM file
Purifying the mapped reads	PREFIX.1_1.bedpe.selected.txt	File with selected PETs. The mapping scores of both tags are above the mapping quality score threshold, and the optional fields contain the annotation tags: XT:A:U, X0:i:1 and X1:i:0.
	PREFIX.1_1.bedpe.selected.unique.txt	PETs file after merging PETs that are mapped to the same genomic positions
	PREFIX.1_1.bedpe.selected.unique.pet.txt	In this file， the mapping result of a tag is simplified into the format: chromosome, end or start (the genome location that is close to the linker sequence), strand, and sort by chromosome and coordinates.
	PREFIX.1_1.bedpe.selected.merged.pet.txt	PETs file after merging the similarly mapped PETs into one PET
	PREFIX.bedpe.selected.pet.txt	PETs file after merging both same-linker PET files
PET classification	PREFIX.ipet	File for inter-ligation PETs, including inter-, intra-chromosomal inter-ligation PETs

	PREFIX.spet	File for self-ligation PETs
	PREFIX.opet	File for other PETs with short distance
Interaction calling	PREFIX.cluster.FDRfiltered.txt	File with statistically-significant chromatin interactions
Peak calling	PREFIX.peak	File with peaks
	PREFIX.peaks.FDRfiltered.txt	File with statistically-significant peaks
Visualization	PREFIX.peak.toUCSC.bedGraph	Peak file which could be visualized with UCSC browser
	PREFIX.cluster.toUCSC.gff	Cluster file which could be visualized with UCSC browser

Table 2. Files used in the visualization report

File name	Description
PREFIX.tag_length_distribution.txt	Distribution of tag length
PREFIX.linker_alignment_score_distribution.txt	Distribution of best alignment scores from the designed linker sequences to the reads
PREFIX.linker_alignment_score_difference_distribution.txt	Distribution of score differences between the best-aligned linker and the second-best aligned linker
PREFIX.linker_composition_distribution.txt	Distribution of linker composition
PREFIX.runningInformation.LinkerFiltering_FastQ_PET.txt	Running information of Linker filtering
PREFIX.runningInformation.Mapping.txt	Running information of Mapping to genome
PREFIX.runningInformation.RemovingRedundancy.txt	Running information of Removing redundancy
PREFIX.runningInformation.Categories.txt	Running information of Classification
PREFIX.runningInformation.clustering.txt	Running information of Peak calling
PREFIX.runningInformation.Peakcalling.txt	Running information of Clustering
PREFIX.runningInformation.Report.txt	Running information of Generating reports
PREFIX.1_1.mapping_statistics.txt	Statistics of mapping mode (Non-mappable, Uniquely-mapped and Others)
PREFIX.1_1.bedpe.qc.dist.txt	Distribution of mapping quality scores
PREFIX.1_1.bedpe.strand.dist.txt	Distribution of strands information
PREFIX.1_1.bedpe.selected.dist.txt	Distribution of counts of PETs that are mapped to the same genomic position
PREFIX.1_1.bedpe.selected.unique.intra-chrom.strand.dist.txt	Distribution of strands information after removing redundancy
PREFIX.bedpe.selected.intra-chrom.distance.txt	Span and strands information of intra-chromosomal PETs
PREFIX.bedpe.selected.intra-chrom.distance.plusplus.txt	Span of PETs with both tags mapped on plus strand in intra-chromosomal PETs

PREFIX.bedpe.selected.intra-chrom.distance.plusminus.txt	Span of PETs with two tags in intra-chromosomal PETs mapped on plus and minus strand respectively
PREFIX.bedpe.selected.intra-chrom.distance.minusplus.txt	Span of PETs with two tags in intra-chromosomal PETs mapped on minus and plus strand respectively
PREFIX.bedpe.selected.intra-chrom.distance.minusminus.txt	Span of PETs with both tags in intra-chromosomal PETs mapped on minus strand in intra-chromosomal PETs
PREFIX.PET_count_distribution.txt	Statistics file of PET counts' distribution
PREFIX.basic_statistics.txt	Basic statistics of PETs number
PREFIX.running_time.txt	The running time of each section
ChIA-PET_Tool_Report	A html report containingstatistics of information during data processing with ChIA-PET Tool

Files in the output directory:

(1) PREFIX.1_1.bedpe (These explanation is from http://bedtools.readthedocs.org/en/latest/content/general-usage.html)

This file is used to illustrate the information of pair-end tag sequence mapped to genome. This file is getting from unique-mapping PETs.

chrom1	start1	end1	chrom2	start2	end2	name	score	strand1	strand2
chr5	175964094	175964115	chr5	175964271	175964291	SRR372741.25	25	-	+
chr1	166569355	166569376	chr12	56223587	56223608	SRR372741.36	37	+	+
chr14	71969478	71969499	chr19	16530595	16530616	SRR372741.94	37	-	-
chr10	108878463	108878484	chr2	184995039	184995059	SRR372741.143	37	-	-
chr10	29904926	29904946	chr11	32678927	32678947	SRR372741.167	37	-	+
chr9	127735487	127735507	chr9	127736244	127736264	SRR372741.176	37	-	+
chr2	191497466	191497487	chr4	83532329	83532350	SRR372741.241	37	+	-
chr17	79680058	79680079	chr17	79680152	79680172	SRR372741.247	37	-	+
chr16	10275488	10275508	chr6	150597540	150597561	SRR372741.251	37	-	+
chr1	152844798	152844818	chr1	152880900	152880921	SRR372741.263	37	-	+

Meaning of the columns:

chrom1: The name of the chromosome on which tag1 exists.

- Use "." for unknown.

start1: The zero-based starting position of the tag1 on chrom1.

- The first base in a chromosome is numbered 0.
- As with BED format, the start position in each BEDPE feature is therefore interpreted to be 1 greater than the start position listed in the feature. This column is required.
- Use -1 for unknown.

end1: The one-based ending position of tag1 on chrom1.

- The end position in each BEDPE feature is one-based.
- Use -1 for unknown.

chrom2: The name of the chromosome on which tag2 exists.

- Use ”.” for unknown.

start2: The zero-based starting position of tag2 on chrom2.

- The first base in a chromosome is 0.
- As with BED format, the start position in each BEDPE feature is therefore interpreted to be 1 greater than the start position listed in the feature. This column is required.
- Use -1 for unknown.

end2: The one-based ending position of tag2 on chrom2.

- The end position in each BEDPE feature is one-based.
- Use -1 for unknown.

name: Definingthe name of the BEDPE feature.

- In this file, name is the reads id gotten from raw fastq format file.

score: The “score” field is the mapping quality score from the BAM alignment.

strand1: Definingthe strand for the first end of the feature.Either ‘+’ or ‘-‘.

- Use ”.” for unknown.

strand2: Defines the strand for the second end of the feature. Either ‘+’ or ‘-‘.

- Use ”.” for unknown.

(2) PREFIX.1_1.bedpe.selected.txt and PREFIX.1_1.bedpe.selected.unique.txt

PREFIX.1_1.bedpe.selected.txt is converted from the bedpe format file PREFIX.1_1.bedpe described above. The changes are at two point: 1) the name of the BEDPE feature is replaced by '.'; and 2) the mapping quality score of each PET is replaced by the mapping cutoff.

PREFIX.1_1.bedpe.selected.unique.txt is a file generated from PREFIX.1_1.bedpe.selected.txtby keeping only one PET from those PETs with both tags mapped at the same positions.

chrom1	start1	end1	chrom2	start2	end2	/	cutoff	strand1	strand2
chr5	175964094	175964115	chr5	175964271	175964291	.	20	-	+
chr1	166569355	166569376	chr12	56223587	56223608	.	20	+	+
chr14	71969478	71969499	chr19	16530595	16530616	.	20	-	-
chr10	108878463	108878484	chr2	184995039	184995059	.	20	-	-
chr10	29904926	29904946	chr11	32678927	32678947	.	20	-	+
chr9	127735487	127735507	chr9	127736244	127736264	.	20	-	+
chr2	191497466	191497487	chr4	83532329	83532350	.	20	+	-
chr17	79680058	79680079	chr17	79680152	79680172	.	20	-	+
chr16	10275488	10275508	chr6	150597540	150597561	.	20	-	+
chr1	152844798	152844818	chr1	152880900	152880921	.	20	-	+

Meaning of the columns:

chrom1: The name of the chromosome on which the tag1 exists.And all the PETs that either side chromosome information unknown are removed.

start1: The zero-based starting position of tag1on chrom1.

end1: The one-based ending position of tag1on chrom1.

chrom2: The name of the chromosome on which the tag2 exists. And all the PETs that either side chromosome information unknown are removed.

start2: The zero-based starting position of tag2 on chrom2.

end2: The one-based ending position of tag2 on chrom2.

/: Using '.' to replace the name of the BEDPE feature. It doesn't have exact meaning.
cutoff: Threshold of score value. Using this value to replace the mapping score.
strand1: Defines the strand for the first end of the feature.
strand2: Defines the strand for the second end of the feature.

(3) PREFIX.1_1.bedpe.selected.unique.pet.txt
This file is used to illustrate the PETs after merging those mapped to the same genome position, which are in a high chance caused by PCR amplification. And the mapped location of the tag sequence is replaced with a point, which is the border between tag and linker.

chrom1	locus1	strand1	chrom2	locus2	strand2
chr10	100003964	-	chr10	100005553	-
chr10	100012097	+	chr10	100012839	-
chr10	100004117	-	chr10	100070385	-
chr10	100022370	-	chr10	100025881	-
chr10	100029170	-	chr10	100056212	-
chr10	100059995	-	chr10	100061239	-
chr10	100069332	-	chr10	100086674	-
chr10	100072260	-	chr10	100173581	-
chr10	100072560	-	chr10	100175402	-
chr10	100092053	-	chr10	100171865	-

Meaning of the columns:
chrom1: The name of the chromosome on which tag1 exists.
locus1: Locus1 is the border between tag1 and linker. Using end1 as the value of location if tag1 is aligned to plus strand; in contrast, start1 is used when tag1 is aligned to minus strand.
strand1: Defines the strand for tag1.
chrom2: The name of the chromosome on which tag2 exists.
locus2: Locus2 is the border between tag2 and linker. Using end2 as the value of location if tag2 is aligned to plus strand, in contrast, start2 is used when tag2 is aligned to minus strand.
strand2: Defines the strand for tag2.

(4) PREFIX.1_1.bedpe.selected.merged.pet.txt
PREFIX.bedpe.selected.pet.txt
PREFIX.ipet
PREFIX.spet
PREFIX.opet
PREFIX.1_1.bedpe.selected.merged.pet.txtis used to illustrate the location of PETs after removing similar position caused by PCR amplification and ultrasonic disruption.
PREFIX.bedpe.selected.pet.txtcontains same linker PETs after removing redundancy.
PREFIX.ipet is the file with inter-ligation PETs.
PREFIX.spetis the file with self-ligation PETs.
PREFIX.opet is the file containing the other PETs with short distance.

chrom1	locus1	strand1	chrom2	locus2	strand2	counts	index
chr10	103618745	-	chr10	103618888	+	1.0	-1

chr10	103619140	-	chr10	103662080	+	1.0	-1
chr10	103633680	-	chr10	103634188	+	1.0	-1
chr10	103634897	-	chr10	103753306	+	1.0	-1
chr10	103635188	-	chr10	103635449	+	2.0	-1
chr10	103640011	-	chr10	103641203	+	1.0	-1
chr10	103647998	-	chr10	103648487	+	1.0	-1
chr10	103657683	-	chr10	103657935	+	1.0	-1
chr10	103661110	-	chr10	103661482	+	1.0	-1
chr10	103661281	-	chr10	103661712	+	1.0	-1

Meaning of the columns:

chrom1: The name of the chromosome on which tag1 exists.

locus1: The location of tag1 after merging similar PETs.

strand1: Defines the strand for tag1.

chrom2: The name of the chromosome on which tag2 exists.

locus2: The location of tag2 after merging similar PETs.

strand2: Defines the strand for tag2.

counts: Counts of similar position PETs mapped to genome.

index: Indexis reserved for future use.

(5) PREFIX.cluster.FDRfiltered.txt

This file contains the statistically-significant interaction clusters and related information.

chrom1	start1	end1	chrom2	start1	end2	ipet counts	type
chr10	5454640	5455156	chr10	5469890	5470405	2	1
chr10	5487515	5489041	chr10	5587424	5589742	9	1
chr10	5488183	5488970	chr10	5655080	5655943	2	1
chr10	5488356	5489176	chr10	5534701	5535565	2	1
chr10	5488670	5489194	chr10	5531320	5531883	2	1
chr10	5489401	5490133	chr10	5638857	5639501	2	1
chr10	5490050	5491267	chr10	5587813	5589259	4	1
chr10	5492324	5492945	chr10	5586500	5587151	2	1
chr10	5504728	5505495	chr10	5588057	5588979	2	1
chr10	5505491	5506273	chr10	5589580	5590267	2	1

distance	anchor1 tag counts	anchor2 tag counts	p-value	p.adjust	-log10(p-value)	-log10(p.adjust)
15249	18	5	2.92e-11	1.46e-10	10.53	9.84
100305	182	390	7.56e-26	3.71e-24	25.12	23.43
166935	159	14	2.18e-08	3.32e-08	7.66	7.48
46367	159	50	2.94e-07	3.40e-07	6.53	6.47
42669	82	35	3.77e-08	5.35e-08	7.42	7.27
149412	41	22	3.61e-09	7.35e-09	8.44	8.13
97877	63	289	3.70e-13	3.38e-12	12.43	11.47
94191	12	57	2.01e-09	4.54e-09	8.70	8.34
83406	20	187	6.30e-08	8.40e-08	7.20	7.08

84041	18	92	1.22e-08	2.03e-08	7.91	7.69

Meaning of the columns:

chrom1: The name of the chromosome on which cluster anchor1 exists.

start1: The start location of cluster anchor1.

end1: The end location of cluster anchor1.

chrom2: The name of the chromosome on which cluster anchor2exists.

start2: The start location ofcluster anchor2.

end2: The end location of cluster anchor2.

ipet counts: Counts of PETs between two anchors of an interaction cluster.

type: Interactive type. 1 represents intra-chromosomal interaction, 0 for inter-chromosomal interaction.

distance: Distance between both two anchors of intra-chromosomal PET. And if the two anchors are located on different chromosomes, the value of distance is 2,147,483,647 (as the largest integer with 32bit, 2^31-1).

anchor1 tag counts: Counts of tag loci fall into cluster anchor1.

anchor2 tag counts: Counts of tag loci fall into cluster anchor2.

p-value: This value represents the statistical significance of the interaction. It is calculated by hyper-geometric distribution.

p.adjust: P.adjust means p-value adjustment with Benjamini-Hockberg method (1995)

-log10(p-value): The negative logarithm of p-value.

-log10(p.adjust): The negative logarithm of adjusted p-value.

(6) PREFIX.peak

This file is used to illustrate the primary peak information obtained from self-ligation PETs.

chrom	start	end	summit start	summit end	peak coverage
chr1	840639	840651	840639	840651	14
chr1	841447	841511	841447	841511	7
chr1	843472	843495	843472	843495	5
chr1	846703	846732	846703	846732	16
chr1	847869	847883	847869	847883	5
chr1	849284	849285	849284	849285	20
chr1	852786	852788	852786	852788	46
chr1	853677	853746	853677	853679	58
chr1	854750	854812	854750	854763	48
chr1	858117	858423	858117	858149	9

Meaning of the columns:

chrom: The name of the chromosome on which peak exists.

start: The start position of peak. This column is used in broad peak.

end: The end position of peak. This column is used in broad peak.

summit start: The start position of peak summit. This column is used in narrow peak.

summit end: The end position of peak summit.

peak coverage: The coverage of the overlapped peaks.

(7) PREFIX.peak.FDRfiltered.txt

This file is used to illustrate the high potential peaks which passed the threshold of adjusted p-value.

chrom	start	end	peak coverage	p-value	p.adjust
chr1	852786	852788	46	1.73e-07	1.8e-06
chr1	853677	853679	58	1.44e-13	2.31e-12
chr1	854750	854763	48	1.2e-07	1.26e-06
chr1	860735	860737	37	9.32e-08	1.03e-06
chr1	877184	877186	23	1.25e-09	1.6e-08
chr1	902065	902068	53	1.08e-40	4.55e-39
chr1	935164	935172	61	4.6e-30	1.36e-28
chr1	936330	936335	46	8.52e-16	1.51e-14
chr1	955915	955938	31	2.29e-17	4.29e-16
chr1	994919	994929	25	5.89e-09	7.26e-08

Meaning of the columns:

chrom: The name of the chromosome on which the peak exists

start: The start position of peak

end: The end position of peak

peak coverage: The coverage of the overlapped peaks.

p-value: This value represents the reliability of a real peak. It is calculated by Poisson distribution.

p.adjust: P.adjust means p-value adjustment with Benjamini-Hockberg method.

Files used in thevisualization report:

(1) PREFIX.linker_alignment_score_distribution.txt

This file is used to illustrate the distribution of best alignment scores from the designed linker sequences to the reads.

alignment score	counts of read1	counts of read2
0	0	99987
1	2815	9249
2	27337	46442
3	24444	46457
4	67866	81805
5	87995	95125
6	103164	105683
7	100254	123026
8	98584	125184
9	84207	120977
10	124532	162743
11	186583	256884
12	135917	204520
13	808328	1041042
14	1521765	1875530

15	33572085	32416084
16	43485247	43744265
17	967484	909505
18	100999	65062
19	57964	28000

Meaning of the columns:

alignment score: Local alignment score of reads with the best matched linker by dynamic programming.

counts of read1: Counts of read1 that are aligned to best matched linker with certain local alignment score.

counts of read2: Counts of read2 that are aligned to best matched linker with certain local alignment score.

(2) PREFIX.linker_alignment_score_difference_distribution.txt

This file is used to illustrate the score difference distribution between the best-aligned linker and the second-best aligned linker. It shows how the barcode distinguishes different linkers.

score difference	counts of read1	counts of read2
0	237941	381176
1	247817	243882
2	282467	347700
3	877610	787925
4	406169	562655
5	79505566	79234232

Meaning of the columns:

score difference: Reads mappedto different linkers will get a set of local alignment scores. This column is the score difference between the best aligned and second-best aligned one.

counts of read1: Counts of read1 with certain score difference.

counts of read2: Counts of read2 with certain score difference.

(3) PREFIX.tag_length_distribution.txt

This file is used to illustrate the tag length distribution. Tags refer to the DNA fragments that are digested by enzyme *Mme*Ifar away from linkers. Such fragments are gotten by removing linker from sequencing reads. The selected tags will be kept for further processing.

tag length	counts of read1	counts of read2
0	0	99987
1	111	188
2	1974	1805
3	3808	4744
4	939	2204
5	1480	2448
6	1923	2922
7	2533	3813
8	7913	8360

9	103553	102500
10	41470	45188
11	34000	37377
12	22210	25660
13	25898	28738
14	30130	31580
15	27110	27857
16	44094	39730
17	73781	50902
18	88013	76375
19	1010611	1006545
20	45191493	45902278
21	34479577	33648812
22	222849	252044
23	37024	42881
24	15408	17240
25	12873	13127
26	14795	15981
27	14197	15407
28	17627	18741
29	13407	14184
30	8027	8469
31	5226	5380
32	2649	2447
33	778	1045
34	89	523
35	0	88

Meaning of the columns:

tag length: The length of tag sequence.

counts of read1: Counts of tag1 with certain tag length.

counts of read2: Counts of tag2 with certain tag length.

(4) PREFIX.linker_composition_distribution.txt

This file is used to illustrate the proportion of each linker combination aligned to the reads.

	A_A	A_B	B_A	B_B	Ambiguous	Total
Numbers	33822895	4744854	4748192	32998692	5242937	81557570
Percentage	41.47%	5.82%	5.82%	40.46%	6.43%	100%

Meaning of the columns:

A_A: "A_A" means the category of PETs that both tags optimally aligned to linker A.

A_B: "A_B" means the category of PETs that tag1 and tag2 optimally aligned to linker A and linker B respectively.

B_A: "B_A" means the category of PETs that tag1 and tag2 optimally aligned to linker B and

linker A respectively.

B_B: "B_B" means the category of PETs that both tags optimally aligned to linker B.

Ambiguous: "Ambiguous" means the category of PETs that not satisfy one of the criteria below: 1) the best linker alignment score is equal to or larger than the **minimum_linker_alignment_score**, 2) the difference between the best and the second best linker alignment scoresis equal to or larger than the **minSecondBestScoreDiff**, 3) tag length should conform to the specified range, and 4) the barcodes in the PETs must completely match the designed barcodes in the linkers.

Total: Counts of the whole PETs.

Attention: A_A and B_B should constitute the majority of the PETsin a good ChIA-PET library as the sample shows.

(5) PREFIX.1_1.mapping_statistics.txt

This file contains the statistics of the pair-end reads mapping results. Each end has three possible mapping results: Non-mappable, Uniquely-mapped and Others, and there are nine combinations from the two ends.

read1 \ read2	Non-mappable	Uniquely-mapped	Others
Non-mappable	46	4146	8459
Uniquely-mapped	6254	4757858	7807698
Others	11679	7163264	14063491

The meanings of the different categories are as follows:

Non-mappable: Mappingpaired-end reads to a reference genome using BWA can obtain a SAM format file. There are two conditions for Non-mappable reads. 1) If the mapping result of a tag in the SAM file doesn't have optional fields (less than 12 columns), it is classified into non-mappabletag. 2) If the mapping result of a tag in the SAM file has optional fields and there is a label XT:A:N, this tag is classified into "non-mappable" category.

Uniquely-mapped: If the mapping quality score MAPQ (5th column) of a tag is more than 20 and the optional fields contain the following labels: XT:A:U(12th column), X0:i:1 (16th column), X1:i:0 (17th column), this tagis classified into "uniquely-mapped" category.

Others: If the mapping result of a tag doesn't meet the criteria above, it is classified into "Others" category, which includes the tags with low mapping scores and multiply-mapped tags.

Attention: In a SAM format file, the mapping results of the PETs appear in pairs. So the first line in the pairs is read1, the second line in pairs is read2.

(6) PREFIX.1_1.bedpe.qc.dist.txt

This file contains the distribution of mapping scores from unique-mapping PETs.

counts	score
15234	25
4669548	37
2	50
73074	60

Meaning of the columns:

counts: Counts of PETs with the corresponding mapping scores.

score: Score value(8th column) from the bedpe format file above.

(7) PREFIX.1_1.bedpe.strand.dist.txt

This file contains the distribution of strand combinations from the uniquely-mapped PETs. There are four types of strand combinations: 1) both tags are mapped to plus strands (++) , 2) tag1 is mapped to plus strand and tag2 is mapped to minus strand (+-), 3) tag1 is mapped to minus strand and tag2 is mapped to plus strand (-+), 4) both tags are mapped to minus strands (--).

counts	strand1	strand2
841994	-	-
2227108	-	+
845768	+	-
842988	+	+

Meaning of the columns:

counts: Counts of certain strand combination.

strand1: Strand information of tag1 (9th column) from the bedpe format file above.

strand2: Strand information of tag2 (10th column) from the bedpe format file above.

(8) PREFIX.1_1.bedpe.selected.dist.txt

This file is about the summary statistics of the redundancies caused by PCR amplification. The redundancy means the PETs that are mapped exactly to the same genomic position, which are from PCR amplification with high chance.

counts	repetitions
3358894	1
566169	2
75928	3
8548	4
841	5
66	6
7	7

Meaning of the columns:

counts: means the number of frequencies appeared in the data.

repetitions: meansthe frequency of PETs mapped to the same genomic position.

(9) PREFIX.1_1.bedpe.selected.unique.intra-chrom.strand.dist.txt

This file is used to illustrate strands combination of intra-chromosomal inter-ligation PETs.

counts	strand1	strand2
169000	-	-
1338210	-	+
172396	+	-
169847	+	+

Meaning of the columns:

counts: Counts of PETs with certain strand combinations. The selected are intra-chromosomal inter-ligation PETs.

strand1: Strand of tag1 from the intra-chromosomal inter-ligation PETs after removing redundancy.

strand2: Strand of tag2 from the intra-chromosomal inter-ligation PETs after removing

redundancy.

(10) PREFIX.bedpe.selected.intra-chrom.distance.txt

This file contains the distance between two tags of the PETs and strand information. Based on the ChIA-PET design and Illumina sequencing principle, there should be more PETs with strand combination -+ at short distance than other combination.

span	strand1	strand2
1589	-	-
1036	-	-
742	+	-
66268	-	-
2505	-	-
3511	-	-
27042	-	-
27157	-	-
1244	-	-
17342	-	-

Meaning of the columns:

span: Span between the two tags of individual intra-chromosomal PETs.

strand1: Strand of tag1 in an intra-chromosomal PETs.

strand2: Strand of tag2 in an intra-chromosomal PETs.

(11) PREFIX.bedpe.selected.intra-chrom.distance.plusplus.txt

PREFIX.bedpe.selected.intra-chrom.distance.plusminus.txt

PREFIX.bedpe.selected.intra-chrom.distance.minusplus.txt

PREFIX.bedpe.selected.intra-chrom.distance.minusminus.txt

These four files are separated from PREFIX.bedpe.selected.intra-chrom.distance.txtaccording to strand combination.

span
90590128
672
527
62489
3631
99731215
108495
1440
52
1014

Meaning of the columns:

span: Span between the two tags of individual intra-chromosomal PETs in certain strand combination.

