

Supplementary Material

GUAVA: a Graphical User interface for the Analysis and Visualization of ATAC-seq data

Mayur Divate, Edwin Cheung*

* **Correspondence:** Corresponding Author: echeung@umac.mo

1 Supplementary Figures and Tables

1.1 Supplementary Figures

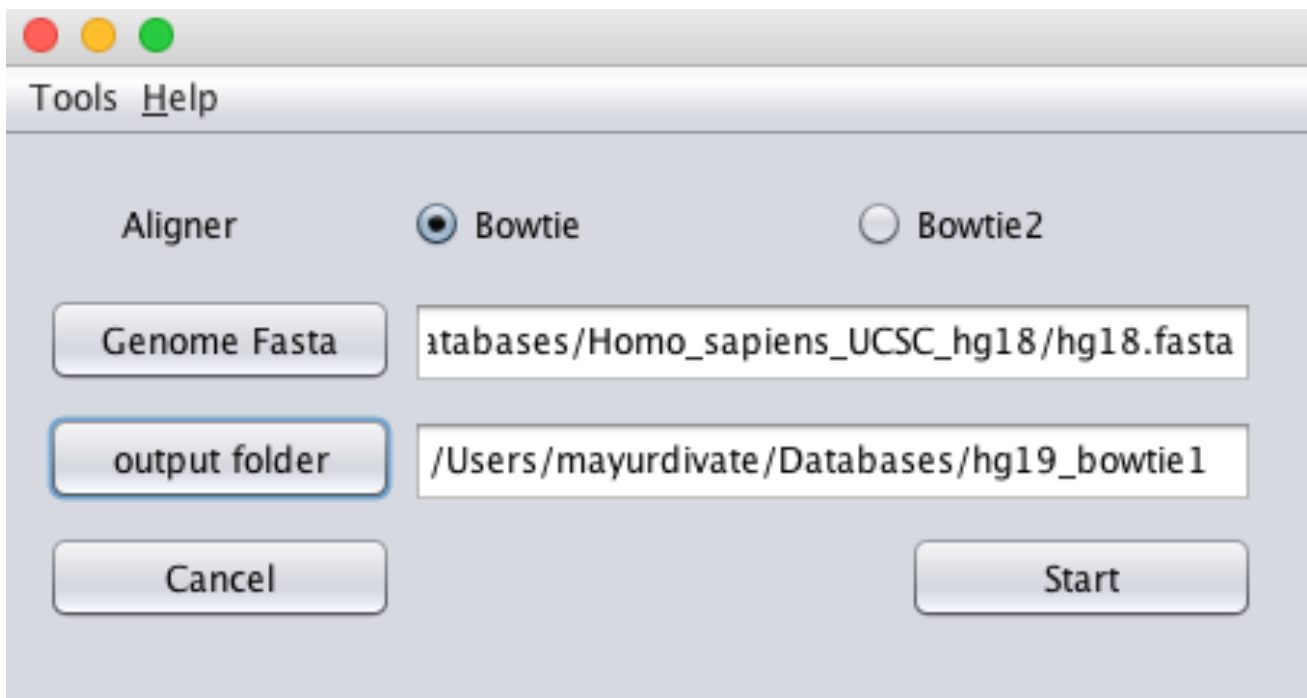
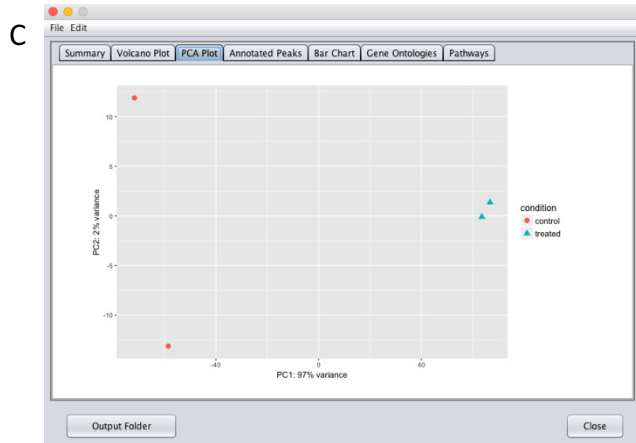
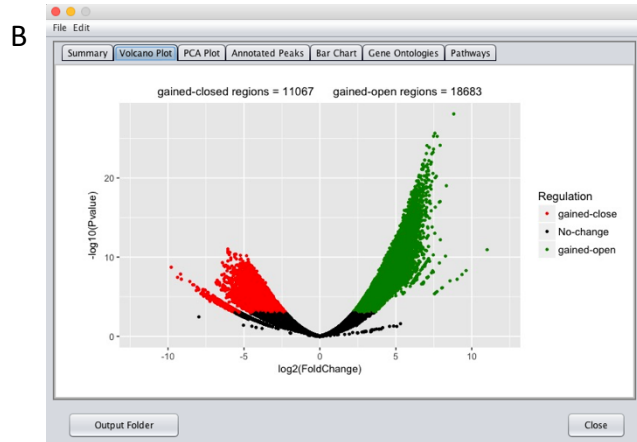
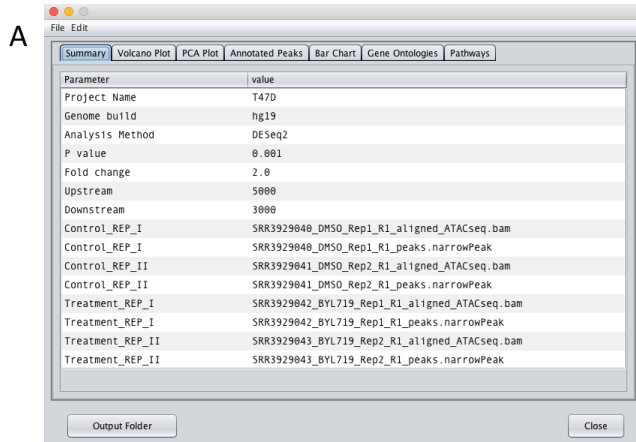
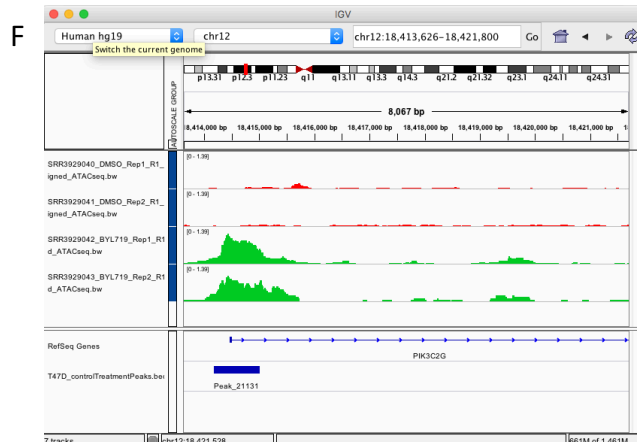
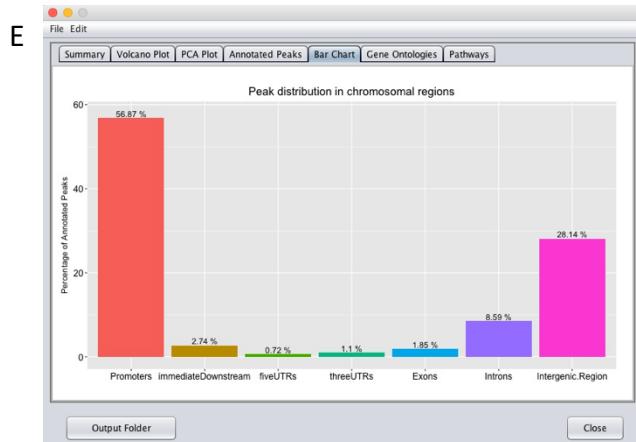


Figure S1. Genome index builder. Input window for the genome index builder program. Here, users choose between Bowtie and Bowtie2. Users also upload the genome fasta file using the 'Genome Fasta' button. Users also select the output folder here. Finally, clicking on the 'Start' button will start the genome index builder program.



D

Chr	Start	End	Length	log2(FC)	P value	adj. P va...	Regulation	Gene Sym...	Distance
chrX	48776111	48776907	797	3.082	7.18E-04	2.68E-03	gained-o...	PI2M	0
chr10	81102002	81102305	304	3.063	2.56E-04	1.13E-03	gained-o...	PPF	4914
chr10	99449252	99449892	641	-3.34	3.33E-04	1.40E-03	gained-c...	AVP1	2236
chr11	75268593	75269722	1180	-3.144	6.54E-05	3.54E-04	gained-c...	SERP1H1	3328
chr12	18414190	18415015	826	4.919	4.65E-10	1.73E-08	gained-o...	PK3C2G	0
chr12	130819...	130820...	425	2.682	4.14E-04	1.69E-03	gained-o...	PIM1	2011
chr14	94759613	94760014	402	3.514	1.74E-05	1.15E-04	gained-o...	SERPINA10	4
chr14	94788955	94789881	1027	3.691	8.82E-06	6.44E-05	gained-o...	SERPINA6	0
chr14	94856539	94857605	1067	5.54	8.40E-16	2.75E-13	gained-o...	SERPINA1	0
chr14	95103054	95103719	666	4.49	7.10E-09	1.69E-07	gained-o...	SERPINA...	3342
chr15	41135960	41136916	957	4.608	1.92E-09	5.66E-08	gained-o...	SPIN1	0
chr15	90459850	90460358	509	3.643	7.52E-06	5.61E-05	gained-o...	SRPN	3627
chr17	26903719	26904410	692	3.181	8.58E-06	6.28E-05	gained-o...	PIG5	4831
chr17	27367060	27367800	741	-5.27	2.87E-05	1.76E-04	gained-c...	PIPOX	2117
chr18	3012404	3014911	2508	-9.176	1.37E-08	2.92E-07	gained-c...	LPIN2	458
chr18	3015092	3015862	871	-7.973	1.30E-06	3.28E-05	gained-c...	LPIN2	3146
chr18	11149229	11150045	817	-3.167	6.46E-04	2.45E-03	gained-c...	PIEZO2	467
chr18	61143904	61144332	429	6.316	3.90E-13	4.59E-11	gained-o...	SERPIN5	0
chr18	61549738	61550433	696	-2.611	7.53E-04	2.79E-03	gained-c...	SERPIN2	4505
chr19	38754784	38755410	627	4.947	2.83E-08	5.03E-07	gained-o...	SPIN2	0
chr1	160000...	160002...	2122	2.413	2.60E-04	1.14E-03	gained-o...	PIGM	0
chr20	36928432	36928809	378	3.766	3.51E-06	2.94E-05	gained-o...	BPI	3742
chr20	39968282	39970221	1940	2.593	8.86E-05	4.63E-04	gained-o...	LPIN3	0
chr21	15588178	15589145	968	2.833	8.80E-04	3.12E-03	gained-o...	LPI	4965
chr5	147214...	147215...	564	3.762	1.90E-07	2.57E-06	gained-o...	SPINK1	3571
chr5	147545...	147546...	1010	4.168	1.11E-08	2.44E-07	gained-o...	SPINK14	1617



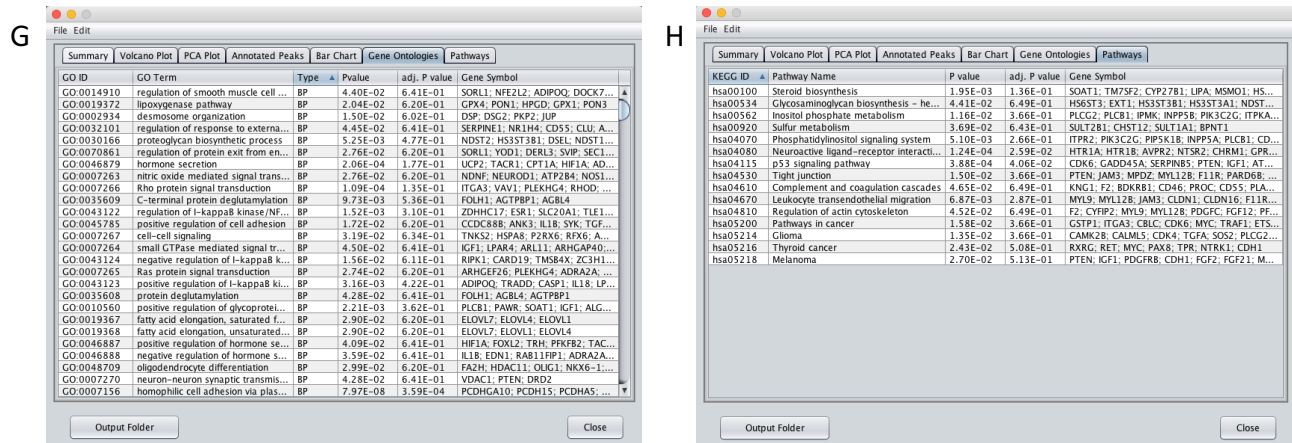


Figure S2. Output interface for GUAVA ATAC-seq differential analysis. A) Input summary. B) Volcano plot indicating the differentially enriched peaks. Red: peaks with increased chromatin accessibility, green: peaks with reduced chromatin accessibility and black: peaks with no significant change in chromatin accessibility. C) Principal component analysis (PCA) plot for the control and treated replicates D) A table containing the annotated differentially enriched ATAC-seq peaks. This tab also provides easy access to IGV to visualize differentially enriched peaks and normalized ATAC-seq signals for each sample E) Bar chart showing the percentage of peaks in various genomic locations such as the promoter, intron, exon, UTR, etc. F) Peak visualization in IGV. G) Over represented gene ontologies and H) KEGG pathways

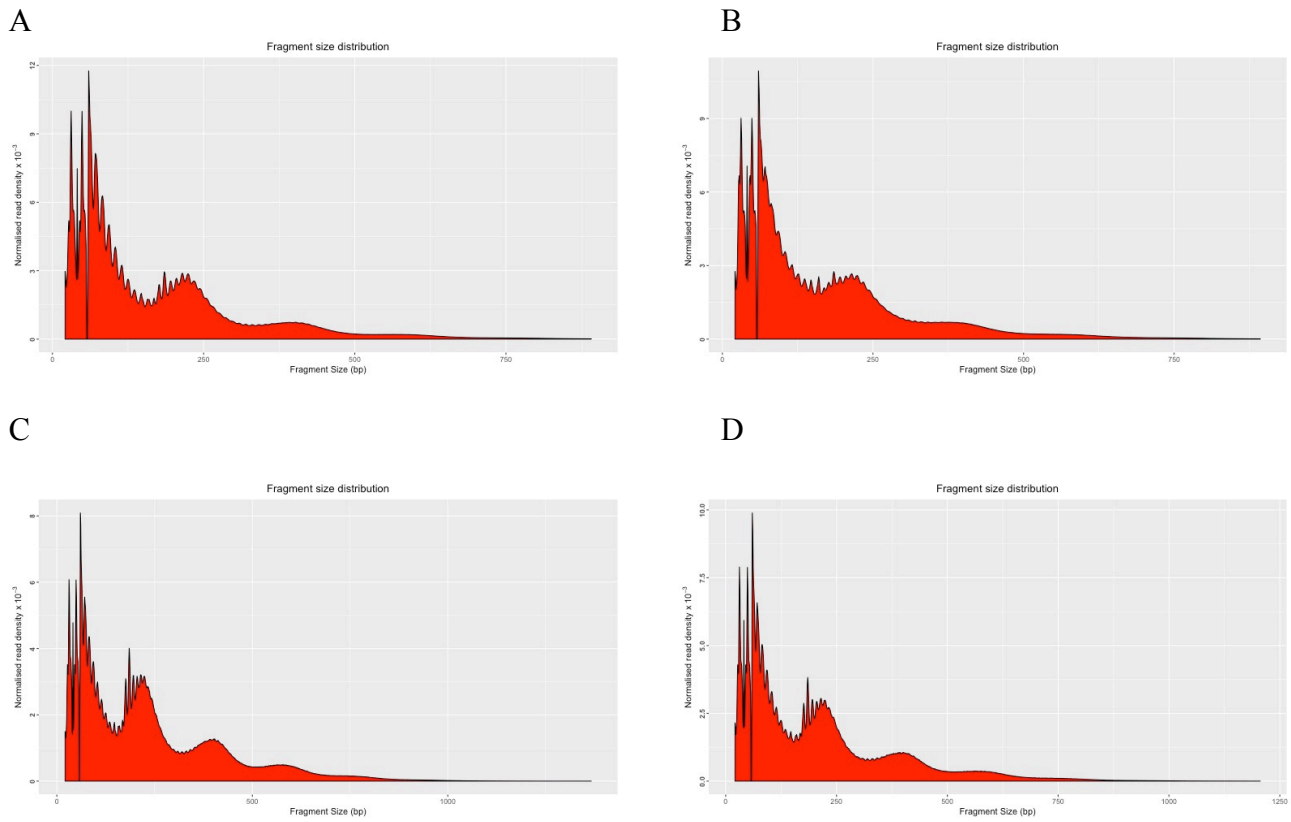


Figure S3. Fragment size distribution graphs. Graphs showing the fragment size distribution of ATAC-seq reads in T47D cells for A) DMSO replicate 1, B) DMSO replicate 2, C) BYL719 replicate 1, and D) BYL719 replicate 2.

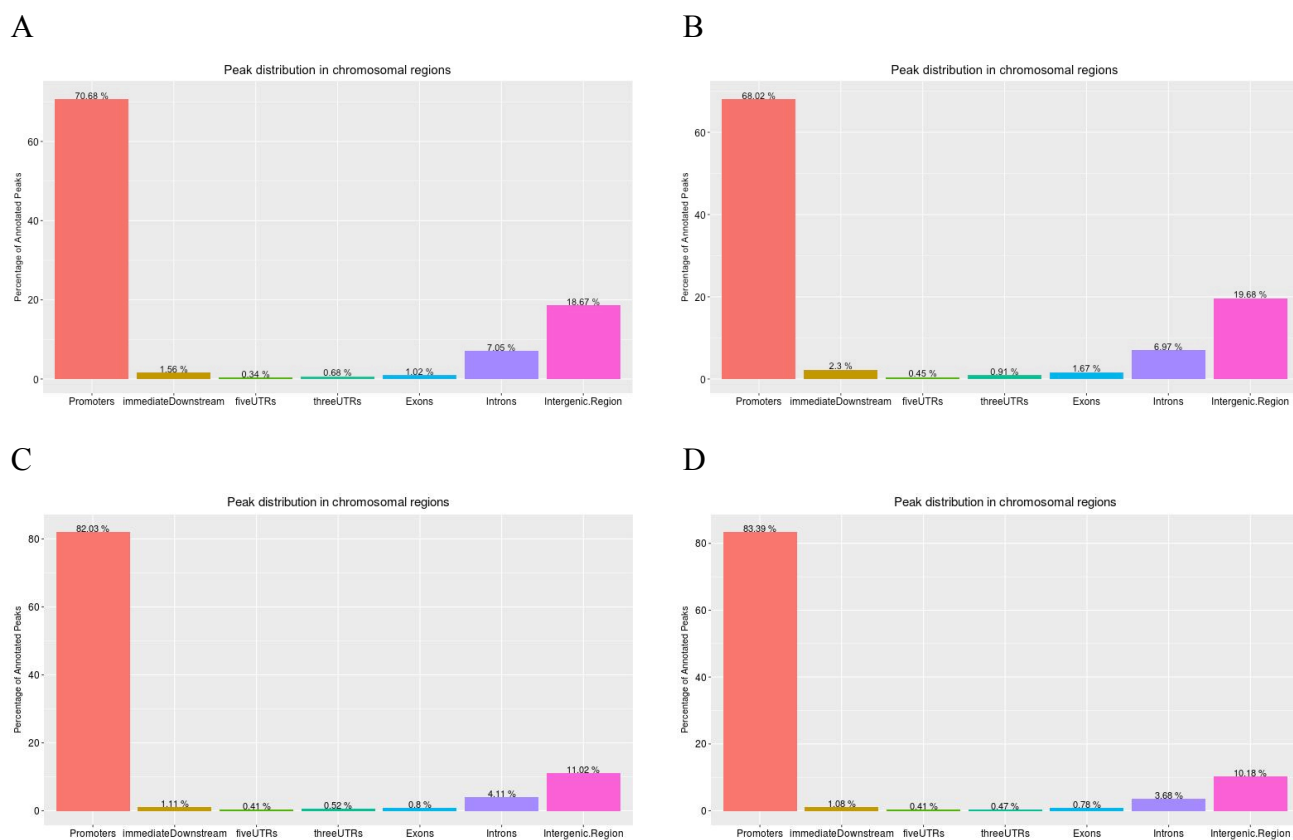


Figure S4. Genomic distribution of ATAC-seq peaks. Bar charts showing the distribution of ATAC-seq peaks for A) DMSO replicate 1, B) DMSO replicate 2, C) BYL719 replicate 1, and D) BYL719 replicate 2.

1.2 Supplementary Tables

Table S1. GEO accession number for the samples from the case study.

Sample	GEO Accession Number	Treatment	Replicate Number
DMSO_Rep1	GSM2241147	DMSO	Replicate I
DMSO_Rep2	GSM2241148	DMSO	Replicate II
BYL719_Rep1	GSM2241149	BYL719	Replicate I
BYL719_Rep2	GSM2241150	BYL719	Replicate II

Table S2. GUAVA alignment and peak calling statistics.

Sample Name	BYL719_Rep1	BYL719_Rep2	DMSO_Rep1	DMSO_Rep2
Total Reads	44892526	29411509	101692391	94208170
Total number of reads aligned	34410274 (76.65%)	22322001 (75.9%)	82027622 (80.66%)	75771166 (80.43%)
Mapping quality filtered reads	7551956 (16.82%)	5059044 (17.2%)	13940905 (13.71%)	10717722 (11.38%)
Total number of reads failed to align	2930296 (6.53%)	2030464 (6.9%)	5723864 (5.63%)	7719282 (8.19%)
Reads Mapped to ChrM	7957443 (17.73%)	5496204 (18.69%)	15552430 (15.3%)	2907706 (3.09%)
Duplicate Reads	12865311 (28.66%)	7001849 (23.81%)	25430014 (25.01%)	13855635 (14.71%)
Total Useful Reads	19952669 (44.45%)	14051507 (47.78%)	55033289 (54.12%)	60956787 (64.7%)
Total number of Peaks identified	65735	56734	151009	147675