LSM3241 CA2

Hunting for Genomic Insertions and their Consequences

Tan Wei Qi A0158131M

27 March 2019

1. Introduction

DNA transposon is a chromosomal segment that can move from one location to another location within the genome. These transpositions often have significant roles in genetic changes and thus are widely used for transposon-based mutagenesis research in different organisms (Munoz-Lopez & Garcia-Perez, 2010). It is an effective and less labour-intensive method to generate large amount of random mutations that offers application purposes in genetic screens for both prokaryotes and eukaryotes (Xu, Bharucha & Kumar, 2011).

This research focuses on transposon-based mutagenesis in the yeast (*Saccharomyces cerevisiae*) genome, specifically in strain S288C. Yeast has been extensively used in transposon mutagenesis studies due to its complete genomic publication in 1996 (Botstein & Fink, 2011).

The aim of this study is to identify the possible alterations caused by one or more insertions of a transposon (Ty5-6p genebank accession U19263.1) in the yeast genome. Mapping and analysis of next generation sequencing data are used to identify the mapped positions and consequences of transposon insertions.

2. Methods

2.1 Setup and Workflow

An altered paired-end sequencing yeast genome (A0158131M.tgz) by insertion of transposon Ty5-6p was provided to conduct the study. FASTA files of reference yeast genome (sacCer.fa) and the transposon (ty5_6p.fa) were also given. The general workflow shown in Figure 1 summarises the data extraction and analysis done on Ubuntu Linux system and Integrative Genomics Viewer (IGV). Packages “fastqc”, “bowtie2” and “samtools” were also installed. The full annotated reproducible code can be found on https://github.com/Tan-WeiQi/LSM3241-CA2.

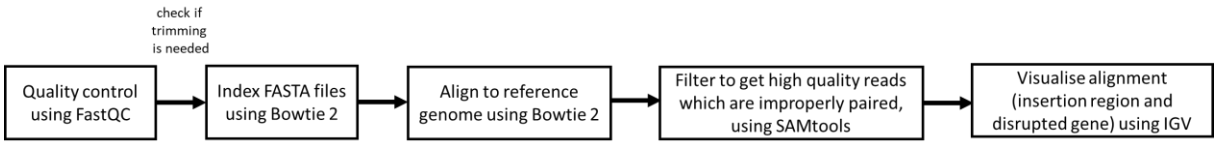


Figure 1: Flowchart representing the workflow of this study

2.2 Assessing Quality using FastQC

FastQC is a tool to enable quality control for high throughput genomic sequence data (Leggett, Ramirez-Gonzalez, Clavijo, Waite & Davey, 2013). Before conducting further analysis, results from FastQC reports for each end (A0158131M_1.fq and A0158131M_2.fq) of the paired-end sequencing genome were accessed to determine if the sequences need to be trimmed using the “trimmomatic” package. Trimmomatic is a flexible and efficient tool to filter and trim poor-quality reads. This step is important as contamination by adapter sequences, duplication and low-quality reads can result in inaccurate mapping and hence suboptimal analysis (Bolger, Lohse & Usadel, 2014).

2.3 Alignment to Reference Genome using Bowtie 2

Before alignment, the FASTA files (sacCer3.fa and ty5_6p.fa) must first be indexed so that aligner can narrow down the potential alignment sites of a query sequence within the genome. This allow reads to be quickly aligned which helps to save both time and memory (Kahveci, Ljosa & Singh, 2004). Since Bowtie 1 is geared towards shorter sequencing reads up to 50 base pairs, Bowtie 2 was used to index instead because it is faster, more sensitive and uses less memory than Bowtie 1. Hence, Bowtie 2 is more suitable for handling longer sequencing reads efficiently in this study (Fonseca, Rung, Brazma & Marioni, 2012).

After indexing, paired-end FASTQ files (A0158131M_1.fq and A0158131M_2.fq) were aligned to the indexed files. Using Bowtie 2 algorithm, 4 parallel search threads were launched simultaneously on different processors and a "very fast" speed was chosen to achieve greater alignment speed for these long sequencing reads (Langmead, Wilks, Antonescu & Charles, 2018). The resulting SAM file is a tab-delimited text file containing all the relevant information of each individual read and its alignment to reference genome. This information is needed for locating the transposon insertions on the yeast genome.

2.4 Post-Alignment Clean Up using SAMtools

SAMtools has many utilities that enable post-processing DNA sequences in SAM, BAM and CRAM files such as indexing and sorting (Li et al., 2009). It was first used to convert the resulting SAM file into a compressed binary BAM file to reduce its size and allow for indexing. Then, it was used to sort the BAM file by coordinates and index it so that its data can be accessed efficiently for visualisation on IGV (Carver et al., 2012).

The BAM file was further filtered using the grep command line (grep -e 'TY5') to only include reads/mates mapped to Ty5-6p transposon. SAM flag value 1 (-f 1) was then included to filter reads and mates which are paired, and SAM flag value 14 (-F 14) was then excluded to filter out paired-ends which are mapped in proper pair, and have their reads unmapped and mates unmapped. This filtering process via grep and SAM flags ensure the final filtered data only contain data which are improperly mapped, where the read is mapped to yeast genome, and its corresponding mate to the Ty5-6p transposon, or vice versa. This will suggest the possible insertion points (Grimm, Hagmann, Koenig, Weigel & Borgwardt, 2013).

Further filtering was done via removing data which has mapping quality score of below 10, so that only data which have its reads/mates accurately mapped are left for analysis (Ruffalo, Koyuturk, Ray & LaFramboise, 2012). The bitwise FLAG values showed in the remaining filtered data provides information on whether the read and mate are on the reverse or forward DNA strand, which allows us to determine the orientation of the insertions.

2.5 Visualisation of Alignment using IGV

Finally, the visualization using IGV helps to locate the region of insertions, the number of insertions, and the possible gene disruptions in the yeast genome due to insertions (Thorvaldsdottir, Robinson & Mesirov, 2012). This was done by loading the yeast genome and indexed BAM file onto IGV.

3. Results and Discussions

3.1 FastQC Report

In this study, no trimming was needed as the paired-ends have high per base sequence quality as shown in Figure 2 where the reads fall mainly on the green region which represents very good quality score.

Figure 2: Per base sequence quality for A0158131M_1.fq (top) and A0158131M_2.fq (bottom)

In addition, adaptor sequences were not found in the reads as shown in Figure 3, hence no removal was needed before mapping.

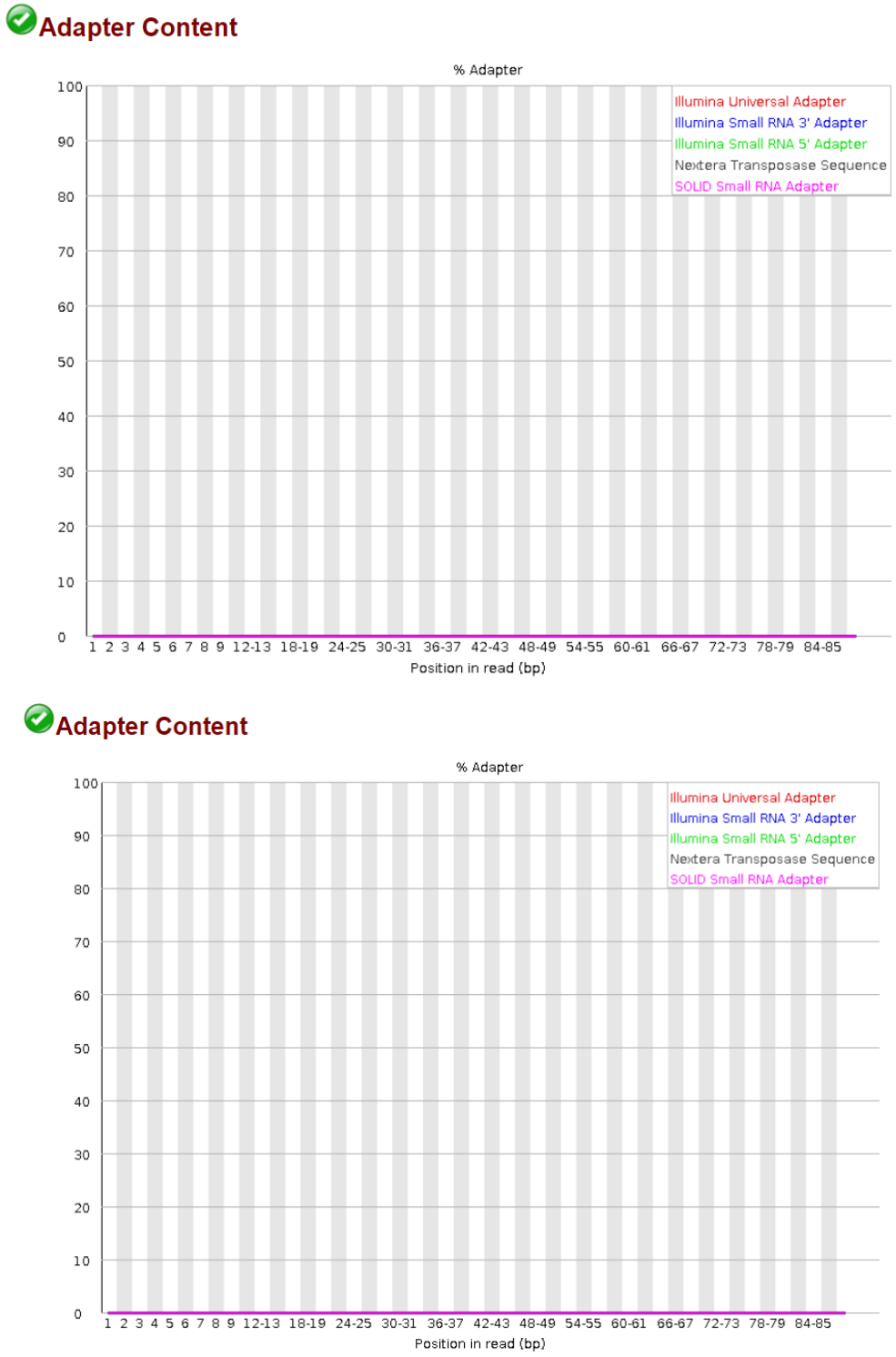


Figure 3: Adaptor Content for A0158131M_1.fq (left) and A0158131M_2.fq (right)

Although K-mer content is warned, the sharp spike in content at the start of read is still considered low at around 6 for both FASTQ files. This Kmer bias at the start may be due to incomplete sampling of possible random primers of the yeast clone provided (Bioinformatics.babraham.ac.uk, 2019). Hence, no trimming is needed.

3.2 Analysing the Location, Orientation and Number of Insertions

The final remaining relevant data that are improperly paired and have quality score above 10 after all the filtering processes mentioned in Methods are shown in Table 1, Table 2 and Table 3 for Chromosome V, Chromosome XII and Chromosome XIV respectively.

QNAME	FLAG	RNAME	POS	MAPQ	MRNM	MPOS
chrV-111274	177	chrV	269771	23	TY5	5106
chrV-112020	177	chrV	269751	42	TY5	5123
chrV-47696	129	chrV	269502	42	TY5	5301
chrV-59366	65	chrV	269500	35	TY5	5290
chrV-71220	129	chrV	269446	40	TY5	5337
chrV-76154	65	chrV	269467	42	TY5	5310

Table 1: Chromosome V data for analysis

QNAME	FLAG	RNAME	POS	MAPQ	MRNM	MPOS
chrXII-116888	81	chrXII	367546	42	TY5	5325
chrXII-170778	81	chrXII	367560	42	TY5	5317
chrXII-60618	97	chrXII	367244	23	TY5	5112

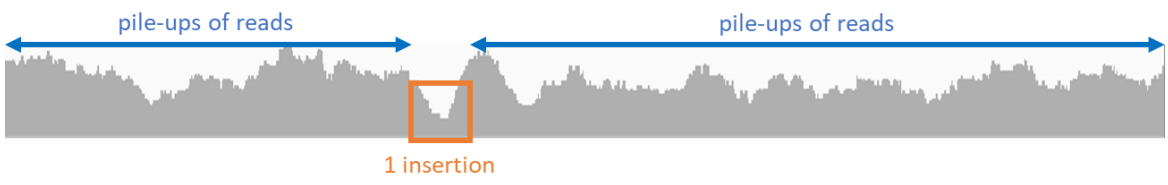
Table 2: Chromosome XII data for analysis

QNAME	FLAG	RNAME	POS	MAPQ	MRNM	MPOS
chrXIV-109550	161	chrXIV	488611	23	TY5	5112
chrXIV-116008	161	chrXIV	488589	23	TY5	5112
chrXIV-154642	145	chrXIV	488869	42	TY5	5280
chrXIV-19604	97	chrXIV	488584	23	TY5	5112
chrXIV-21884	145	chrXIV	488926	42	TY5	5334

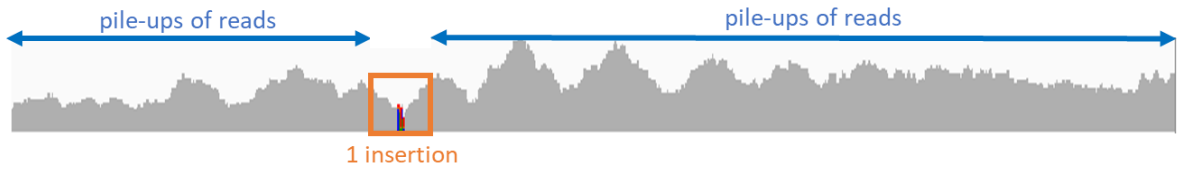
Table 3: Chromosome XIV data for analysis

The 3 tables implied that there are insertions on Chromosome V, XII and XIV. Hence, IGV was used to visualise the alignments to further confirm the presence of insertions and analyse the number of insertions on each chromosome. The POS values in the tables helped us to narrow down the region for analysis on IGV. Figure 4 highlighted that each chromosome only has one insertion as they have pile up of reads on both sides (blue arrows) with only one deep in-between them (orange boxes) which represents the possible area of Ty5-6p transposon insertion.

A



B



C

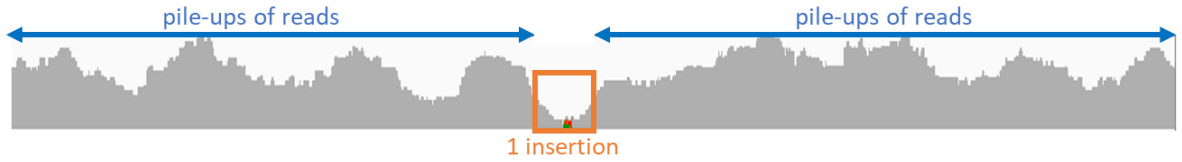
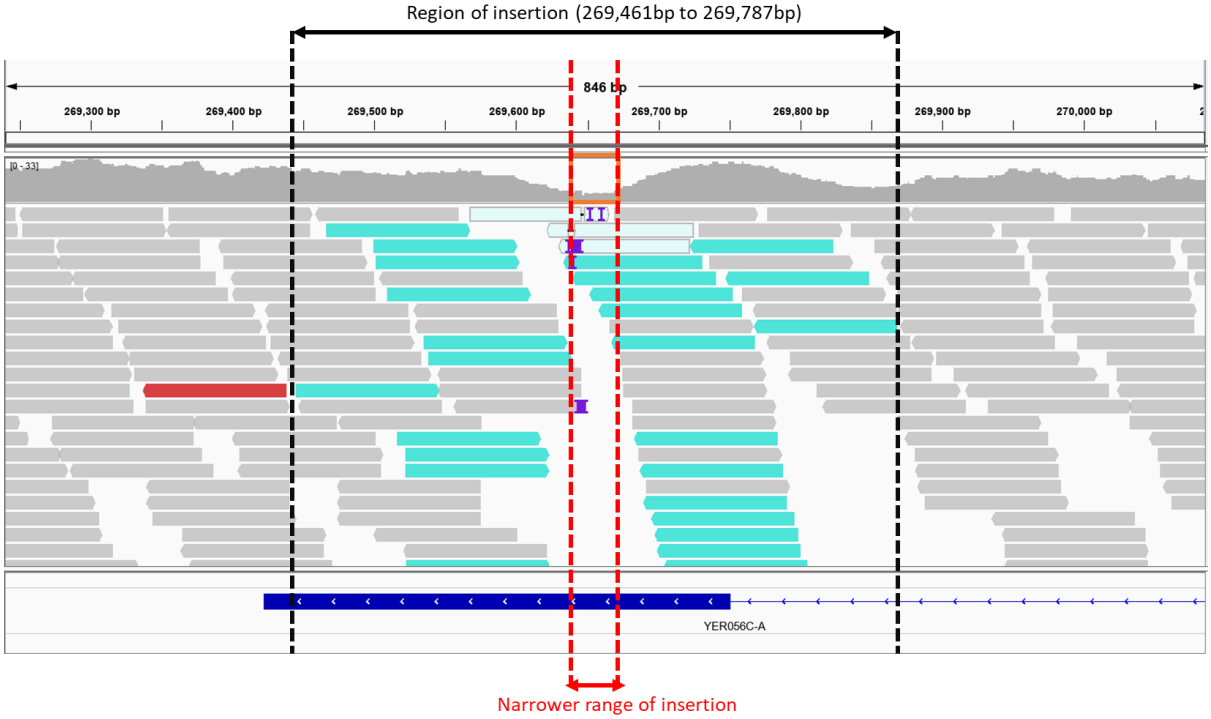


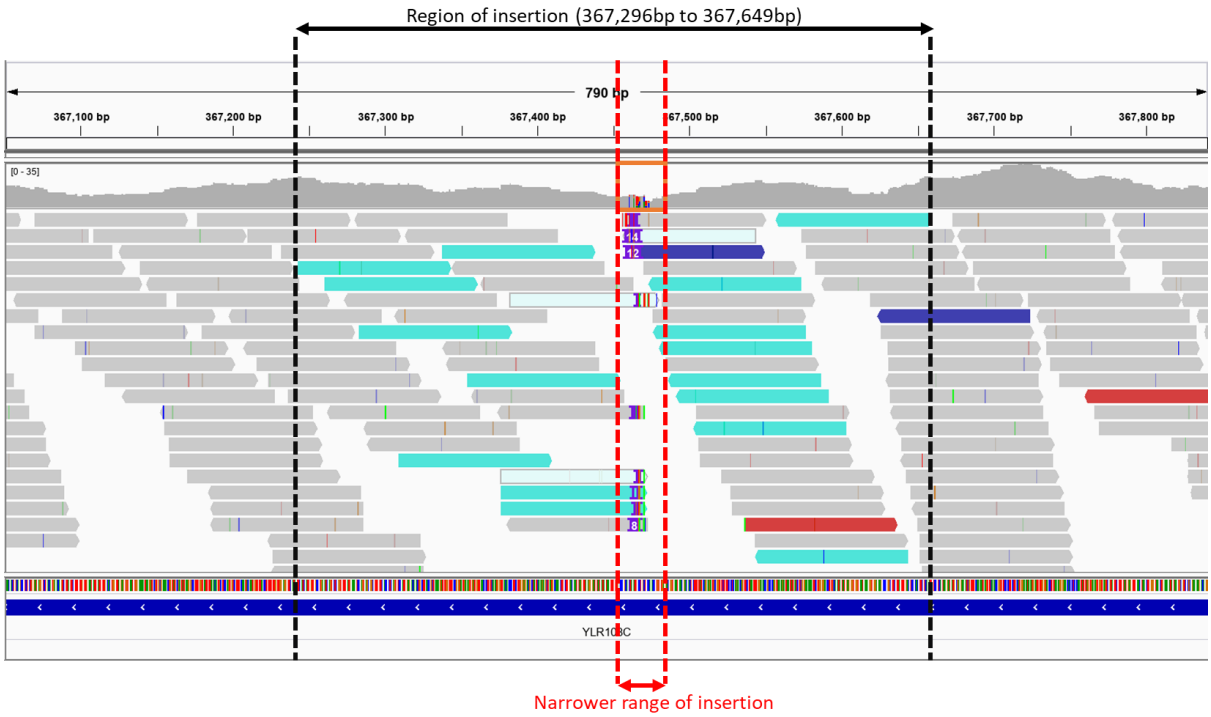
Figure 4: Coverage represented on IGV for (A) chromosome V, (B) chromosome XII, and (C) chromosome XIV, showing only 1 possible insertion for each chromosome

The range of insertion region was determined by taking the furthest start point of cyan coloured read (improper pairing) on both sides, represented by black dotted lines on Figure 5. From the figure, we can conclude that the most probable region of insertion on chromosome V, XII and XIV is 269461bp to 269787bp, 367296bp to 367649bp, and 488606bp to 488999bp on yeast genome respectively. By comparing these regions to the deeps (orange boxes) in Figure 4, a narrower probable range of insertion is represented by red dotted lines, which is roughly in the middle of the broader range (black dotted lines) for each chromosome.

A



B



C

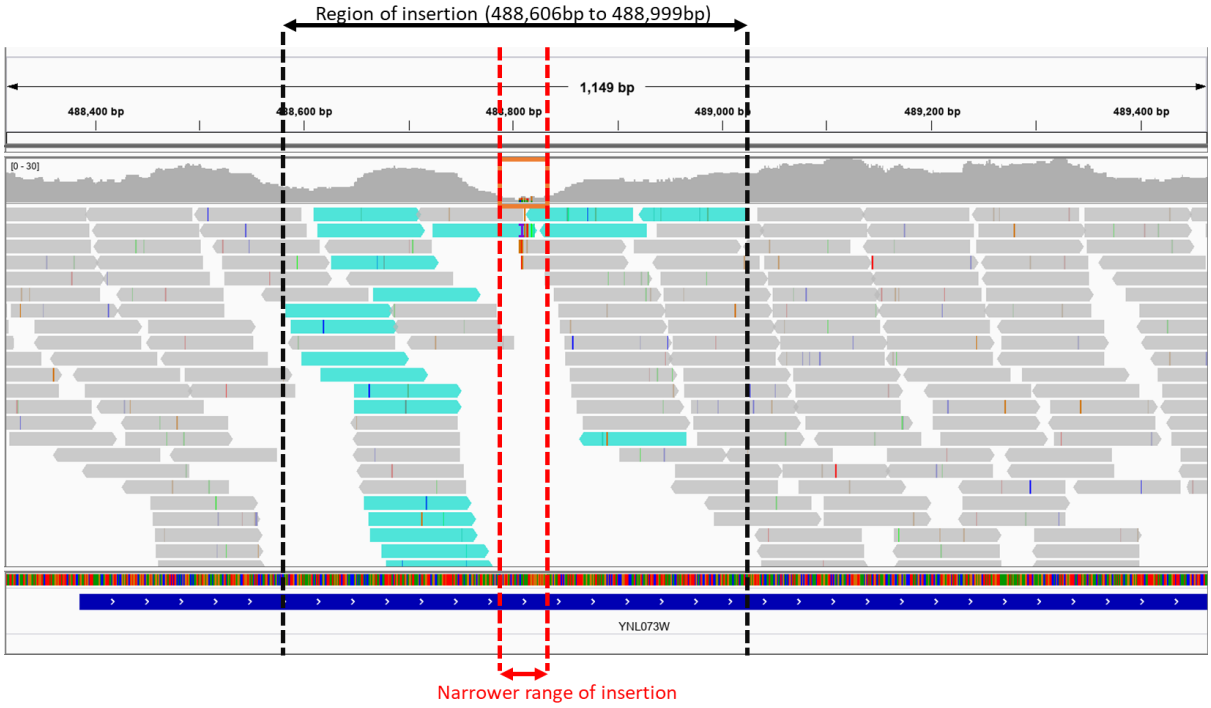


Figure 5: Visualisation of alignment on IGV for (A) chromosome V, (B) chromosome XII, and (C) chromosome XIV, showing the probable region/range of insertion

The orientation of insertions was determined by the combination of properties encoded by SAM flags represented as bitwise FLAG value shown in Tables 4, 5 and 6, containing only the final high-quality reads. If both read and mate are mapped to the reverse(-)/forward(+) strand,

the transposon has a reverse insertion since DNA is synthesised and read from 5' to 3' direction. If the read is mapped to the reverse strand and mate mapped to forward strand, or vice versa, the transposon has a non-reverse/normal insertion. As shown in Table 4, 5, and 6, Ty5-6p transposon was inserted reversely in chromosome V and inserted normally/non-reversely in chromosome XII and XIV.

QNAME	FLAG	MAPQ	Read on which strand	Mate on which strand
chrV-111274	177	23	-	-
chrV-112020	177	42	-	-
chrV-47696	129	42	+	+
chrV-59366	65	35	+	+
chrV-71220	129	40	+	+
chrV-76154	65	42	+	+

Table 4: SAM flags properties of chromosome V

QNAME	FLAG	MAPQ	Read on which strand	Mate on which strand
chrXII-116888	81	42	-	+
chrXII-170778	81	42	-	+
chrXII-60618	97	23	+	-

Table 5: SAM flags properties of chromosome XII

QNAME	FLAG	MAPQ	Read on which strand	Mate on which strand
chrXIV-109550	161	23	+	-
chrXIV-116008	161	23	+	-
chrXIV-154642	145	42	-	+
chrXIV-19604	97	23	-	+
chrXIV-21884	145	42	+	-

Table 6: SAM flags properties of chromosome XIV

3.3 Potential Consequences of Transposon Insertions

As seen from Figure 5 previously, Ty5-6p transposon insertion at chromosome V, XII, and XIV, will disrupt the coding region (exon) of genes named RPL34A (YER056C-A), YLR108C, and MSK1 (YNL073W) respectively.

RPL34A (YER056C-A) is a gene located on chromosome V (269423bp – 270185bp) and on the reverse strand. It codes for ribosomal 60S subunit protein L34A in the budding yeast (Ensembl, 2019). This large subunit of the ribosome has a ribosomal catalytic site called peptidyl transferase center (PTC) which catalyses the formation of peptide bonds, hence forming nascent polypeptide chain by joining amino acids transported by the tRNAs (Ben-Shem et al., 2011). Therefore, when transposon is inserted in this region and disrupt the gene coding region, it may prevent the formation of large ribosomal subunit in yeast strain S288C. This may hinder a complete translation process to occur as the small ribosomal subunit requires

large ribosomal large subunit to synthesise proteins together, which are needed by the yeast cell (Lindahl et al., 2019).

YLR108C is a “predicted” gene found on chromosome XII (366667bp – 368124bp) and on the reverse strand (Ensembl, 2019). It codes for a BTB (BR-C, ttk, and bab)/POZ (Pox virus and Zinc finger) domain-containing hypothetical protein which is of unknown function (Uniprot, 2019). It is also a green fluorescent protein (GFP)-fusion protein localised to the nucleus and increases in abundance in response to DNA replication stress (Saccharomyces Genome Database, 2019). Speculations such as the transposon insertion affecting cellular function of yeast such as transcriptional regulation and cytoskeleton dynamics are plausible due to the known functions of BTB/POZ domain in other organisms (Stogios, Downs, Jauhal, Nandra & Privé, 2005). However, there is still lack of research and experimental evidences to draw any concrete conclusion regarding the consequence of Ty5-6p transposon insertion in this region of the yeast genome (NCBI, 2019).

MSK1 (YNL073W) is a gene found on chromosome XIV in yeast genome, coding for mitochondrial lysine-tRNA synthetase (Saccharomyces Genome Data, 2019). It is responsible for aminoacylation of mitochondrial tRNA for protein synthesis in yeast (Sepuri, Gorla and King, 2012). Therefore, when a Ty5-6p transposon is inserted in this region, this gene may be disrupted and cause decrease in production of mitochondrial lysine-tRNA synthetase. When this happens, less tRNA will be “charged/loaded” to form aminoacyl-tRNA. Hence, ribosomes cannot transfer amino acids from the tRNA to a growing peptide. This hinders the translation process and thus the protein synthesis in yeast (Tarassov, Entelis and Martin, 1995).

4. Conclusion

In conclusion, there is a total of three Ty5-6p transposon insertions identified in the yeast clone given (A0158131M.tgz). Chromosomes V, XII and XIV have one insertion each, at relative mapped positions 269461bp to 269787bp, 367296bp to 367649bp, and 488606bp to 488999bp, on the yeast reference genome (sacCer3) respectively. The orientation of Ty5-6p transposon insertion is reverse insertion on chromosome V and non-reverse (normal/forward) insertions on both chromosomes XII and XIV.

The genes disrupted by transposons insertion at these regions are RPL34A (YER056C-A), YLR108C, and MSK1 (YNL073W) on chromosome V, XII and XIV respectively. RPL34A codes for ribosomal 60S subunit protein L34A responsible for catalysing peptide bond formation during translation process in yeast. YL8108C codes for a BTB/POZ domain-containing hypothetical protein in yeast, which has yet to have concrete known functions. Finally, MSK1 codes for mitochondrial lysine-tRNA synthetase in yeast which is responsible for mitochondrial tRNA aminoacylation during protein synthesis. Therefore, gene disruptions caused by transposon insertion in these regions are likely to affect the protein synthesis process of yeast as both RPL34A and MSK1 code for proteins playing vital role in the translation process. Figure 6 summarises the main findings of this study conducted.

Chromosomes	Number of Insertions	Insertion Locations	Insertion Orientations	Gene Affected	Protein Coded by Gene
V	1	269461bp – 269787bp	Reverse	RPL34A (YER056C-A)	Ribosomal 60S subunit protein L34A
XII	1	367296bp – 367649bp	Non-reverse	YLR108C	BTB/POZ domain-containing hypothetical protein
XIV	1	488606bp – 488999bp	Non-reverse	MSK1 (YNL073W)	Mitochondrial lysine-tRNA synthetase

Figure 6: Summary of findings from this study

Quality control was ensured throughout the process of this study. Data extraction (coding) on Ubuntu was reran multiple rounds to ensure there is no discrepancy in the codes used. Results collected for each round were also compared to ensure consistency, so that no mistakes were made. Quality control for the sequences via FastQC and SAMtools also ensure high quality and accuracy of reads for analysis, which in turn brings about more accurate conclusions. Analysis was also made efficient by optimising the speed needed via tools like Bowtie 2.

However, there are limitations to this study. Firstly, the exact positions of Ty5-6p transposon insertions could not be derived from this study alone. But the exact position of insertion can be guaranteed to be within those ranges concluded previously because it ranges from the start point of furthest improper paired reads from both sides. Another limitation is that no concrete consequence can be drawn about the disruption of gene YLR108C by transposon insertion at chromosome XII. This is due to the lack of research and experimental evidences about this hypothetical protein/“predicted” gene.

Hence, further studies can be done to ensure more accuracy and specificity in pinpointing the exact insertion positions on the chromosomes, as well as, the implications of disrupted YLR108C gene. This can in turn lead to a more holistic and accurate report of Ty5-6p transposon-based mutagenesis in yeast genome of strain S288C.

5. Bibliography

Munoz-Lopez, M., & Garcia-Perez, J. (2010). DNA Transposons: Nature and Applications in Genomics. *Current Genomics*, *11*(2), 115-128. doi: 10.2174/138920210790886871 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2874221/

Xu, T., Bharucha, N., & Kumar, A. (2011). Genome-Wide Transposon Mutagenesis in Saccharomyces cerevisiae and Candida albicans. *Methods In Molecular Biology*, 207-224. doi: 10.1007/978-1-61779-197-0_13 https://www.ncbi.nlm.nih.gov/pubmed/21815095

Botstein, D., & Fink, G. (2011). Yeast: An Experimental Organism for 21st Century Biology. *Genetics*, *189*(3), 695-704. doi: 10.1534/genetics.111.130765 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3213361/

Leggett, R., Ramirez-Gonzalez, R., Clavijo, B., Waite, D., & Davey, R. (2013). Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Frontiers In Genetics*, *4*. doi: 10.3389/fgene.2013.00288 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3865868/

Bolger, A., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. doi: 10.1093/bioinformatics/btu170 https://academic.oup.com/bioinformatics/article/30/15/2114/2390096

Kahveci, T., Ljosa, V., & Singh, A. (2004). Speeding up whole-genome alignment by indexing frequency vectors. *Bioinformatics*, *20*(13), 2122-2134. doi: 10.1093/bioinformatics/bth212 https://academic.oup.com/bioinformatics/article/20/13/2122/242407

Fonseca, N., Rung, J., Brazma, A., & Marioni, J. (2012). Tools for mapping high-throughput sequencing data. *Bioinformatics*, *28*(24), 3169-3177. doi: 10.1093/bioinformatics/bts605 https://academic.oup.com/bioinformatics/article/28/24/3169/245777

Langmead, B., Wilks, C., Antonescu, V., & Charles, R. (2018). Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics*, *35*(3), 421-432. doi: 10.1093/bioinformatics/bty648 https://academic.oup.com/bioinformatics/article/35/3/421/5055585

Carver, T., Harris, S., Otto, T., Berriman, M., Parkhill, J., & McQuillan, J. (2012). BamView: visualizing and interpretation of next-generation sequencing read alignments. *Briefings In Bioinformatics*, *14*(2), 203-212. doi: 10.1093/bib/bbr073 https://academic.oup.com/bib/article/14/2/203/208017

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., & Homer, N. et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. doi: 10.1093/bioinformatics/btp352 https://academic.oup.com/bioinformatics/article/25/16/2078/204688

Fonseca, N., Rung, J., Brazma, A., & Marioni, J. (2012). Tools for mapping high-throughput sequencing data. Bioinformatics, 28(24), 3169-3177. doi: 10.1093/bioinformatics/bts605 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3614465

Ruffalo, M., Koyuturk, M., Ray, S., & LaFramboise, T. (2012). Accurate estimation of short read mapping quality for next-generation genome sequencing. *Bioinformatics*, *28*(18), i349-i355. doi: 10.1093/bioinformatics/bts408 https://academic.oup.com/bioinformatics/article/28/18/i349/249968

Thorvaldsdottir, H., Robinson, J., & Mesirov, J. (2012). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings In Bioinformatics*, *14*(2), 178-192. doi: 10.1093/bib/bbs017 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3603213/

Bioinformatics.babraham.ac.uk. (2019). Kmer Content. [online] Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3%20Analysis%20Modules/11%20Kmer%20Content.html [Accessed 31 Mar. 2019].

Gene: RPL34A (YER056C-A) - Summary - Saccharomyces cerevisiae - Ensembl genome browser 95. (2019). Retrieved from http://asia.ensembl.org/Saccharomyces_cerevisiae/Gene/Summary?db=core;g=YER056C-A;r=V:269423-270185;t=YER056C-A_mRNA

Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., & Yusupov, M. (2011). The Structure of the Eukaryotic Ribosome at 3.0 A Resolution. *Science*, *334*(6062), 1524-1529. doi: 10.1126/science.1212642 http://science.sciencemag.org/content/334/6062/1524

Lindahl, L., Gregory, B., Rahman, N., Bommakanti, A., Shamsuzzaman, M., & Thapa, M. et al. (2019). The small and large ribosomal subunits depend on each other for stability and accumulation. doi: http://dx.doi.org/10.1101/384362

Gene: YLR108C - Summary - Saccharomyces cerevisiae - Ensembl genome browser 95. (2019). Retrieved from http://asia.ensembl.org/Saccharomyces_cerevisiae/Gene/Summary?g=YLR108C;r=XII:366667-368124;t=YLR108C_mRNA

YLR108C - BTB/POZ domain-containing protein YLR108C - Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast) - YLR108C gene & protein. (2019). Retrieved from https://www.uniprot.org/uniprot/Q12259

YLR108C | SGD. (2019). Retrieved from https://www.yeastgenome.org/locus/S000004098

Stogios, P., Downs, G., Jauhal, J., Nandra, S., & Privé, G. (2005). *Genome Biology*, *6*(10), R82. doi: 10.1186/gb-2005-6-10-r82 https://genomebiology.biomedcentral.com/articles/10.1186/gb-2005-6-10-r82

YLR108C hypothetical protein [Saccharomyces cerevisiae S288C] - Gene - NCBI. (2019). Retrieved from https://www.ncbi.nlm.nih.gov/gene/?term=850798

Sepuri, N., Gorla, M. and King, M. (2012). Mitochondrial Lysyl-tRNA Synthetase Independent Import of tRNA Lysine into Yeast Mitochondria. *PLoS ONE*, 7(4), p.e35321. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3335127/

Tarassov, I., Entelis, N. and Martin, R. (1995). Mitochondrial import of a cytoplasmic lysine-tRNA in yeast is mediated by cooperation of cytoplasmic and mitochondrial lysyl-tRNA synthetases. *The EMBO Journal*, 14(14), pp.3461-3471. https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1460-2075.1995.tb07352.x

6. Appendix

6.1 FastQC

Full FastQC Reports:
file:///C:/Users/Tan%20Wei%20Qi/Desktop/CA2/results/fastqc/A0158131M_1_fastqc.html
file:///C:/Users/Tan%20Wei%20Qi/Desktop/CA2/results/fastqc/A0158131M_2_fastqc.html

A0158131M_1:

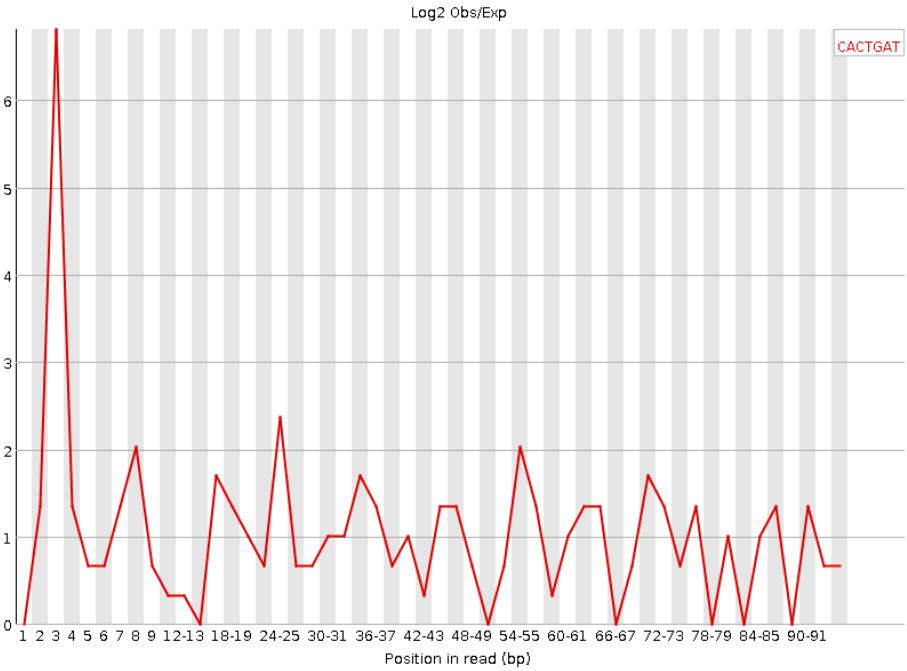
Summary

- Basic Statistics
- Per base sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content
- Kmer Content

Basic Statistics

Measure	Value
Filename	A0158131M_1.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1208690
Sequences flagged as poor quality	0
Sequence length	100
%GC	38

Kmer Content



Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
CACTGAT	690	0.0073325736	6.8115945	3

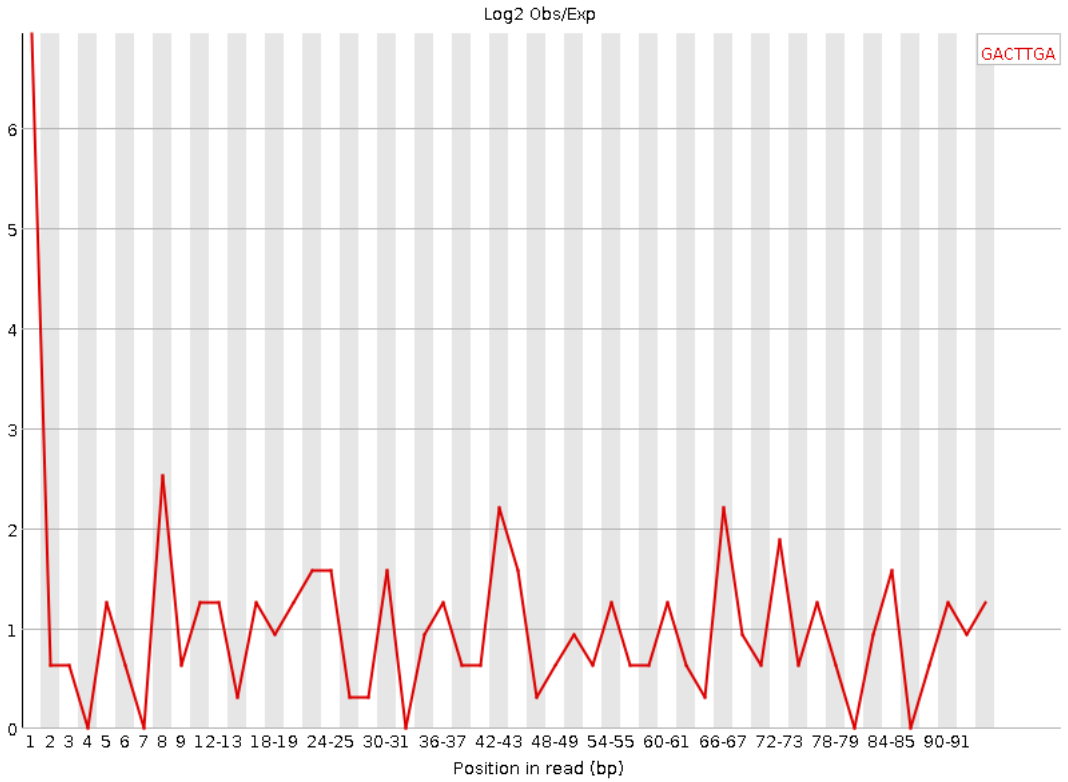
A0158131M_2:

Summary

Basic Statistics

Measure	Value
Filename	A0158131M_2.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1208690
Sequences flagged as poor quality	0
Sequence length	100
%GC	38

Kmer Content



Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
GACTTGA	745	0.0020016334	6.939597	1

6.2 Full Data and Filtered Data

CSV files of data: https://github.com/Tan-WeiQi/LSM3241-CA2/blob/master/CSV%20files%20of%20all%20data.csv

Filtered CSV file of high-quality data: https://github.com/Tan-WeiQi/LSM3241-CA2/blob/master/Filtered%20CSV%20file%20of%20high-quality%20data.csv

6.3 IGV Alignments for Filtered BAM File

Chromosome V

Chromosome XII

Chromosome XIV