UNIVERSITI SAINS MALAYSIA
GERAN PENYELIDIKAN UNIVERSITI PENYELIDIKAN
LAPORAN AKHIR

THE STUDY OF TRANSGENE INTERGRATION HOTSPOTS IN MAMMALIA EXPRESSION SYSTEM

PENYELIDIK

PROFESOR MADYA DR. SHAHARUM SHAMSUDDIN

PENYELIDIK BERSAMA

KHOO JO LYNN

2015

F0555

ORANG FRGS – P3(R)

Kod Projek :	FRGS/FASA1-2009/(BIDANG)/(NAMA IPT)/(NO.RUJ. KPT)

FINAL REPORT
FUNDAMENTAL RESEARCH GRANT SCHEME (FRGS)
Laporan Akhir Skim Geran Penyelidikan Asas (FRGS) IPT
Pindaan 1/2009

RESEARCH TITLE : **The study of transgene integration hotspots in mammalian expression system**
Tajuk Penyelidikan — *Kajian terhadap spot panas integrasi `transgene' dalam sistem pengekspresian mamalia*

3

PROJECT LEADER : **Associate Prof. Shaharum Shamsuddin**
Ketua Projek

PROJECT MEMBERS : 1. **Khoo Jo Lynn (MSc – Graduated Sept 2012)**
(including GRA) 2.
Ahli Projek

ROJECT ACHIEVEMENT *(Prestasi Projek)*

ACHIEVEMENT PERCENTAGE			
Project progress according to milestones achieved up to this period	**0 - 50%**	**51 - 75%**	**76 - 100%**
Percentage			X

RESEARCH FINDINGS		
Number of articles/ manuscripts/ books	**Indexed Journal**	***Non-Indexed Journal***
	1 (MJMS - submitted)	
Paper presentations	**International**	***National***
		1
Others (Please specify)		

HUMAN CAPITAL DEVELOPMENT			
Human Capital	**Number**		Others (Please specify): **MSc thesis :** *`Characterization of transgene integration hotspot in NS0 (Non-secreting) Myeloma Cells expression system'*
	On-going	**Graduated**	
PhD Student			
Masters Student		1	
Undergraduate Students			
Temporary Research Officer			
Temporary Research Assistant			
Total	1		

EXPENDITURE *(Perbelanjaan)*

C

ADDITIONAL RESEARCH ACTIVITIES THAT CONTRIBUTE TOWARDS DEVELOPING SOFT AND HARD SKILLS
(Aktiviti Penyelidikan Sampingan yang menyumbang kepada pembangunan kemahiran insaniah)

D

International

Activity	Date (Month, Year)	Organizer
(e.g : Course/ Seminar/ Symposium/ Conference/ Workshop/ Site Visit)	-	-

National

Activity	Date (Month, Year)	Organizer
(e.g : Course/ Seminar/ Symposium/ Conference/ Workshop/ Site Visit)	Symposium (Dec, 2011)	UMS, Kota Kinabalu

PROBLEMS / CONSTRAINTS IF ANY *(Masalah/ Kekangan sekiranya ada)*

E None

RECOMMENDATION (*Cadangan Penambahbaikan*)

F

A number of putative MARs have been identified in this study. Although *in vitro* and/or *in vivo* verific
needs to be done to identify the true MARs, these MARs would be useful tools in the future.
This may include :

1) Develop an expression vector that consists of a true Matrix Attachment Region along with a protei
increase the yield of the protein.
2) Determine the mechanism of integration of transgenes.
3) Facilitate the design of a site – specific integration expression vector

RESEARCH ABSTRACT – Not More Than 200 Words*(Abstrak Penyelidikan – Tidak Melebihi 200 patah perkataan)*

G

Mammalian expression system is very crucial for the production of proteins in numerous scientific and commercial areas as they are able to perform proper protein folding as compared to other hosts. Matrix attachment regions (MARs) are concentrated with transcription factor binding sites and have a strong effect on the level of expression of transgenes, thus highly studied to increase protein production. In this study, an expression vector carrying green fluorescent protein was transfected into mouse myeloma NS0 cell line in order to allow random integration of the vector into the host genome. Conceptually, if the vector is integrated into a transcriptionally active site within the genome, green fluorescent protein will be highly or constitutively expressed. Transfected cells were then subjected to sorting in order to select the high producing cells using ClonePix FL (CPL). CPL is an automated high–throughput clone selection instrument. DNA sequences in the genome flanking the integrated vector in the high producing cells were subsequently retrieved using 'genome walking' method. Following this, the flanking regions were sequenced and characterized using bioinformatics tools. Functional analysis was carried out by entering the sequences into three different programs (SMARtest, MARfinder and MARscan) to determine the presence of Matrix Attachment Regions. In addition, biochemical characterization was also done using Chromatin Immunoprecipitation (ChIP) Assay. As a result, putative MAR elements were identified in mammalian expression system with different lengths of MAR regions between the three algorithms. MAR binding proteins that were used in ChIP assay were also shown to interact with the mouse genome.

Date
Tarikh

COMMENTS, IF ANY/ ENDORSEMENT BY RESEARCH MANAGEMENT CENTER (RMC)
(Komen, sekiranya ada/ Pengesahan oleh Pusat Pengurusan Penyelidikan)

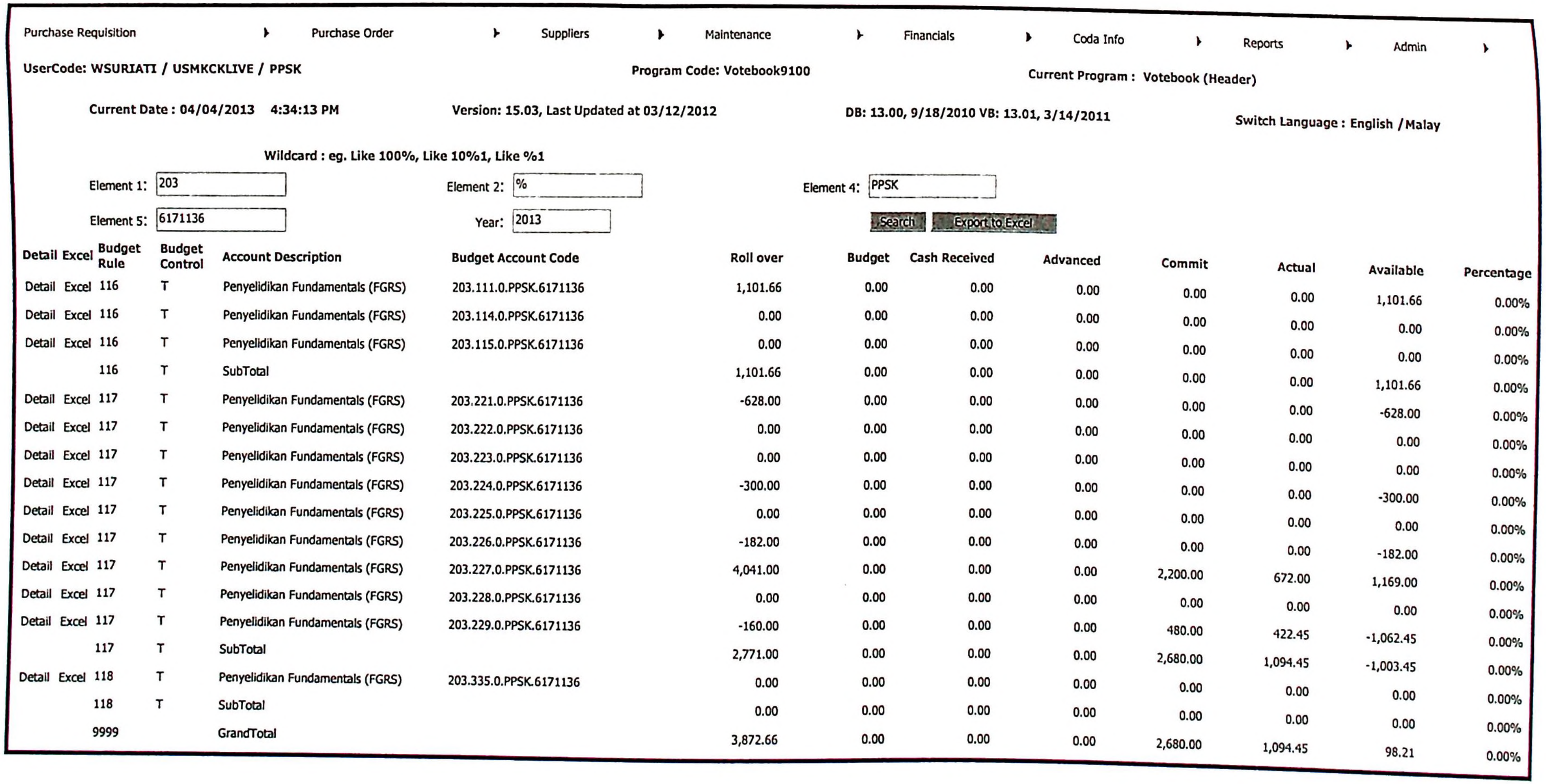
H ....

....

....

Name:
Nama:

Date:
Tarikh:



UserCode: WSURIATI / USMKCKLIVE / PPSK Program Code: Votebook9100 Current Program : Votebook (Header)

Current Date : 04/04/2013 4:34:13 PM Version: 15.03, Last Updated at 03/12/2012 DB: 13.00, 9/18/2010 VB: 13.01, 3/14/2011 Switch Language : English / Malay

Wildcard : eg. Like 100%, Like 10%1, Like %1

Element 1: 203 Element 2: % Element 4: PPSK

Element 5: 6171136 Year: 2013

Detail	Excel	Budget Rule	Budget Control	Account Description	Budget Account Code	Roll over	Budget	Cash Received	Advanced	Commit	Actual	Available	Percentage
Detail	Excel	116	T	Penyelidikan Fundamentals (FGRS)	203.111.0.PPSK.6171136	1,101.66	0.00	0.00	0.00	0.00	0.00	1,101.66	0.00%
Detail	Excel	116	T	Penyelidikan Fundamentals (FGRS)	203.114.0.PPSK.6171136	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
Detail	Excel	116	T	Penyelidikan Fundamentals (FGRS)	203.115.0.PPSK.6171136	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
		116	T	SubTotal		1,101.66	0.00	0.00	0.00	0.00	0.00	1,101.66	0.00%
Detail	Excel	117	T	Penyelidikan Fundamentals (FGRS)	203.221.0.PPSK.6171136	-628.00	0.00	0.00	0.00	0.00	0.00	-628.00	0.00%
Detail	Excel	117	T	Penyelidikan Fundamentals (FGRS)	203.222.0.PPSK.6171136	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
Detail	Excel	117	T	Penyelidikan Fundamentals (FGRS)	203.223.0.PPSK.6171136	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
Detail	Excel	117	T	Penyelidikan Fundamentals (FGRS)	203.224.0.PPSK.6171136	-300.00	0.00	0.00	0.00	0.00	0.00	-300.00	0.00%
Detail	Excel	117	T	Penyelidikan Fundamentals (FGRS)	203.225.0.PPSK.6171136	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
Detail	Excel	117	T	Penyelidikan Fundamentals (FGRS)	203.226.0.PPSK.6171136	-182.00	0.00	0.00	0.00	0.00	0.00	-182.00	0.00%
Detail	Excel	117	T	Penyelidikan Fundamentals (FGRS)	203.227.0.PPSK.6171136	4,041.00	0.00	0.00	0.00	2,200.00	672.00	1,169.00	0.00%
Detail	Excel	117	T	Penyelidikan Fundamentals (FGRS)	203.228.0.PPSK.6171136	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
Detail	Excel	117	T	Penyelidikan Fundamentals (FGRS)	203.229.0.PPSK.6171136	-160.00	0.00	0.00	0.00	480.00	422.45	-1,062.45	0.00%
		117	T	SubTotal		2,771.00	0.00	0.00	0.00	2,680.00	1,094.45	-1,003.45	0.00%
Detail	Excel	118	T	Penyelidikan Fundamentals (FGRS)	203.335.0.PPSK.6171136	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
		118	T	SubTotal		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
		9999		GrandTotal		3,872.66	0.00	0.00	0.00	2,680.00	1,094.45	98.21	0.00%

TECHNICAL REPORT

CHARACTERIZATION OF TRANSGENE INTEGRATION HOTSPOTS IN MAMMALIAN EXPRESSION SYSTEM

Project Code : 203/PPSK/6171136

Table of Contents

Problem Statement & Objectives

Hotspot regions in the genome that permit integration of transgene share common properties in their sequences such as conserved regions or protein binding sites. In this study we intended :

1) To determine the integration site (hotspot) for a transgene in the mammalian cell in which could enhance recombination protein expression.
2) To characterize the flanking region of the integration site.
3) To biochemically characterize the interaction between the matrix attachment region's binding proteins with the genome

Research Flowchart

Transformation of phrGFP II - I® into *E. coli* for propagation.

Transfection of the linearized phrGFP II - I® into mammalian cell.

Clonal selection for the positive clones.

ClonePix to sort out high producer clones.

Mapping of transgene flanking region via genome walking.

Perform DNA cloning and sequencing. Sequence analysis via Clustal W and BLAST to identify the hotspot.

Virtual walking up to 100 Kb via NCBI.

Functional characterisation of sequences using bioinformatics tools to determine the presence of Matrix Attachment Regions (MARs).

***In vitro* biochemical characterisation using Chromatin Immunoprecipitation Assay.**

Findings & Research Results

Introduction of Vitality® phrGFP II – I Vector into Mouse Myeloma NS0 Cells

Transfection was done using FuGENE® HD Transfection Reagent (Roche, USA) with some modifications to the manufacturer's protocol. Transfection was carried out in 6 – well plates with 3×10^5 cells/ml in each well. After transfection was done, cells were transferred to T – 25 flasks, resuspended with growth media with antibiotic selection and checked under microscope every day. Mouse myeloma NS0 cells are semi – adherent cells. As so, the cells will naturally adhere at the surface of the flask. Although some of the viable cells will also be suspended in the media, all of the non – viable cells will be suspended in the media. Figure 1 depicts the viable and non – viable cells one week after transfection. As the flasks were checked every day, it was observed that cells in the negative control flasks gradually died off as each day passed. This is because cells in negative control flasks were not inserted with the Vitality® phrGFP II – I vector which contains an antibody resistance gene.

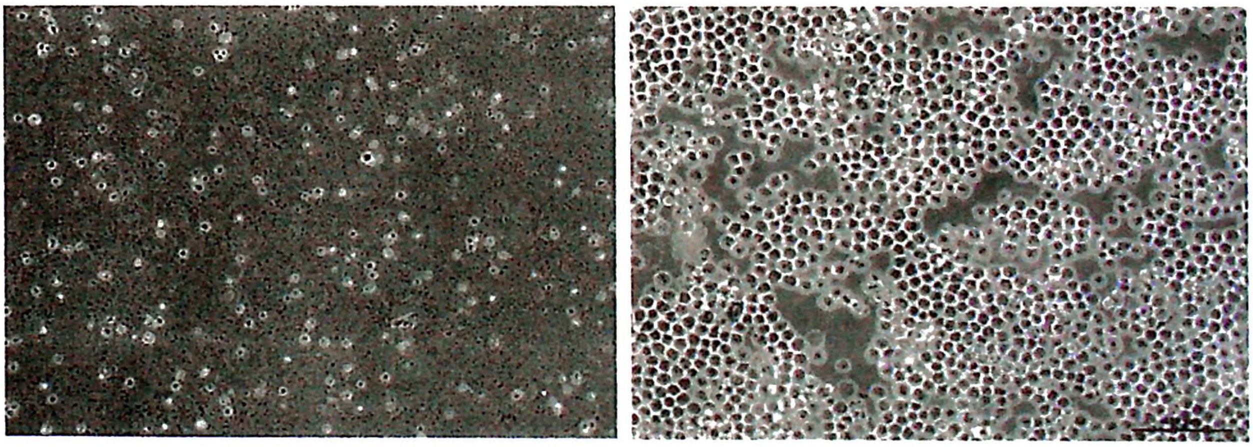


Figure 1 : The left image represents the non – viable cells from the negative control and the right image represents a mixture of viable and non – viable cells from transfection (under 10× magnification). Viable cells are assumed to be transfected following the death of all the cells in the negative control flasks.

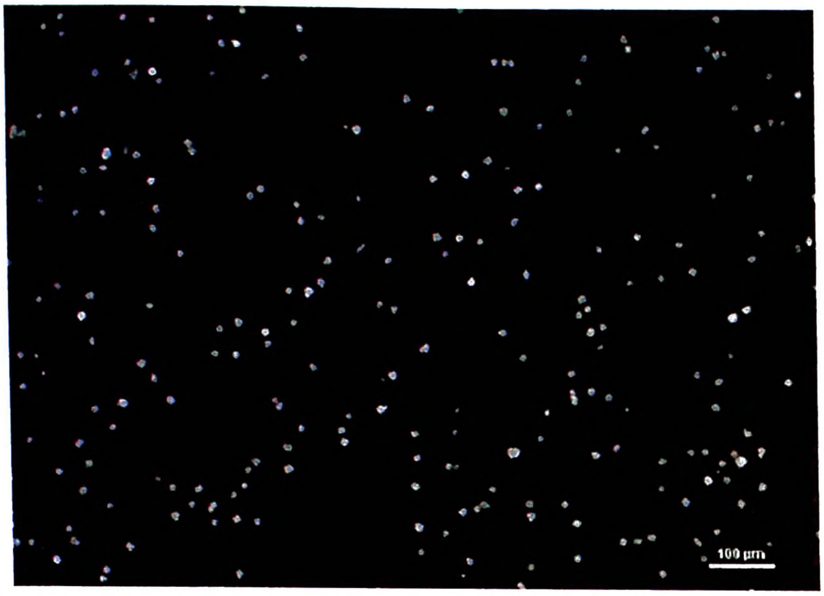


Figure 2 : Transfected cells imaged under fluorescent microscope (10 × magnification).

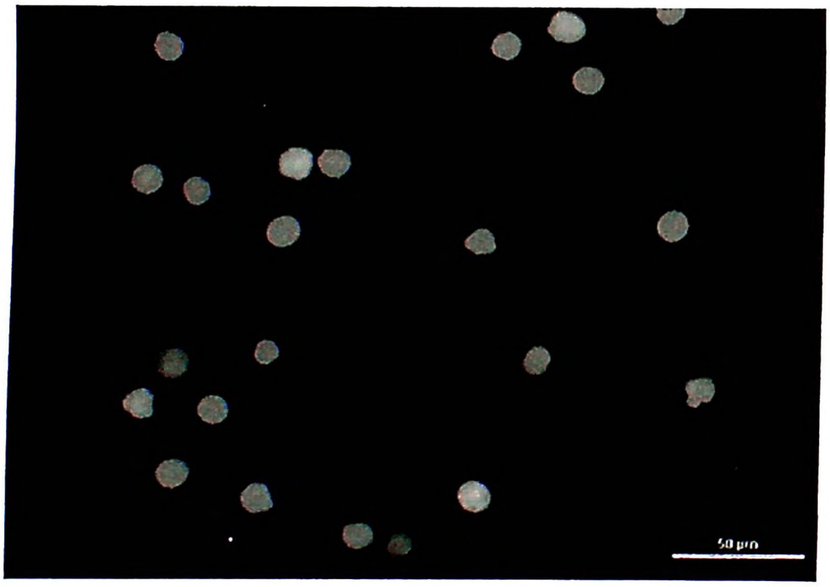


Figure 3 : Transfected cells imaged under fluorescent microscope (40 × magnification).

Sorting of High Producer Clones using ClonePix^{TM} FL

In this study, ClonePix^{TM} FL was used to pick for high – producing clones. Each of the Genetix PetriWell – 6 plates was first screened under white light and subsequently under FITC exposure at 200 milliseconds. Figure 4 represents colonies that were screened under white light and FITC exposure using ClonePix^{TM} FL. Colonies that were shown under white light are the total number of colonies which grow in the well, whereas colonies that were shown under FITC exposure are the colonies which are integrated with GFP gene. Each of the colonies shown under FITC exposure has different intensity of FITC intensity. From the two images taken under white light and FITC exposure, colonies which are integrated with the GFP gene can be distinguished from the colonies that are not integrated with the GFP gene. After high – producing clones were picked and transferred to 96 well plate, the colonies that are able to grown in the 96 well plate were then scaled up and expanded by transferring the clones to 24 well plate, 6 well plate, T – 25 and T – 75 flasks. While clones were in T – 75 flasks, they were seeded into semi – solid media once again and screened using ClonePix^{TM} FL to check for its homogeneity/monoclonality. Figure 5 shows the images of the second round of screening for randomly selected clones. All of the colonies that were imaged under white light were also seen to be imaged under FITC exposure at 200 milliseconds. From these images, it can be concluded that colonies picked by ClonePix^{TM} FL are homogenous or have a clonality of more than 99 %. The stability of clones was determined by the fluorescence of GFP. Clones that were not stable were discarded and the stable clones were used for subsequent studies. Figure 6 depicts the fluorescent of GFP at passage 25 under fluorescent microscope with 10 × magnification. The cells that are shown here are still exhibiting fluorescence at passage 25. As so, it can be concluded that the clones are stable clones as the cells are still producing GFP.

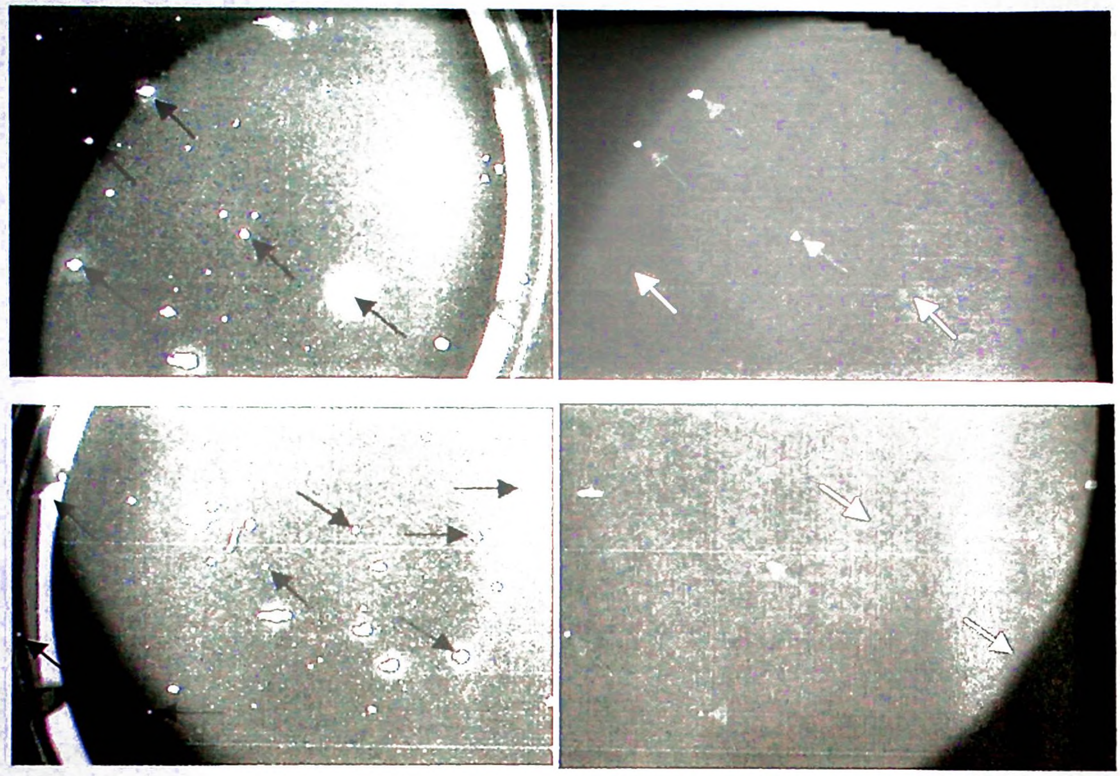


Figure 4 : The left diagram represents the colonies of cells imaged under white light. The right diagram represents the colonies imaged under FITC exposure. Each green spot imaged under FITC exposure represents colonies that are integrated with GFP gene, each with different intensity of FITC signal. The comparison images taken under white light and fluorescent exposure enables us to directly identify colonies with (red arrows) and without (white arrows) the integrated GFP gene.

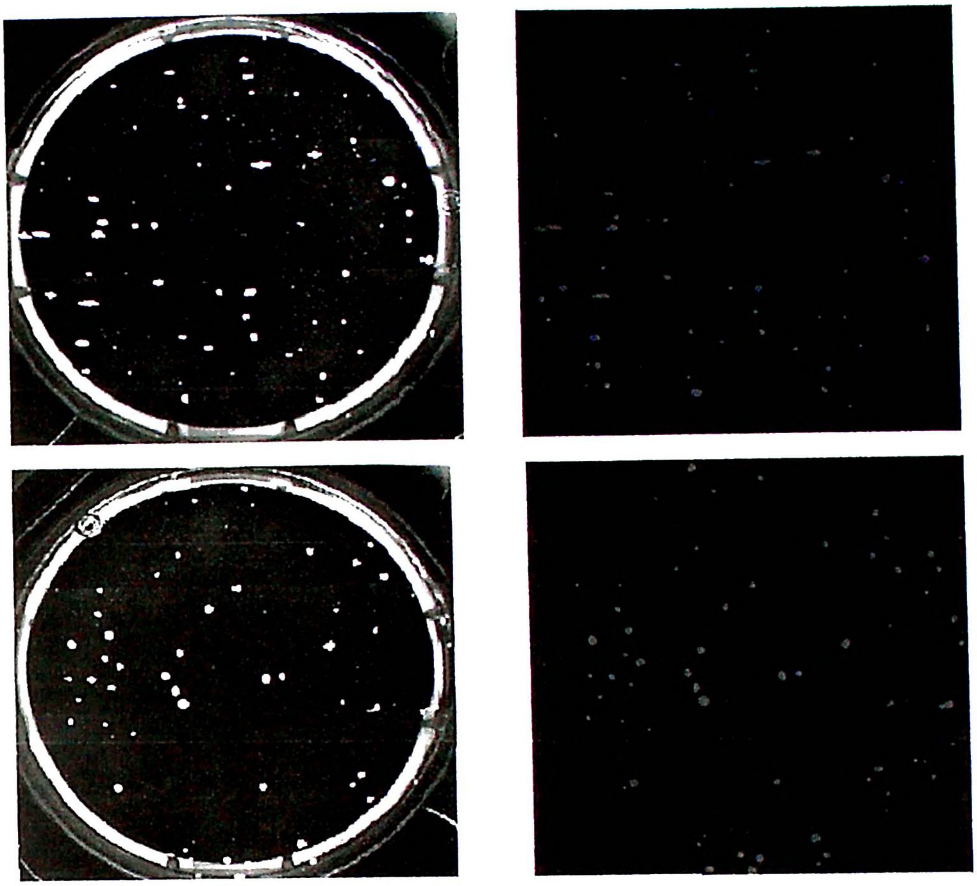


Figure 5 : Representative clones in the left diagram were imaged under white light, whereas the representative clones in the right diagram were imaged under FITC exposure. These representative colonies of cells originated from a single parental which was previously picked by ClonePix^{TM} FL and then re – plated into 96-well plate. The cell populations were then screened again by ClonePix^{TM} FL to demonstrate homogeneity. This study proved that the ClonePix^{TM} FL picked single parental colony is homogenous or having monoclonality of more than 99 %.

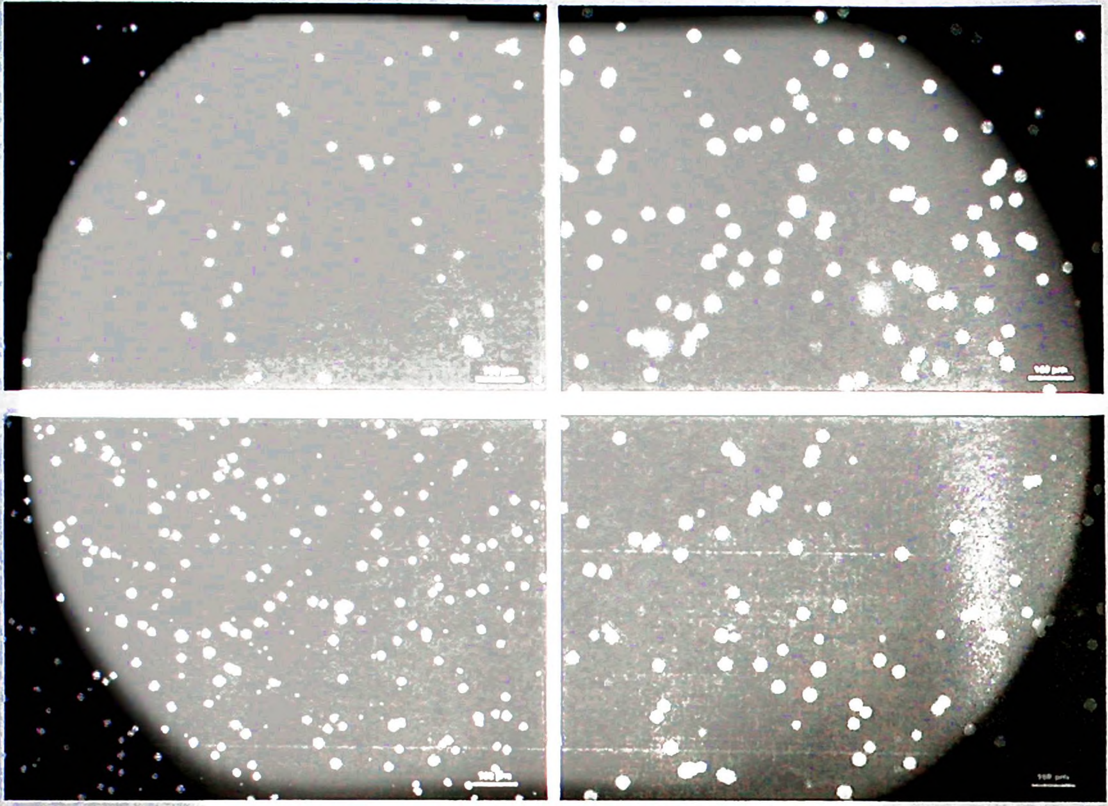


Figure 6 : Fluorescence of GFP of stable representative clones picked by ClonePix^{TM} FL at passage 25 under fluorescent microscope with 10 × magnification.

Mapping of Transgene Flanking Region via DNA Walking

DNA walking was done to map out the transgene flanking region. DNA walking was carried out by using DNA Walking *SpeedUp*^{TM} Premix Kit II (Seegene, Korea) using manufacturer's protocol. Primers were designed based on the sequences of hrGFP with the help of Primer3 software and also NCBI Blast. Target – specific primers were designed in sets of three primers as DNA walking uses nested polymerase chain reaction (PCR) method. Each set of primer was designed in a way that the last primer used for the nested PCR was situated approximately 50 – 200 bp from the end of the hrGFP sequence. Each of the primers designed was then blasted against NCBI Blast to verify that the primers have no similarities to mouse genome before synthesizing the primers.

A total of 3 polymerase chain reactions (PCR) are needed for one cycle of DNA walking. The second and third PCR for each DNA walking are nested PCR which involve the nested primers to bind or amplify a secondary target within the first PCR product and produce a new PCR product that is shorter than the first product.

It was shown that all clones for Part I DNA walking did not show any bands when agarose gel electrophoresis was done. This also includes the positive and negative controls. The unavailability of visible bands might be caused by the size of the amplified region. The amplified region of Part I DNA walking might be too small in size to be visualised on the agarose gel (Seegene Incorporated, 2006). However, after Part I DNA walking products were purified and used as template for the Part II DNA walking, visible bands were observed for all clones including the positive control, with the exception of negative control. Although duplicates were done for each of the clones, there were differences in the length of amplified regions for the duplicates. This occurrence is due to the random primers that were used. As the random primers are short (4 – 5 bp) and non – specific, these primers can anneal to any region of the genome, thus producing different lengths of amplified regions.

Most of the results from Part III DNA walking did not show any visible bands when agarose gel electrophoresis was done, with the exception of positive control. Eventually, a few of the samples that only showed visible bands in Part II DNA walking were cloned into a cloning vector, plasmid isolated and sequenced. From the sequencing results, it was observed that part of the hrGFP II Open Reading Frame (ORF) sequence was cut off and replaced by mouse genome. In cases where no visible bands were observed in Part III DNA walking, products from Part II DNA walking were used for further studies. After DNA walking was done, visible bands that were seen on the agarose gel were extracted, cloned and sequenced.

Lane M : 1kb DNA ladder

Lane 1: First PCR product for clone 2A1 sample 1

Lane 2 : First PCR product for clone 2A1 sample 2

Lane 3 : First PCR product for clone 2A2 sample 1

Lane 4 : First PCR product for clone 2A2 sample 2

Lane 5 : First PCR product for clone 2A3 sample 1

Lane 6 : First PCR product for clone 2A3 sample 2

Lane 7 : First PCR product for clone 2A4 sample 1

Lane 8 : First PCR product for clone 2A4 sample 2

Lane 9 : First PCR product for clone 2A5 sample 1

Lane 10 : First PCR product for clone 2A5 sample 2

Lane 11 : First PCR product for clone 2A6 sample 1

Lane 12 : First PCR product for clone 2A6 sample 2

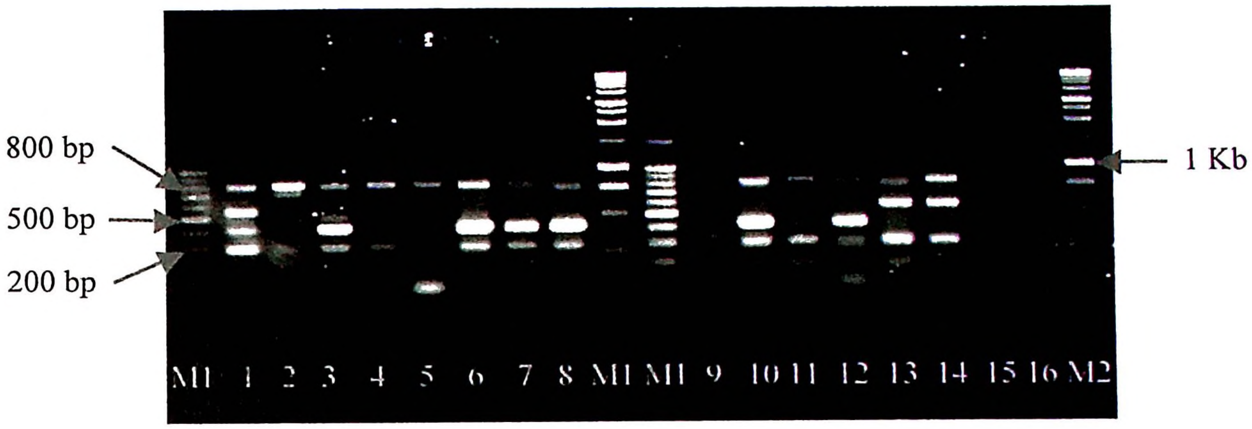
Lane 13 : First PCR product for negative control sample 1

Lane 14 : First PCR product for negative control sample 2

Lane 15 : First PCR product for positive control sample 1

Lane 16 : First PCR product for positive control sample 2

Figure 7 : Analytical 1.0 % (w/v) agarose gel electrophoresis for first PCR product of DNA walking.



Lane M1 : 100 bp DNA ladder

Lane M2 : 1 kb DNA ladder

Lane 1 and 2 : Second PCR product for clone 2A1

Lane 3 and 4 : Second PCR product for clone 2A2

Lane 5 and 6 : Second PCR product for clone 2A3

Lane 7 and 8 : Second PCR product for clone 2A4

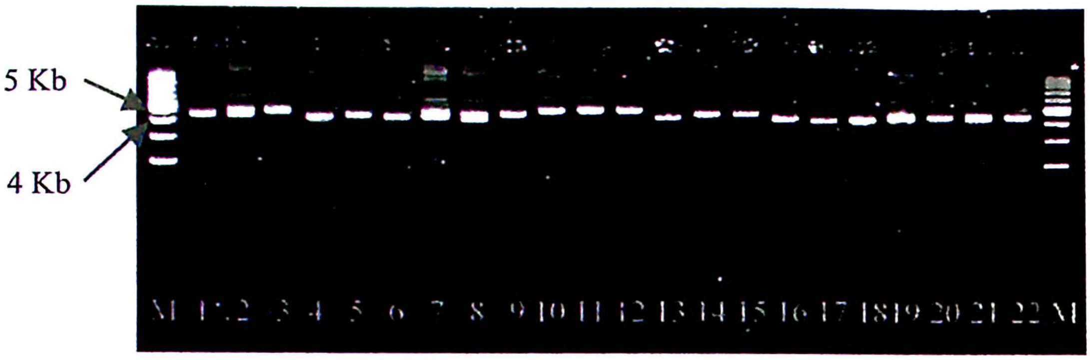
Lane 9 and 10 : Second PCR product for clone 2A5

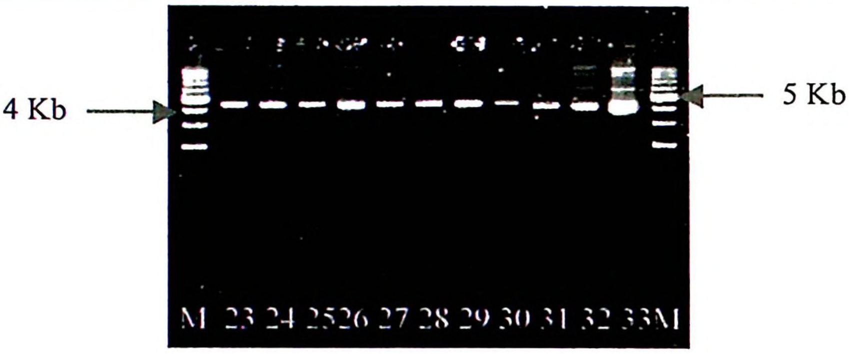
Lane 11 and 12 : Second PCR product for clone 2A6

Lane 13 and 14 : Second PCR product for positive control

Lane 15 and 16: Second PCR product for negative control

Figure 8 : Analytical 1.0 % (w/v) agarose gel electrophoresis for second PCR product of DNA walking.





Lane M : Supercoil DNA ladder

Lane 1 : Isolated plasmid for clone 2A1 sample 1 band 1 from DNA walking part II

Lane 2 : Isolated plasmid for clone 2A1 sample 1 band 2 from DNA walking part II

Lane 3 : Isolated plasmid for clone 2A1 sample 1 band 3 from DNA walking part II

Lane 4 : Isolated plasmid for clone 2A1 sample 1 band 4 from DNA walking part II

Lane 5 : Isolated plasmid for clone 2A1 sample 2 band 1 from DNA walking part II

Lane 6 : Isolated plasmid for clone 2A2 sample 1 band 1 from DNA walking part II

Lane 7 : Isolated plasmid for clone 2A2 sample 1 band 2 from DNA walking part II

Lane 8 : Isolated plasmid for clone 2A2 sample 1 band 3 from DNA walking part II

Lane 9 : Isolated plasmid for clone 2A2 sample 1 band 4 from DNA walking part II

Lane 10 : Isolated plasmid for clone 2A2 sample 2 band 1 from DNA walking part II

Lane 11 : Isolated plasmid for clone 2A2 sample 2 band 2 from DNA walking part II

Lane 12 : Isolated plasmid for clone 2A3 sample 1 band 1 from DNA walking part II

Lane 13 : Isolated plasmid for clone 2A3 sample 1 band 2 from DNA walking part II

Lane 14 : Isolated plasmid for clone 2A3 sample 2 band 1 from DNA walking part II

Lane 15 : Isolated plasmid for clone 2A3 sample 2 band 2 from DNA walking part II

Lane 16 : Isolated plasmid for clone 2A3 sample 2 band 3 from DNA walking part II

Lane 17 : Isolated plasmid for clone 2A4 sample 1 band 1 from DNA walking part II
Lane 18 : Isolated plasmid for clone 2A4 sample 1 band 2 from DNA walking part II
Lane 19 : Isolated plasmid for clone 2A4 sample 1 band 3 from DNA walking part II
Lane 20 : Isolated plasmid for clone 2A4 sample 2 band 1 from DNA walking part II
Lane 21 : Isolated plasmid for clone 2A4 sample 2 band 2 from DNA walking part II
Lane 22 : Isolated plasmid for clone 2A4 sample 2 band 3 from DNA walking part II
Lane 23 : Isolated plasmid for clone 2A5 sample 2 band 1 from DNA walking part II
Lane 24 : Isolated plasmid for clone 2A5 sample 2 band 2 from DNA walking part II
Lane 25 : Isolated plasmid for clone 2A5 sample 2 band 3 from DNA walking part II
Lane 26 : Isolated plasmid for clone 2A6 sample 1 band 1 from DNA walking part II
Lane 27 : Isolated plasmid for clone 2A6 sample 1 band 2 from DNA walking part II
Lane 28 : Isolated plasmid for clone 2A6 sample 1 band 3 from DNA walking part II
Lane 29 : Isolated plasmid for clone 2A6 sample 1 band 4 from DNA walking part II
Lane 30 : Isolated plasmid for clone 2A6 sample 2 band 1 from DNA walking part II
Lane 31 : Isolated plasmid for clone 2A6 sample 2 band 2 from DNA walking part II
Lane 32 : Isolated plasmid for clone 2A6 sample 2 band 3 from DNA walking part II
Lane 33 : Isolated plasmid for clone 2A6 sample 2 band 4 from DNA walking part II

Figure 9 : Analytical 0.8 % (w/v) agarose gel electrophoresis for plasmid isolation.

Sequence Analysis

BioEdit program was used to identify the inserts after acquiring the results from DNA sequencing. Inserts were then aligned against the target – specific primer with the remaining GFP gene specific sequence using ClustalW Multiple Alignment option. . Figure 10 shows the alignment of sequences from DNA sequencing with the remaining hrGFP II Open Reading Frame (ORF) sequence whereas Figure 11 shows the alignment of sequences from DNA sequencing with part of the hrGFP II ORF sequence. After compiling the results obtained from BioEdit, a pie chart was drawn to depict the percentage of sequences with the whole hrGFP II ORF sequence and sequences with part of the hrGFP II ORF sequence.

Figure 12 shows that there is a higher percentage for sequences with part of the hrGFP II ORF sequence (84 %), while sequences with the whole hrGFP II ORF sequence is only 16 %. Inserts were blast on NCBI Nucleotide Blast against host genome to ensure that the sequences from inserts are from mouse genome. Results showed that when an insert was blast against mouse genome, there was a 99% identity to mouse genome. All of the results showed a 99 % identity to mouse genome, with integration sites mostly at chromosome 6 and 10. Other integration sites include chromosome 4, 11 and 13. Virtual DNA walking was done to each of the sequences obtained from DNA walking. This step was done by walking 50 Kb upstream and 50 Kb downstream of the sequence obtained, making the sequence a total of 100

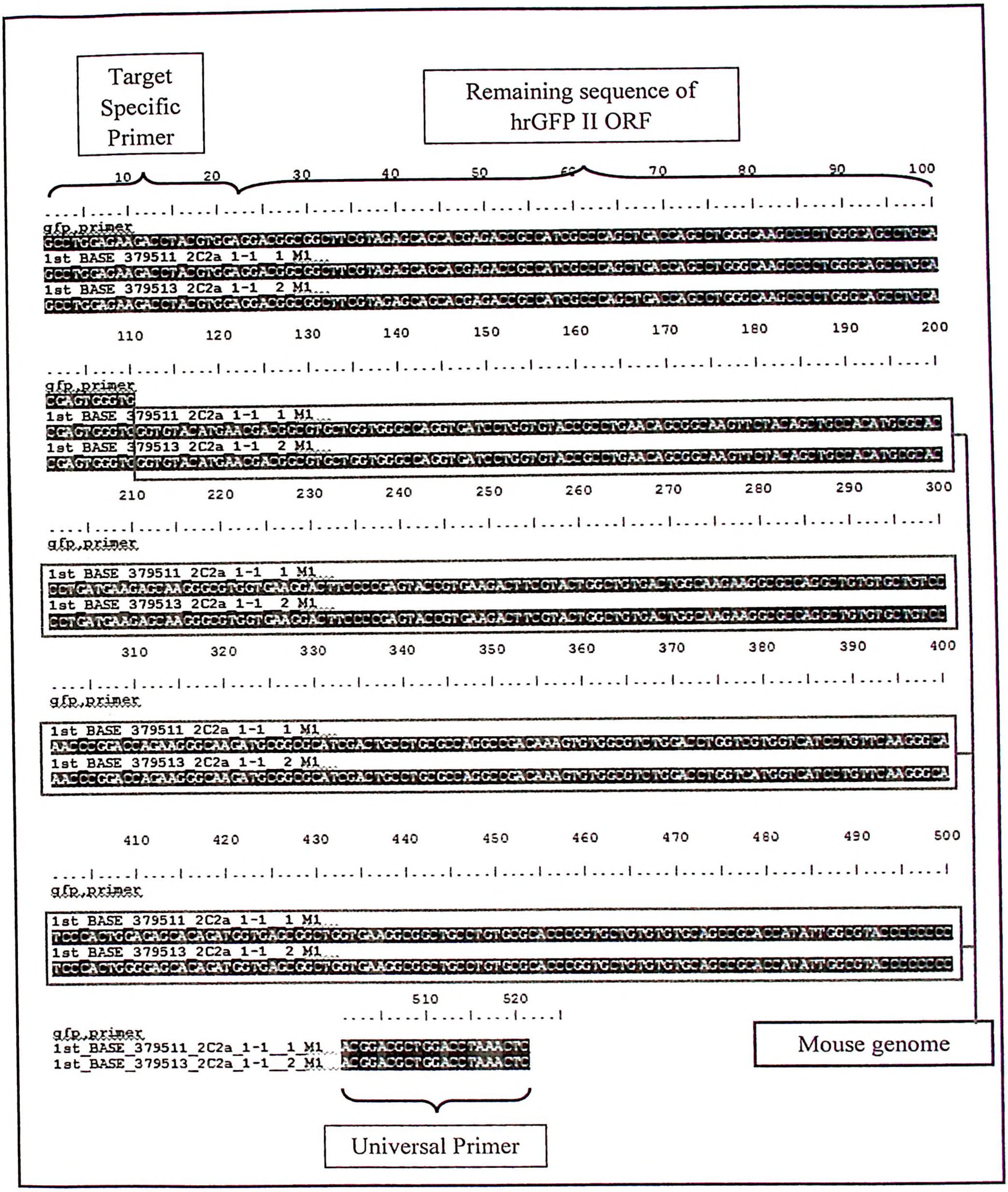


Figure 10 : The alignment of sequences from DNA sequencing with target – specific primer and the remaining hrGFP II ORF sequence (whole).

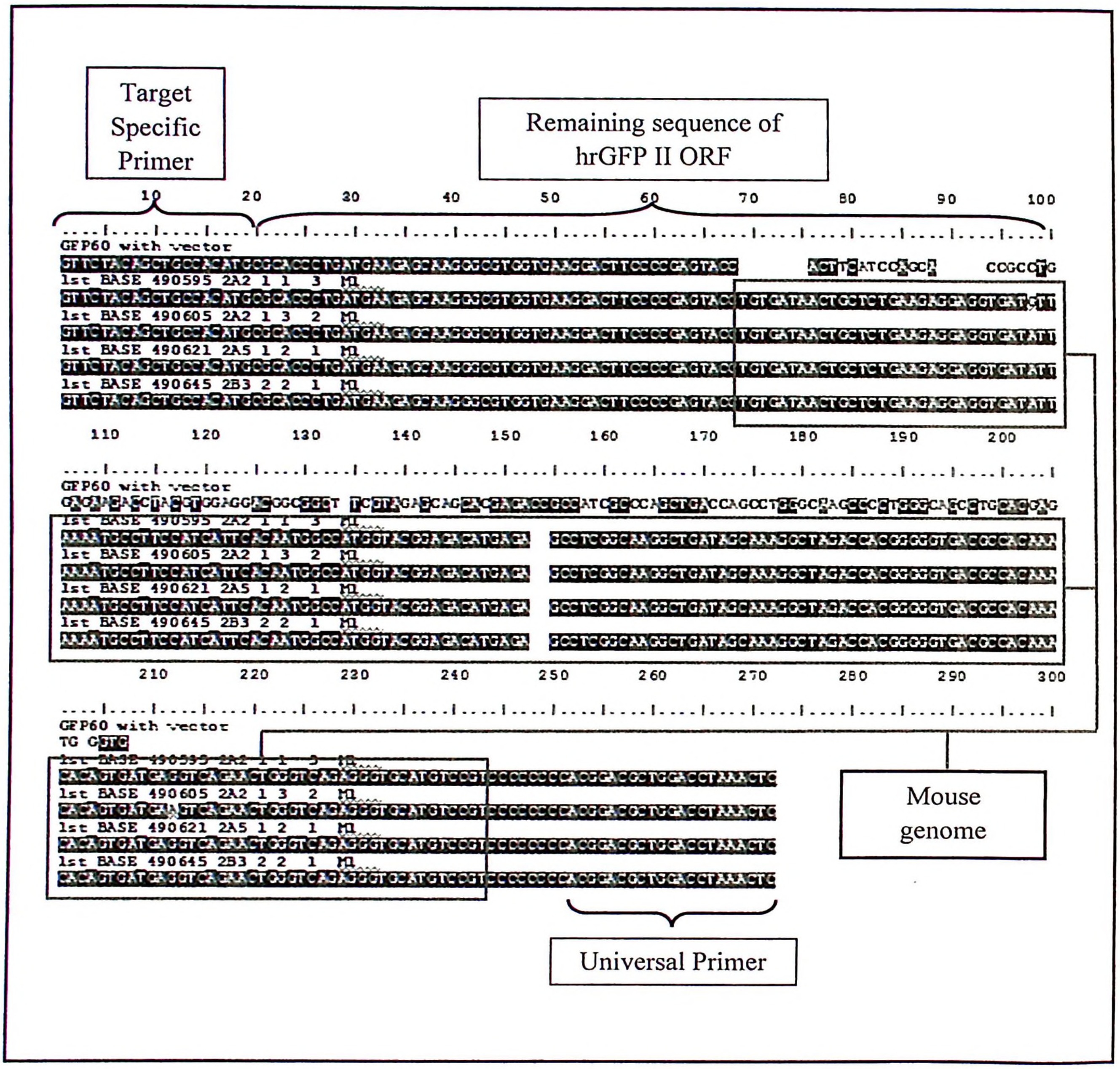


Figure 11 : The alignment of sequences from DNA sequencing with target – specific primer and part of the remaining hrGFP II ORF sequence.

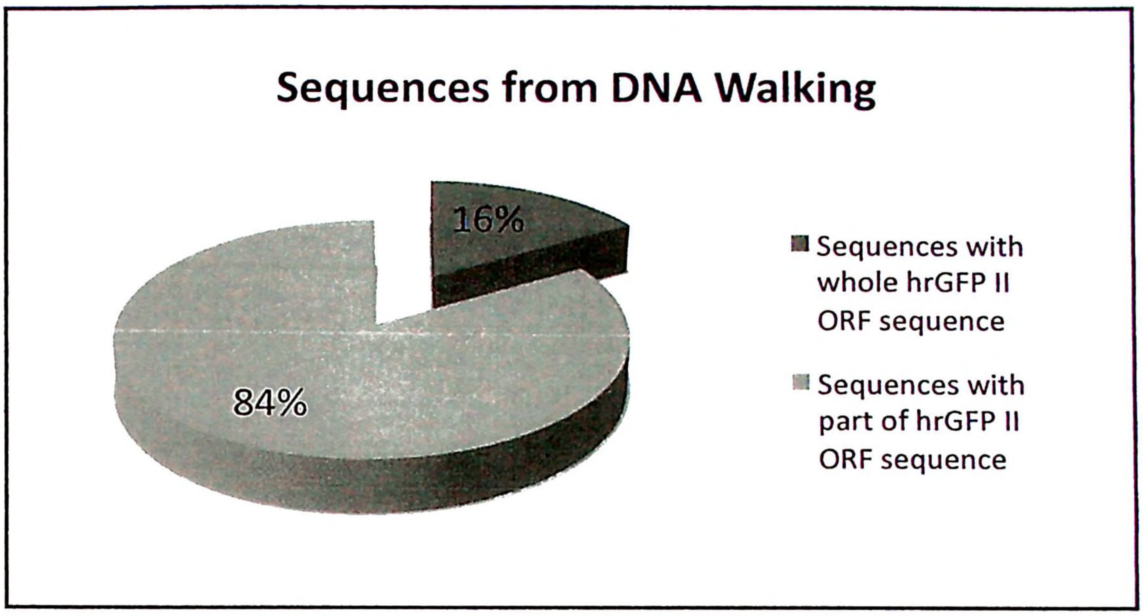
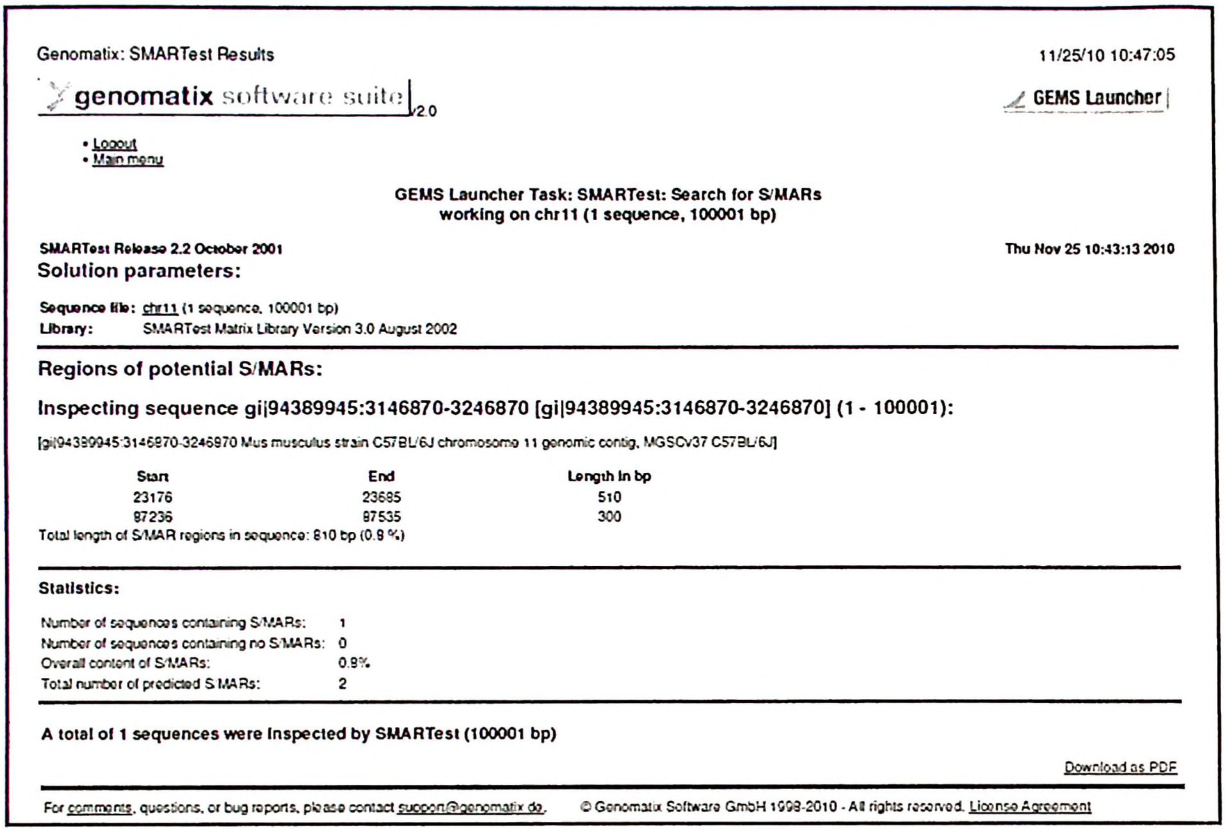


Figure 12 : The pie chart depicts the comparison of sequences derived from DNA walking with the whole hrGFP II ORF sequence and sequences with part of the hrGFP II ORF.

Bioinformatics Characterisation of Sequences from Virtual Walking

100 Kb sequences from virtual walking were uploaded to three different programs for the identification of putative Matrix Attachment Regions (MARs). These three programs are SMARTest, MARFINDER and MARSCAN, which are based on different approaches. Results from SMARTest, MARFRINDER and MARSCAN were collected and compiled into a diagram to enable better examination of the overlapping putative MARs that were identified. Overlapping putative MARs are assumed to have higher probability to turn out to be true MARs.



Genomatix: SMARTest Results 11/25/10 10:47:05

GEMS Launcher Task: SMARTest: Search for S/MARs
working on chr11 (1 sequence, 100001 bp)

SMARTest Release 2.2 October 2001 **Thu Nov 25 10:43:13 2010**

Solution parameters:

Sequence file: chr11 (1 sequence, 100001 bp)
Library: SMARTest Matrix Library Version 3.0 August 2002

Regions of potential S/MARs:

Inspecting sequence gi|94389945:3146870-3246870 [gi|94389945:3146870-3246870] (1 - 100001):

[gi|94389945:3146870-3246870 Mus musculus strain C57BL/6J chromosome 11 genomic contig, MGSCv37 C57BL/6J]

Start	End	Length in bp
23176	23685	510
87236	87535	300

Total length of S/MAR regions in sequence: 810 bp (0.8 %)

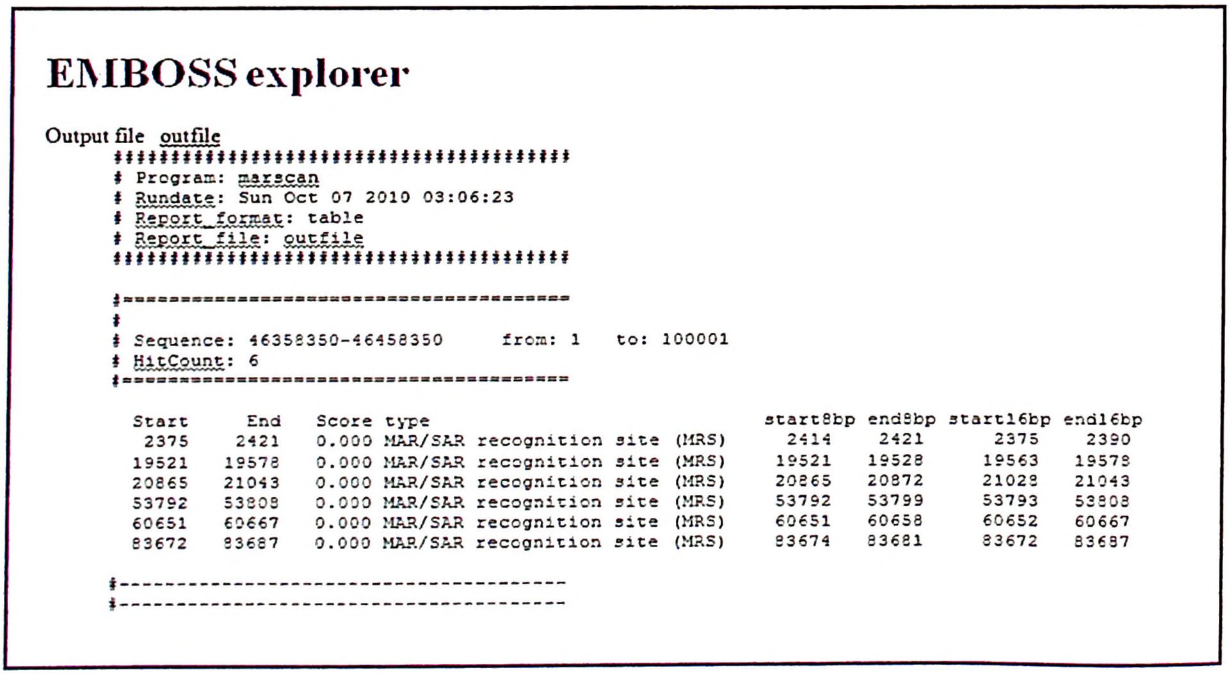
Statistics:

Number of sequences containing S/MARs:	1
Number of sequences containing no S/MARs:	0
Overall content of S/MARs:	0.8%
Total number of predicted S/MARs:	2

A total of 1 sequences were inspected by SMARTest (100001 bp)

For comments, questions, or bug reports, please contact support@genomatix.de. © Genomatix Software GmbH 1998-2010 - All rights reserved. License Agreement

Figure 13 : The results of putative MAR displayed by SMARTest software.



EMBOSS explorer

Output file outfile

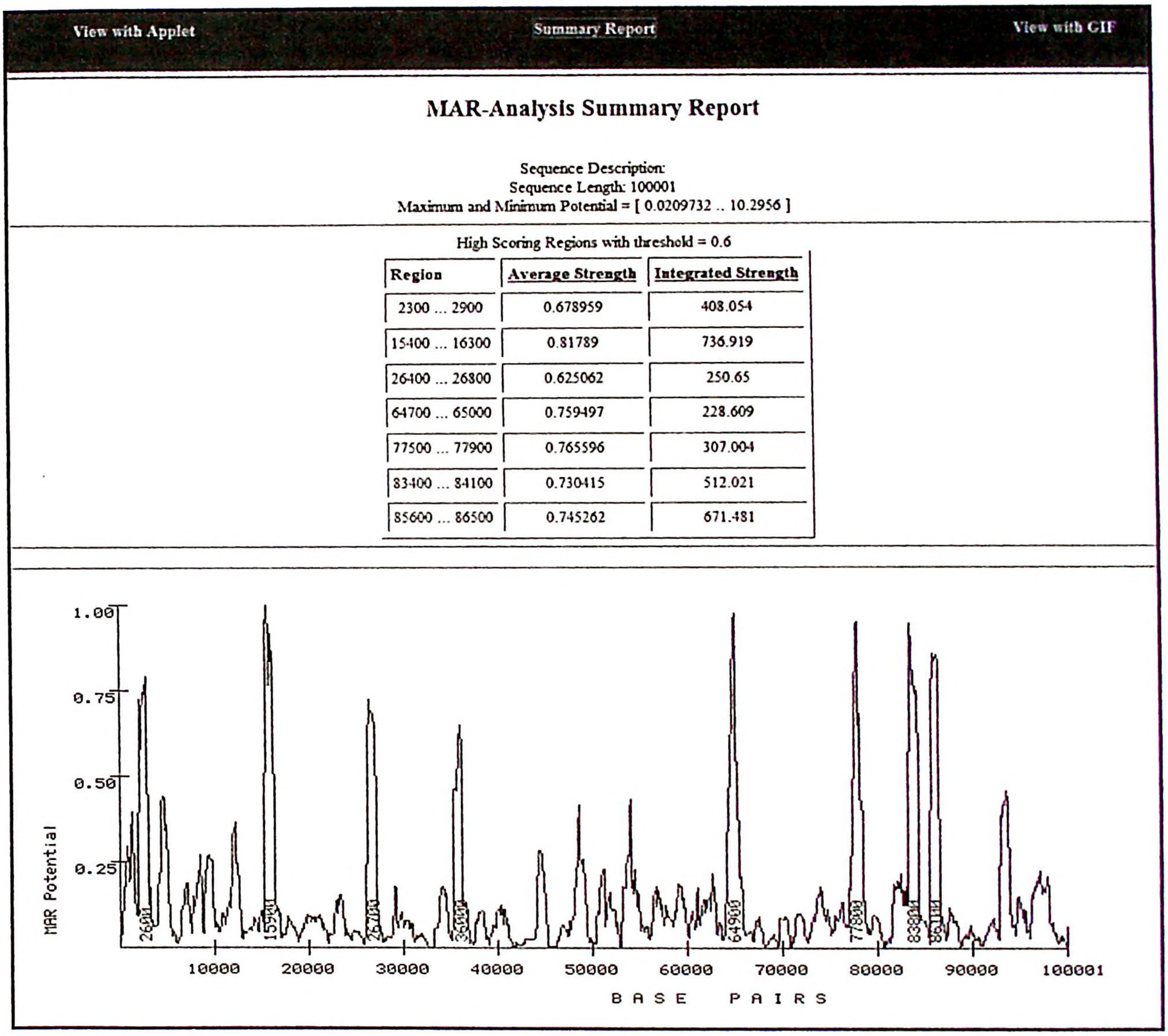
```
########################################
# Program: marscan
# Rundate: Sun Oct 07 2010 03:06:23
# Report_format: table
# Report_file: outfile
########################################

#=======================================
#
# Sequence: 46358350-46458350     from: 1   to: 100001
# HitCount: 6
#=======================================

  Start      End   Score type                                start8bp  end8bp start16bp end16bp
   2375     2421   0.000 MAR/SAR recognition site (MRS)        2414    2421      2375    2390
  19521    19578   0.000 MAR/SAR recognition site (MRS)       19521   19528     19563   19578
  20865    21043   0.000 MAR/SAR recognition site (MRS)       20865   20872     21028   21043
  53792    53808   0.000 MAR/SAR recognition site (MRS)       53792   53799     53793   53808
  60651    60667   0.000 MAR/SAR recognition site (MRS)       60651   60658     60652   60667
  83672    83687   0.000 MAR/SAR recognition site (MRS)       83674   83681     83672   83687

#---------------------------------------
#---------------------------------------
```

Figure 14 : The results of potential MAR displayed by MARSCAN.



MAR-Analysis Summary Report

Sequence Description:
Sequence Length: 100001
Maximum and Minimum Potential = [0.0209732 .. 10.2956]

High Scoring Regions with threshold = 0.6

Region	Average Strength	Integrated Strength
2300 ... 2900	0.678959	408.054
15400 ... 16300	0.81789	736.919
26400 ... 26800	0.625062	250.65
64700 ... 65000	0.759497	228.609
77500 ... 77900	0.765596	307.004
83400 ... 84100	0.730415	512.021
85600 ... 86500	0.745262	671.481

Figure 15 : The results of potential MAR displayed by MARFINDER.

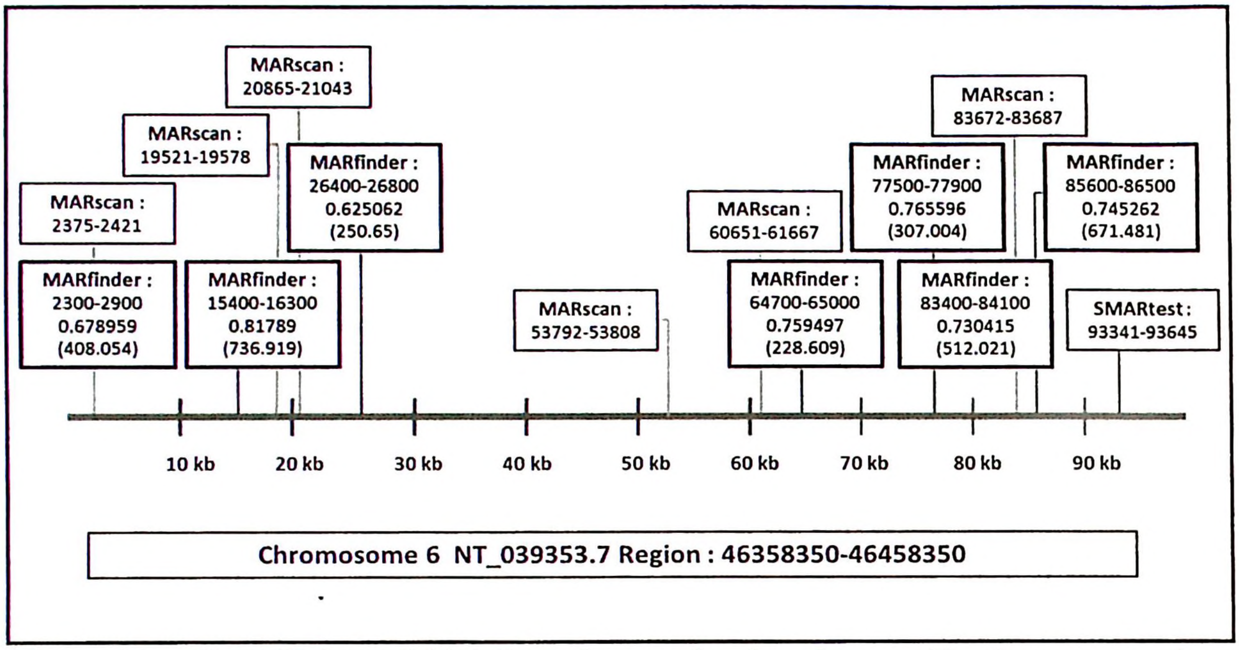


Figure 16 : Compilation of bioinformatics results for clones with chromosome 6 as integration site.

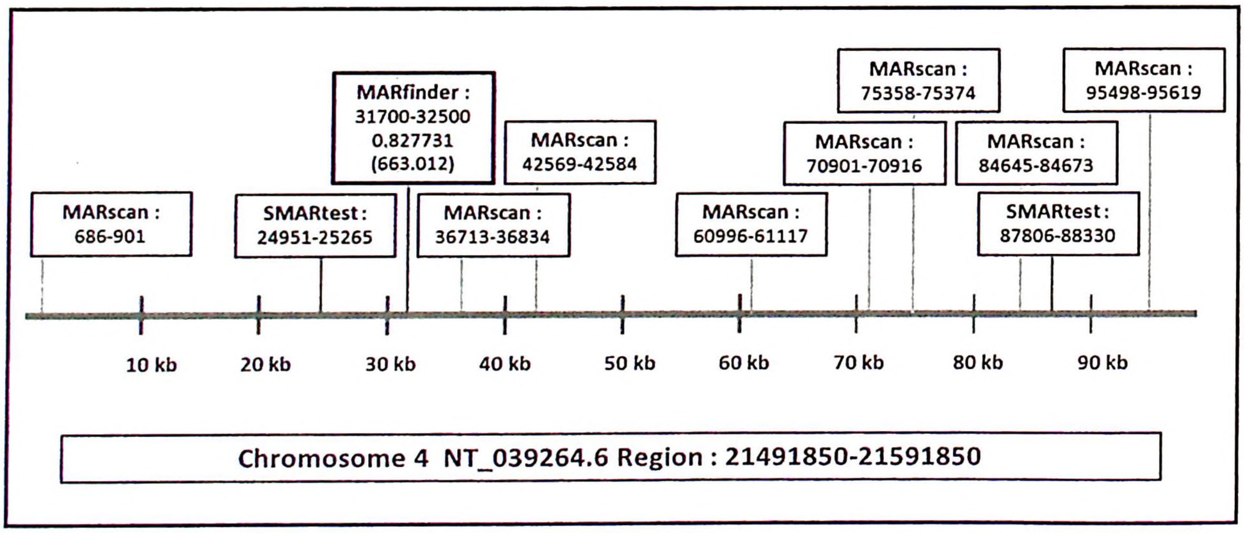


Figure 17 : Compilation of bioinformatics results for clones with chromosome 4 as integration site.

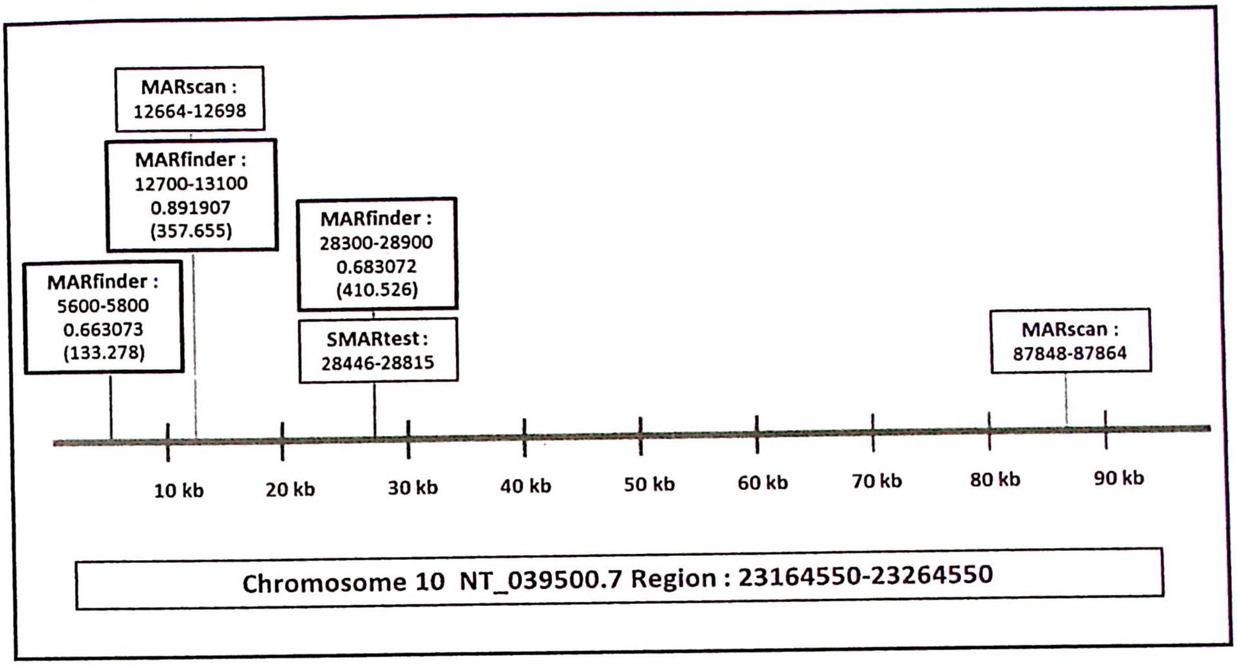


Figure 18 : Compilation of bioinformatics results for clones with chromosome 10 as integration site.

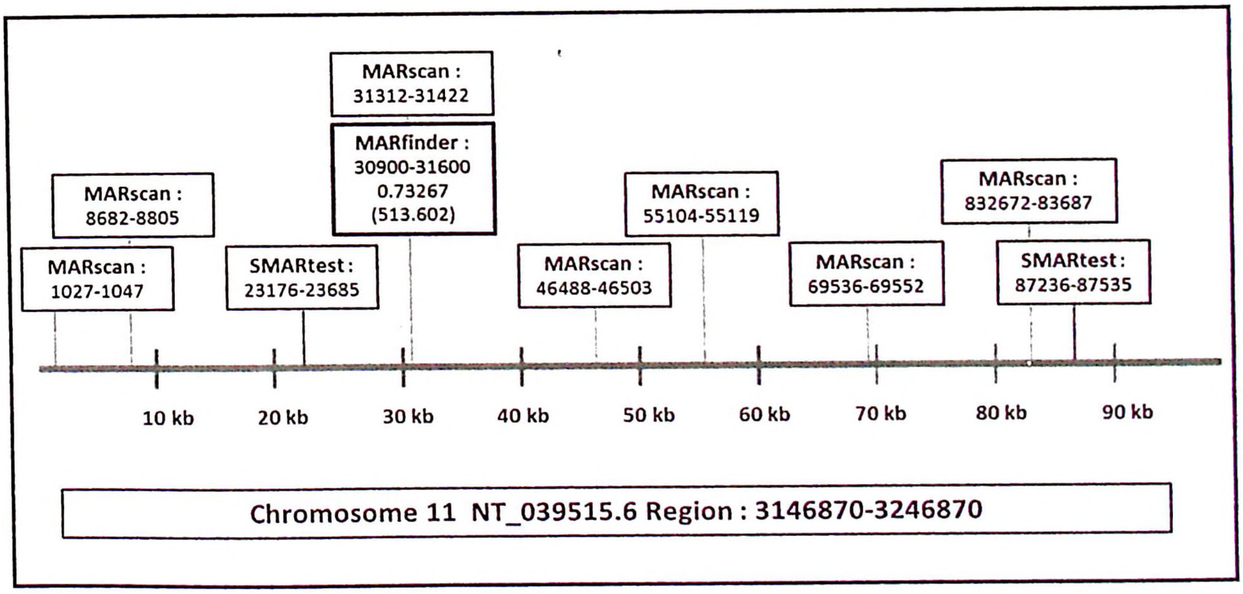


Figure 19 : Compilation of bioinformatics results for clones with chromosome 11 as integration site.

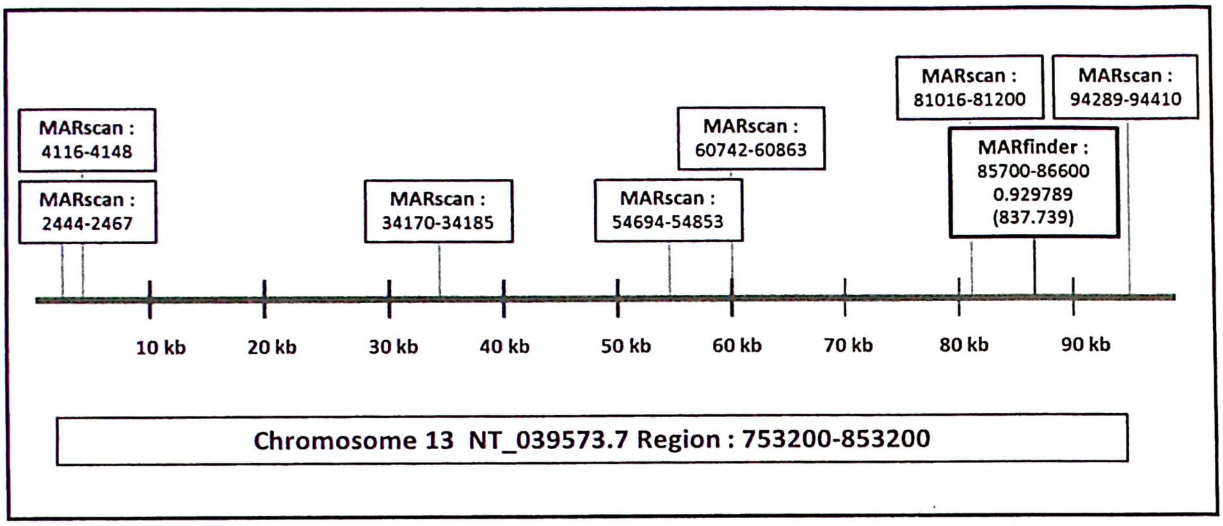
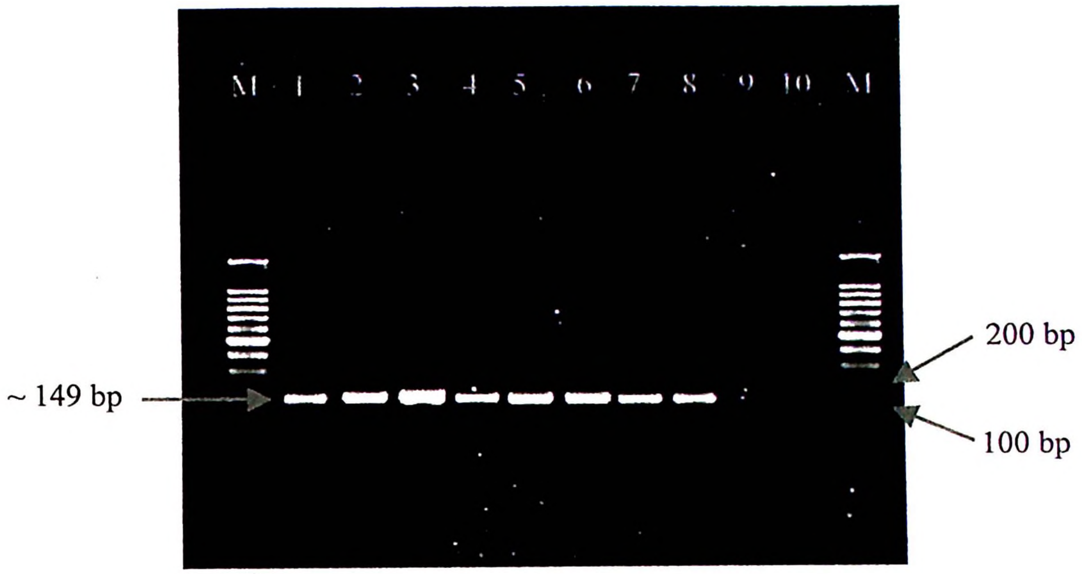


Figure 20: Compilation of bioinformatics results for clones with chromosome 13 as integration site.

Chromatin Immunoprecipitation Assay

ChIP assay was done to 5 of the chosen clones, with each clone showing integration at a different chromosome. All samples were immunoprecipitated with SATB and CTCF antibody and showed interaction between the matrix attachment region's binding proteins with the genome. Figure 21 and 22 depicts an example of the result shown for one sample.



Lane M : 100 bp DNA ladder

Lane 1 : Clone 2C18 with 50 K cells per IP

Lane 2 : Clone 2C18 with 100 K cells per IP

Lane 3 : Clone 2C18 with 200 K cells per IP

Lane 4 : Input control for Clone 2C18 with 50 K cells per IP

Lane 5 : Input control for Clone 2C18 with 100 K cells per IP

Lane 6 : Input control for Clone 2C18 with 200 K cells per IP

Lane 7 and Lane 8 : Positive control

Lane 9 and Lane 10: Negative control

Figure 21 : Analytical 1.0 % (w/v) agarose gel electrophoresis for Clone 2C18 immunoprecipitated with SATB1 antibody.