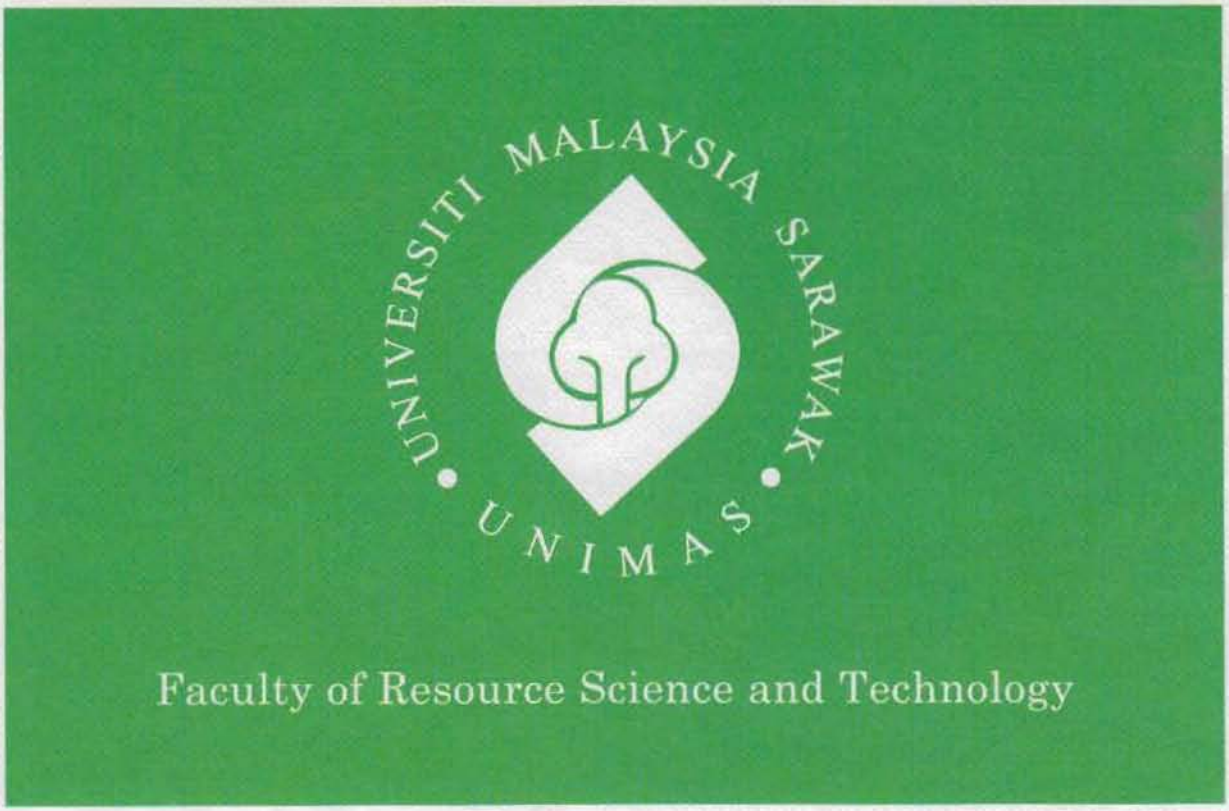

CLONING AND ACTIVITY DETERMINATION OF ENHANCER-LIKE SEQUENCE FROM PROBOSCIS MONKEY (Nasalis larvatus)

CHUNG PHEY LIN
(46585)

Bachelor of Science with Honour
(Resource Biotechnology)
2017

Cloning and Activity determination of
Enhancer-like sequence from Proboscis monkey (*Nasalis larvatus*)

Chung Phey Lin (46585)

A Thesis Submitted in Partial Fulfilment of the Requirement of
The Degree of Bachelor of Science with Honors (Resource Biotechnology)

Supervisor: Dr Chung Hung Hui

Resource Biotechnology Programme
Department of Molecular Biology

Faculty of Resource Science and Technology
University Malaysia Sarawak
2017

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (□)

Final Year Project Report ☑

Masters ☐

PhD ☐

DECLARATION OF ORIGINAL WORK

This declaration is made on the 12 June 2017.

Student's Declaration:

I CHUNG PHEY LIN, 46585, Faculty of Resource Science and Technology hereby declare that the work entitled, **CLONING AND ACTIVITY DETERMINATION OF ENHANCER-LIKE SEQUENCE FROM *Nasalis larvatus*** is my original work. I have not copied from any other students' work or from any other sources with the exception where due reference or acknowledgement is made explicitly in the text, nor has any part of the work been written for me by another person.

19 06 2017
Date submitted

Chung Phey Lin (46585)
Name of the student (Matric No.)

Supervisor's Declaration:

I, _____ (SUPERVISOR'S NAME) hereby certify that the work entitled _____ (TITLE) was prepared by the aforementioned or above mentioned student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of _____ (PLEASE INDICATE THE DEGREE TITLE), and the aforementioned work, to the best of my knowledge, is the said student's work

Received for examination by: _____ Date: _____

(Name of the supervisor)

I declare that Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*

☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*

☐ **OPEN ACCESS**

Validation of Project/Thesis

I hereby duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic and research purposes only and not for other purposes.
- The Centre for Academic Information Services has the lawful right to digitize the content to be uploaded into Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis if required for use by other parties for academic purposes or by other Higher Learning Institutes.
- No dispute or any claim shall arise from the student himself / herself neither a third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student himself/herself without first obtaining approval from UNIMAS.

Student's signature

14 Oct 2017

Supervisor's signature

Current Address:

Notes: * If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the date of restriction indicated, and the reasons for the confidentiality and restriction.

[The instrument was prepared by The Centre for Academic Information Services]

Acknowledgements

First and foremost, it is my pleasure to thank all of those who helped make this thesis probable. Utmost, I would like to express my sincere gratitude to my supervisor Dr. Chung Hung Hui for the continuous support towards my bachelor study. For his patience, motivation, immense knowledge as well as guidance to me throughout the project and writing of this thesis. Besides my supervisor, I would like to thank also the rest of the lecturers who have allowed me to use the equipment and machines in their respective lab. They are Ap Dr. Edmund Sim Ui Hang, Dr. Lee Kui Soon, Dr. Dayang and Dr. Samuel Lihan.

Furthermore, I would also like to express my deepest appreciation to my post-graduates from Animal Biotechnology Lab especially Leonard Lim Whye Kit, for his knowledge, guidance, responsible and care. Others including Hui Ying, Amil, Jill, Cyntia, Shamil, and Shek li.

Significantly, I would like to show gratitude to my lab mates. With a special mention to Vivien Ch'ng and Lee Ying, Woan ting, Alice, Pei Ling for the stimulating discussion, for sleepless nights having together before deadlines and for all the happiness during the projects. Also, I thanks Aril, Abigail, Fatin, Rosierra, Kai Jia, Kang Young, Yeh Qi, Jeffrey who tolerate and help me throughout my research. Without their friendship, this project would not be completed.

Last but not the least, I take this opportunity to acknowledge my lovely family members for their unceasing encouragement, mental and financial support as well as attention. I am blessed to have accompanied and adviced from them during my desperate moment. I am very fortunate to have such a wonderful people around. I could not imagine without their patient and sacrifice, I would definitely suffer in completing this thesis.

Cloning and activity determination of Enhancer-like sequence from Proboscis monkey (*Nasalis larvatus*)

Chung Phey Lin

Resource Biotechnology Programme
Faculty of Science and Technology University
Malaysia Sarawak

ABSTRACT

Enhancer is a distal cis-regulatory element that regulate spatiotemporal gene expression by integrating with transcription factors (TFs). Proboscis monkey (*Nasalis larvatus*) is endemic and potential to be a suitable study model due to its proximate chromosome karyotypess with human. Currently, potential *in silico* software for prediction of enhancer from vast DNA sequence is still in increasing trend of development. The aim of this study is to isolate enhancer sequence from *N. larvatus* and to determine its TFs bounded. In this research, a strong putative enhancer (683bp) with >90% confidence level was selected via iEnhancer-2L software for further PCR amplification. The PCR amplicon was purified and cloned into pGEM-T® Easy vector followed by pGL 3.0 SV40 basic vector. Subsequently, transformed into E.coli XL1 Blue with transformation efficiency of 2.1 X10^4 transformants/ug. Double digested restriction enzymes *BamHI* and *SalI* was conducted to verify the propagation of plasmid with insert DNA. Sequencing result and BLASTn analysis indicated that the putative enhancer has 79% Query similarity to the expected one. Further transcription factors binding site search using MATCH™ 1.0 public and Alibaba 2.1. This includes AP-1, C/EBP beta, NF1, HNF-1, ER receptor and SP1. This shows that the putative enhancer might play a role in regulating liver specific genes. Due to time and reagents constraint, the determination of enhancer activity through transient transfection was not able to be performed. However, future validation on the usability of this approach can be supported by functional assays.

Keyword: Enhancer, Transcription factors, *Nasalis larvatus,* gene-regulation.

ABSTRAK

Enhancer adalah unsur cis selia distal yang mengawal gen spatiotemporal dengan mengintegrasikan dengan faktor-faktor transkripsi (TFs). Proboscis monkey (*Nasalis larvatus*) adalah endemik dan berpotensi untuk menjadi model kajian yang sesuai kerana ada proksimat karyotypes kromosom dengan manusia. Kini, potensi dalam in silico untul ramalan enhancer daripada urutan DNA luas masih dalam penigkatan trend pembangunan. Tujuan kajian ini adalah untuk mengasingkan urutan enhancer dari N. larvatus dan untuk menentukan TFs yang terbatas. Dalam kajian ini, potensi enhancer (683bp) dengan >90% tahap keyakinan telah dipilih melalui perisian iEnhancer-2Lbagi lagi PCRpenguatan. PCR Amplicon telah disucikan dan diklon ke dalam 'pGEM-T® Easy vector' diikuti oleh 'PGL 3.0 SV40 Basic vector'. Selepas itu, tranform menjadi E.coli XL1 biru dengan kecekapan transformasi 2.1 X10 ^ 4 transformants / ug. Double enzim sekatan dicerna *BamHl* dan *SalI* telah dijalankan untuk mengesahkan penyebaran plasmid dengan DNA memasukkan. Keputusan sequencing dan BLASTn menunjukkan bahawa penambah putatif mempunyai 79% persamaan dengan yang diharapkan. Faktor transkripsi lagi mengikat carian tapak menggunakan MATCH™ 1,0 awam dan Alibaba 2.1. Ini termasuk AP-1, C / EBP beta, NF1, HNF-1, ER reseptor dan SP1. Ini menunjukkan bahawa penambah putatif mungkin memainkan peranan dalam mengawal hati gen tertentu. Oleh kerana masa dan reagen kekangan, penentuan aktiviti penambah melalui transfection fana tidak dapat dilakukan. Walau bagaimanapun, pengesahan masa depan pada kebolehgunaan pendekatan ini boleh disokong oleh asai berfungsi.

Keyword: Enhancer, Faktor-faktor transkripsi, *Nasalis larvatus,* gene-regulation.

TABLE OF CONTENT

LIST OF TABLE

LIST OF FIGURES

LIST OF ABBREVIATION

AGE	Agarose Gel Electrophoresis
CRM	Cis-regulatory module
ddH20	Double distilled water
DNA	Deoxyribonucleic Acid
EDTA	Ethylene diamine tetraacetic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
NCBI	National Centre for Biotechnology Information
NHP	Non-human Primate
PCR	Polymerase Chain Reaction
PLB	Passive Lysis Buffer
PSA	Prostate cancer
RNA	Ribonucleic Acid

SV40	Simian Virus 40
Taq	*Thermus Aquaticus*
TAE	Tris-acetate EDTA
TF	Transcription factor

CHAPTER 1.0 INTRODUCTION

An enhancer is a *cis*-regulatory element that up-regulates eukaryotic gene expression providing binding sites for transcription factors and their cofactors. Besides enhancer, there are also others regulatory elements that can control the gene activity such as promoter and silencer. However, in this studies, enhancer is more emphasized. It is also discovered that enhancer can activate promoter in *cis*- or *trans*- position and it is located long-distance away from the transcription start site (Erokhin *et al.*, 2015). Significantly, enhancer is a stepping stone to increase the rate of transcription of a gene activities of cell development, proliferation and division (Dickel *et al.*, 2014).

Most enhancer are bounded with more than one transcription factors (TFs) in gene regulation. The interaction between different TFs and DNA binding motif of enhancer is a key step in mediating the gene expression of each gene independently in specific tissue and cell. Mutation in enhancer nucleotide sequence may alter the activity or availability of TFs, and subsequently these changes may contribute to evolution of the gene expression (Wittkopp, 2010). In addition, the alteration of enhancer element would also significantly act as a predominant source of divergence across closely related species (Wilson *et al.*, 2008). Once the enhancer sequence is predicted, transcription factor encoded with the binding site of enhancer can be verified *in silico*. However, the differences in nucleotide sequence of putative enhancer sequence with other species could hardly link to the mutation of particular gene expression. Hence, further studies is necessary for researcher to conduct the further studies in the relationships between enhancer sequence variation, TFs as well as their gene expression.

In the first human genome project, the size of the human genome was discovered to be over 2.6 billion base pair (Brown, 2002). The completion of human genome draft has enhanced the curiosity of scientists to further explore other organism's genome. Proboscis monkey (*Nasalis larvatus*) is a non-human primate endemic to island of Borneo (Ridzwanali *et al*, 2009). Despite reported result of 90% and above genetic similarities between human and non-human primate (chimpanzee), there was also other related research demonstrated a close relationship in chromosome painting karyotype between human and *N. larvatus* (Bailet, 2014; Bigoni *et al.*, 2003). The hybridization pattern of human chromosomes onto *N. larvatus* chromosome had made this primate an interesting subject to study. Even there is a published genome database of *N. larvatus* in NCBI genebank, but the study of spatiotemporal gene regulation in this non-human primate has not fully explored.

There is still a broad study to use computational approaches in prediction of the enhancer sequence with the flood of DNA sequences generated in the genomic age. This research is focusing on the prediction of putative enhancer *N. larvatus* using iEnhancer-2L algorithm and verifying the enhancer activity in specific cell line. The hypothesis is that there is presence of putative strong enhancer sequence in *N. larvatus* chromosome 18 and it can be clone into pGL 3.0 SV40 reporter vector. Moreover, the activity of the enhancer can be measured by luciferase assay. The research outcome may also widen the scope of study on the gene regulation of enhancer, its predicted transcription factor binding sites and in future its activity in reporter assay.

The objectives of this research are:

1. To isolate predicted enhancer-like sequence from *N. larvatus.*
2. To clone the putative enhancer into luciferase reporter pGL3.0 SV40 basic vector.
3. To identify the transcription factors binding sites *in silico.*

CHAPTER 2.0 LITERATURE REVIEW

2.1 Enhancer

In mammalian species, enhancer is a key element in initiating and driving precise gene expression through coordination of activator protein transcription factors (TFs). These interactions are arranged in a compact structure with chromatin bounded. Significant evolution on the expression level of enhancer sequence could either over-expression or under-expression resulting in alteration of gene transcription, subsequently cause appearance of the human disease. The first enhancer discovered was simian virus 40 (SV40) in 1981 by Walter Schaffner in a virus genome (Figure 2.1) (Copper, 2000). While the first human enhancer was identified in mammalian B lymphocytes that modulated immunoglobulin (Ig) heavy chain gene expression (Levine, 2010). Both of these discoveries had become the keystone in biology information into developing a detailed understanding of the mechanisms of enhancer in activating gene regulation. In addition, there is still variety of enhancers remain controversial and undefined in different species.

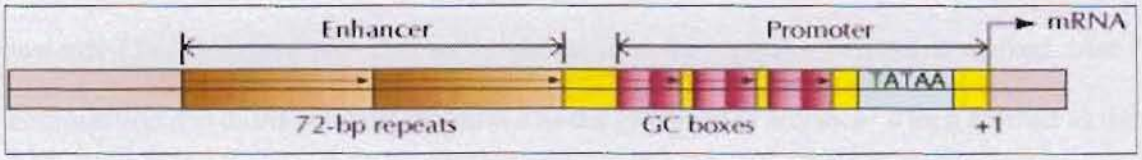

Figure 2.1: SV40 enhancer. (Adapted from http://www.ncbi.nlm.nih.gov/books/NBK9904/)

An enhancer may lie more than a thousand bases away from the target gene in either linear space or three-dimensional space. However, it is possible for it to interact with close proximity promoter. DNA looping mechanism (Figure 2.2) has been clearly studied on how the enhancer is able to drive gene expression distantly from transcription start site of a promoter. It serves as an operational platform to integrate activator protein such as transcription factors protein followed by binding of co-activator mediator complex (TFID). Lastly, recruitment of RNA Polymerase II at promoter binding site will initiate the gene transcription of particular cell type (Ong & Corces, 2011;

Spitz & Furlong, 2012). The formation of DNA looping between all the mediators are stabilized by cohesin protein.

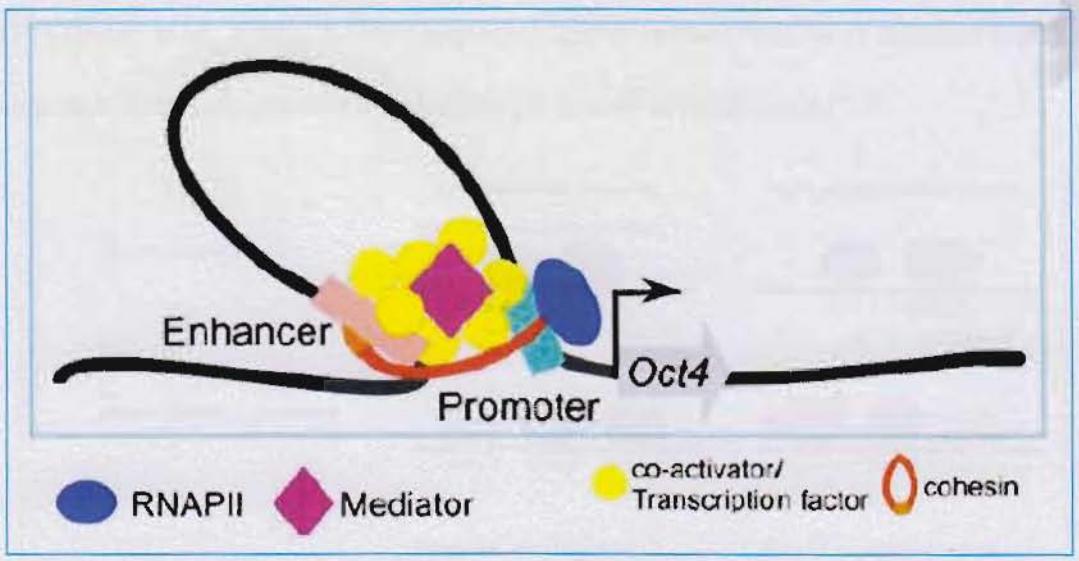

Figure 2.2: Formation of DNA looping. Oct4 represents the gene in Embryonic stem cell (Adapted from Ong & Corces, 2011).

Cooperative recruitment between enhancer and transcription factors in either directly or indirectly interact with the transcriptional machinery can be inferred from motif architecture of enhancer. The specific arrangement of the DNA-binding motif of enhancer towards TFs has indeed provided an insights to how the regulatory elements worked. Motif composition and motif position are known as the grammar of enhancer which defined as the specific arrangement of TFBs towards it (Spitz & Furlong, 2012). Motif composition is a binding motif that located in between enhancers for recruitment on specific TFs that driven expression in selected cell types (Spitz & Furlong, 2012). Analysing on only a few TF occupancy on enhancer will be sufficient to observe on the temporal-spatial of enhancer activity.

Secondly, motif positioning often referred to the motif orientation, relative position and distance of TF within an enhancer (Figure 2.3) (Spitz & Furlong, 2012). A proper orientation of binding motif will allow a prospective protein-protein interaction and thereby promote cooperative binding of cofactors with the transcriptional machinery. Therefore,

understanding in enhancer grammar is effective to generate multiple standard outputs in gene regulation, yet in similar enhancer output, it can be generated from various recruitment of TFs (Zinzen *et al.*, 2009). It is not surprising that in an identification of enhancer regulatory sequence, there are cooperated with different as well as repetition of TFs.

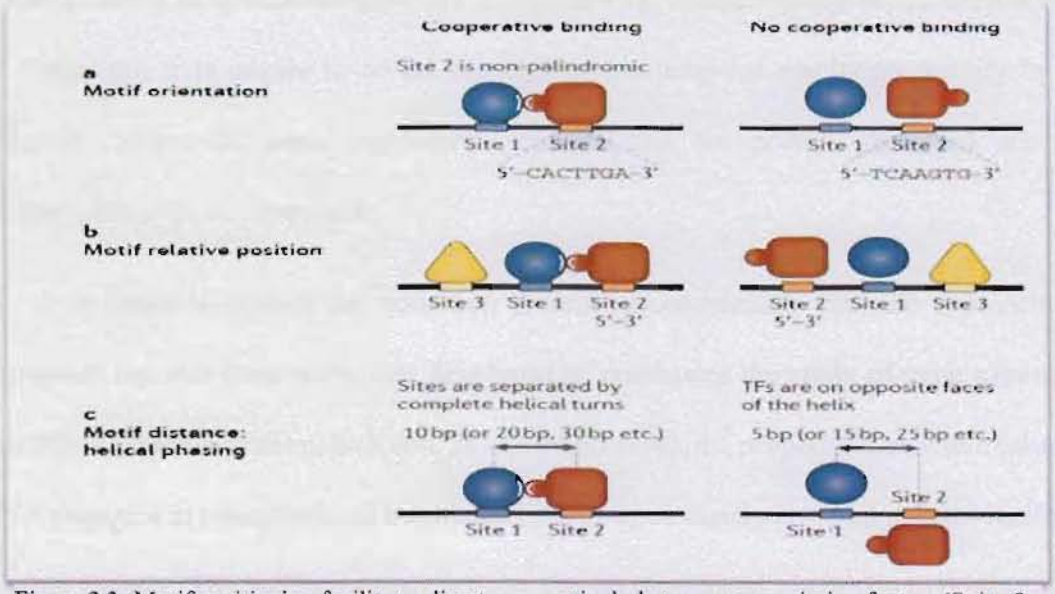

Figure 2.3: Motif positioning facilitates direct cooperatively between transcription factors (Spitz & Furlong, 2012).

The mutation of non-coding enhancer could lead to species phenotypic or genotypic evolution (Villar *et al.*,2015). According to Cortney *et al.*, 2013 and Xiao *et al.*, 2012, they compared few selected mammals that have marked mutation in enhancer regulatory elements are rapidly change during evolution especially while observing the evolutionarily stable gene expression patterns of the species. The divergences of morphology and genetic differences between particular mammals are well hypothesized as the caused by the mutation in highly conserved enhancer segments among the species genome. Yet, the activity of enhancer can be restricted in the different cell line, at a particular time and environment condition. Therefore, it is profoundly vital to identify the enhancers and understand the mechanism driving enhancer evolution recruited with the binding motif TFs.

2.2 Enhancer identification

In past century, despite the incomplete source of several organism genomes, comparative genomics approaches have shown to help in identifying the putative functional enhancer (Visel *et al.*, 2007). Comparative genomic is the basic method used for regulatory region across a variety of species comparisons. Unfortunately, comparative genomic approach has its limitation. It is unable to reveal the high spatio-temporal resolution activity by the enhancer. Moreover, some regulatory elements that are poorly conserved are also undetectable with this approach.

In order to replace the limitation present in comparative genomic approach, the transgenic reporter-gene assay was developed to emphasize the study of gene expression activity of enhancer in animals (Kvon, 2013). In this assay, the property of an active enhancer DNA sequence to transcriptional mechanism can be immediately detected with the luciferase reading activity. The luciferase gene activity can be recognized via in situ hybridization or fluorescent proteins staining (Kvon, 2013). The knowledge on enhancer regulating gene expression via particular approach has been greatly improved with the first insertion of transgenesis regarding transposable elements in fruit flies (Stanojevic *et al.*, 1991). The result demonstrated that the sequence of *eve* stripe 2 enhancer yielding abundances of TFBs has driven expression of *eve* gene within the limits of endogenous *eve* stripe 2 (Stanojevic *et al.*, 1991). The sequence of eve stripe 2 enhancer is the successful output between the association of enhancer-reporter fusion and P-element mediated transformation (Stanojevic *et al.*, 1991).

Experimental approach of verifying enhancer activity has to be correlated with several tools to search for putative enhancer sequence. A number of bioinformatic tools are now available to identify the conserved non-coding elements in the genome sequence. In

this study, an overview of computational approach such as iEnhancer-2L software was provided to interrogate and browse the enhancer sequence in available genome species.

iEnhancer-2L is known as two-layer predictor which the first layer is to analyse the identity of enhancer element in a query sequence (Liu *et al.*, 2015). Second layer predictor is to allow the categorized of enhancer strength into different subgroup including the strong enhancer, week enhancer as well as inactivate enhancer (Liu *et al.*, 2015). It has been listed as a user-friendly web server which out most the performance of other enhancer's identification software. These consists of gkm0SVM (Ghandi *et al.*, 2014) and kmer (Ioshikhes *et al.*,1996).

After that, aligning the predicted sequence with other species has been prevalently developed in relevant programs. The most common one is blastn pairwise alignment generated from https://blast.ncbi.nlm.nih.gov/Blast.cgi . The purpose of this step is to obtain the best match of the sub-region between several genomes whereas the false positive result is possible due to the non-consideration of the regions outside of the query sequence. Therefore, the most common method to achieve highly conserved region is by scoring best conservation rank in multiple alignment program with different closely selected species (http://www.ebi.ac.uk/Tools/msa/clustalo/). Last but not least, the identification on predicted enhancer is heavily relied on the TFs and hence further computer analysis tools with specific characteristic motif will ensure a more detail studies in the DNA-binding site of an enhancer.

2.3 Proboscis Monkey (*Nasalis larvatus*)

Figure 2.4: Proboscis monkey. Adapted from http://www.factzoo.com/mammals/proboscis-monkey-one-big-nose.html

N. larvatus (Figure 2.4) are old world monkey with unique long nosed characteristic and talented in swimming. It is usually range in multi-male groups of average twenty monkeys per group. Under IUCN Red List of Threatened Species, it is known as endangered species with a population of fewer than 1000 numbers in Sarawak (Meeijaard *et al.*, 2008). The decline in population trend of proboscis monkeys is mainly due to the damage of habitat caused by human activity. According to Meijaard and nijman, 2000, conservation of proboscis monkey is preferable in Kalimantan than Sabah and Sarawak area in order to ensure the long-term survival of animal as the population in Kalimantan is still in significant threat.

Non-human primate (NHP) and human share close phylogenetic relationship and physiological have indicated that primate model is a suitable animal model (Phillips *et al.*, 2014). The previous study in the genetic variation, compare to rodent model, NHP is more valuable and relevant as a model organism in research in the biomedical field. This is then further emphasized in human health and diseases including prostate cancer (PSA) (Phillips *et al.*, 2014). It was found that patients who was diagnosed with PSA was found to have deficiency of the activator factor KLK3 gene. This gene was vital to activate the downstream signaling cascade to produce PSA protein in order to inhibit prostate cancer (Karr *et al.*, 1995). Yet, in Old World monkeys and apes subgroup species, they both demonstrated with

a particular KLK3 gene in the genome that depicted that this may be a foundation source for an experiment in studying on the epidemiology of cancer.

In 2003, a research on chromosome painting of proboscis monkey was conducted to compare the karyotype features with the human chromosome. The result demonstrated that in chromosome 6 of proboscis monkey have derived reciprocal translocation from human chromosome 1 and 19 whereas chromosome 17 of proboscis monkey was reciprocal translocation from human chromosome 14 and 15 (Bigoni *et al.*, 2003). The in situ hybridization outcome was displayed in Figure 2.5. According to all the research information available, it is said that there is a proximity of relationship between proboscis monkey and human. Thus, in this thesis, we are aiming to find a correlation between the *N. larvatus* chromosomal karyotype with human chromosome.

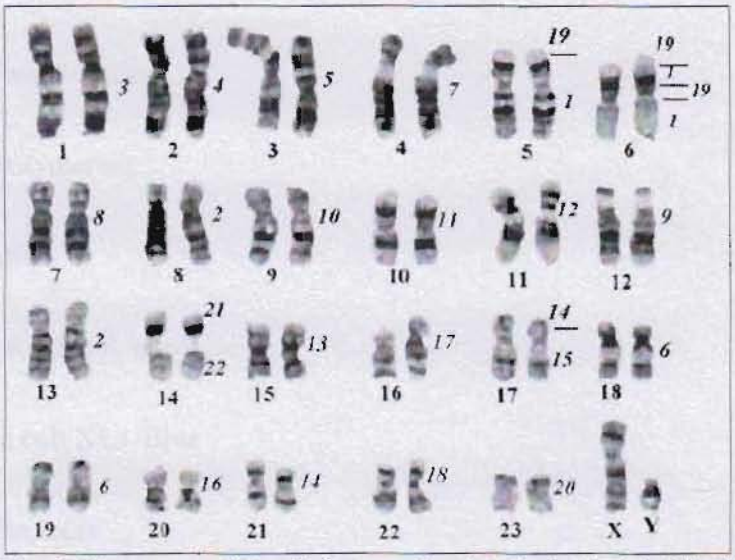

Figure 2.5: G-banded karyotype of Nasalis larvatus showing the in situ hybridization result. (Bigoni et al., 2003) .