Pertanika J. Trop. Agric. Sc. 42 (2): 627 - 647 (2019)

PERTANIKA JOURNALS

TROPICAL AGRICULTURAL SCIENCE

Journal homepage: http://www.pertanika.upm.edu.my/

Isolation and Characterization of Putative Liver-specific Enhancers in Proboscis Monkey (*Nasalis larvatus*)

Leonard Whye Kit Lim[1]*, Hung Hui Chung[1], Yee Ling Chong[1] and Nung Kion Lee[2]

[1]Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
[2]Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

ABSTRACT

Enhancers are indispensable DNA elements responsible for elevation of gene transcriptional efficiency that regulates biological processes tightly at various developmental stages, linking them to numerous genetic diseases. Discovering the enhancer landscape of the genome will not only benefit mankind, but also aid in conservation researches involving endangered non-human primates such as the proboscis monkey. As one of the most ancient colobine endemic to Borneo Island, the proboscis monkey offers a wide spectrum of unique and exclusive characteristics that distinguish it from other primates. This study has successfully isolated 13 liver-specific enhancers from this primate and tested for their activities in HepG2 and A549 cell line. The TFBS-enriched regions such as pairs of AP-1, clusters of C/EBP-β and triplets of HNF-1 in enhancers contributed to enhancer activities whereas huge clusters of HNF-3β possess suppressing effects, but generally these regions contributed to the cell specificities of enhancers. It is hoped that this study serves as a stepping stone in knowledge enrichment on this primate and future conservation researches.

Keywords: Computational approach, conservation, enhancer, liver, proboscis monkey

ARTICLE INFO

Article history:
Received: 14 November 2018
Accepted: 30 January 2019
Published: 30 May 2019

E-mail addresses:
lwkl1993@gmail.com (Leonard Whye Kit Lim)
hhchung@unimas.my (Hung Hui Chung)
ylchong@unimas.my (Yee Ling Chong)
nklee@unimas.my (Nung Kion Lee)
* Corresponding author

INTRODUCTION

Enhancers are DNA elements capable of elevating transcriptional efficiency of the genes they regulate regardless of their orientation and locality in the genome (Khoury & Gruss, 1983; Lim et al., 2018a). There are many roles that require the involvements of the enhancers for the determination of phenotypes especially

ISSN: 1511-3701
e-ISSN: 2231-8542

in various biological processes and developmental stages (Kleftogiannis et al., 2015; Pennacchio et al., 2015). Enhancers (size varying from 50 bp to 1.5 kbp), when activated, can work in tandem with multiple counterparts of their own to activate gene transcriptions up to 1 Mbp away from their positions (Blackwood & Kadonaga, 1998; Pennacchio et al., 2015).

The first enhancer, SV40 enhancer, was discovered almost 37 years ago and it was found that deletion of this element led to at least 100-fold decrement of early gene expression of T antigen (Banerji et al., 1981; Benoist & Chambon, 1981; Gruss et al., 1981). Since then, the discovery of enhancers had progressed throughout the decades utilizing both experimental (reporter assays and high-throughput assays) as well as computational approaches (Cao & Yip, 2016). Experimental approaches such as enhancer trap and transient transgenesis involve reporter plasmids being expressed in the host cells for detection (Kvon, 2015). These approaches (although some that involve high-throughput technologies) can only be conducted on limited number of cellular conditions and cellular properties at one go (Shlyueva et al., 2014). Contrary to the experimental approaches, the computational approaches offer rapid, low-cost, no context restrictions and less labour-intensive alternatives to conduct genome-wide enhancer identification via various machine learning methods (Cao & Yip, 2016). A myriad of computational enhancer prediction tools which employed one or integration of the many enhancer features such as histone modifications, sequence feature, motif signal feature as well as open chromatin feature, had successfully achieved high predictive accuracies (>90%) (He & Jia, 2016, 2017; Kim et al., 2016; Lim et al., 2018a; Liu et al., 2015; Liu et al., 2016). Interestingly, Omar et al. (2017) had attempted to consolidate results from five enhancer predictor tools (namely LS-GKM, GMFR-CNN, CSI-ANN, DeepBind and iEnhancer-2L) employing different enhancer features (Alipanahi et al., 2015; Firpi et al., 2010; Ghandi et al., 2014; Liu et al., 2015; Wong et al., 2016) and had achieved improved accuracies.

The link between enhancer and diseases has been reported in many human-related diseases such as X-linked deafness type 3 (CFN3), coronary heart disease, prostate cancer, Hirschsprung disease and preaxial polydactyly (de Kok et al., 1996; Emison et al., 2005; Grice et al., 2005; Lettice et al., 2002; McPherson et al., 2007; Yegnasubramanian et al., 2011). Kim et al. (2011) took one step further by comparing how enhancer variations between different ethnic groups were correlated to drug responses, focusing on liver-specific enhancers found surrounding nine major liver membrane transporter genes. These enhancer researches are not only beneficial towards the wholesomeness of humankind, but it also channels towards a better understanding of the enhancer landscape of other non-human primates for conservation purposes. One of the non-human primates that is interesting to look upon is none other than one of the most ancient primate colobine of all, the proboscis monkey (*Nasalis larvatus*).

The proboscis monkey, a species endemic to the coastal regions of Borneo, is a reddish-brown skin-coated colobine with a very distinctive large nose (in males) and huge pot-shaped belly (Groves, 2001). There are various characteristics of this primate that are clearly distinguishable from its primate counterparts that spikes interest into the effort of bringing it into the limelight of primate enhancer research (Lim et al., 2019). For instance, its morphological characteristics and behaviours such as terrestrial movement, highly-distinctive sexual dimorphism, proportions of the extremities and skull structure suggest its ancient status that is associated to that of *Rhinopithecus* (Peng et al., 1993). Besides, the digestive physiology (in its four-chambered stomach) such as the regurgitation/remastication (R/R) of this foregut fermenter allows it to possess lysozymes that are believed to have adaptive convergence with that of the ruminants (Bigoni et al., 2003; Matsuda et al., 2011; Stewart et al., 1987). Despite that this primitive primate is being listed as endangered by the International Union for Conservation of Nature (IUCN) (Meijaard et al., 2008), the conservation research on this primate is still at its infancy and requires immediate attention. Furthermore, with the completion of the genome sequencing of the proboscis monkey (Abdullah et al., 2014; Tamrin, 2016), this study aimed to isolate genome-wide liver-specific enhancers from the proboscis monkey and further characterize them. As most enhancer studies to date involving non-human primates like the chimpanzee are focused on evolutionary studies (Boyd et al., 2015; Prescott et al., 2015), it is hoped that this study acts as a stepping stone towards the establishment of conservation research on the proboscis monkey (an ancient "long-neglected" endemic primate), similar to that of the toxicology gene expression studies on endemic *Rasbora* fish (Lim et al., 2018b), antibody production against elephant endotheliotropic herpesvirus (EEHV) (Kochagul et al., 2018) and pathogen battling in white-nose syndrome in bats (Palmer et al., 2018).

MATERIALS AND METHODS

Computational Enhancer Prediction Data Collection

The computational enhancer prediction data collection was conducted with three different approaches (Supplementary Figure 1). First, the chromosome 18 sequences of the proboscis monkey were retrieved from GenBank database (Accession number: CM003007.1). The strong enhancer regions (>300 bp, flanked by weak and non-enhancer regions) identified by iEnhancer-2L using default parameters (Liu et al., 2015) were selected. Second, the genome-wide human liver membrane transporter gene enhancer data was obtained from Kim et al. (2011). The enhancers are aligned with genome of rhesus macaque and chimpanzee using ECR browser (Ovcharenko et al., 2004) and only enhancers achieving above 70% conservation were selected. Third, the enhancers (>300 bp) identified by Omar et al. (2017) were obtained for the next step. All enhancers identified from the three different

approaches were pooled and ranked based on the liver-specific TFBS per nucleotide frequency enumerated by MATCH (Kel et al., 2003) using liver-specific profile and other default parameters. Thirteen top ranked proboscis monkey enhancers (pme) were selected for further procedures. A total of four, five and four enhancers were selected from the first, second and third approach respectively, and they were termed pme001 to pme004, pme101 to pme105 and pme201 to pme204 accordingly.

Isolation and Cloning of Enhancers

A total of 13 pairs of primers (each added with *Bam*H I and *Sal* I restriction sites at the 5' end of forward and reverse primers, respectively) were designed using Primer3 (Untergasser et al., 2012) to isolate all 13 candidate enhancers from the proboscis monkey genome (Table 1). The conditions of the primers were examined using OligoCalc (Kibbe, 2007) to ensure they are optimal in terms of GC content (35-60%), melting temperatures as well as absence of

Table 1
Gene-specific primers designed to isolate enhancers and their characteristics

Enhancer	Sequences (5'-3')	Length (bp)	GC Content (%)
pme001	F: **AA**GGATCCTGGGCCTCGTAGTTAAA	25	48.0
	R: **AT**GTCGACACACACACTTCTACGGT	25	48.0
pme002	F: **AT**GGATCCGACTGTGCTTTTCCCCTG	26	53.8
	R: **AA**GTCGACTTTTGTTGTTGATGCTGTTG	28	39.3
pme003	F: **AA**GGATCCTTTGACTCCATGTCTCAC	26	46.2
	R: **AA**GTCGACCAAGACAATGGGGAAAAT	26	42.3
pme004	F: **AT**GGATCCAGCACTGGGACTGATA	24	50.0
	R: **AT**GTCGACAGTGGGGACTTTTGTTGTT	27	44.4
pme101	F: **AT**GGATCCGGCAGGAGAATTGCTTGAA	27	48.1
	R: **AT**GTCGACAAAATTAGCTGGGCCTGGT	27	48.1
pme102	F: **AT**GGATCCTACCCAAATAGTGCTTGCTG	28	46.4
	R: **AA**GTCGACGCCACATTTCAAGTGCTCA	27	48.1
pme103	F: **AA**GGATCCTTTCCAATCTGACCAGGTG	27	48.1
	R: **AT**GTCGACTACCCTGAAACTTTGCTGA	27	44.4
pme104	F: **AT**GGATCCGTGGCTCTCAGTTTCCTG	26	53.8
	R: **AA**GTCGACAGGCATGAGCCACTACAT	26	50.0
pme105	F: **AT**GGATCCCTTCACAACCAACGTTCAT	27	44.4
	R: **AT**GTCGACTATGGAAGGAGCCTTTGG	26	50.0
pme201	F: **AA**GGATCCATGTGTCATGCGTGTGTA	26	46.2
	R: **AA**GTCGACTGACACGTCACATACGAAAA	28	42.9
pme202	F: **AA**GGATCCGGGTAGACAGCAAGGACA	26	53.8
	R: **AA**GTCGACGGCAAACTGCTTCAGGAT	26	50.0
pme203	F: **AA**GGATCCACATTTGGCAGACATAGT	26	42.3
	R: **AA**GTCGACGAGGTATGTGTCCAAAGCAA	28	46.4
pme204	F: **AA**GGATCCCAATCACCTCTCACCA	24	50.0
	R: **AA**GTCGACTCCCAAATCTTTCCTTGG	26	46.2

Restriction enzyme sites (underlined) where BamH I site was added to 5' end of each forward primer whereas Sal I site was added to 5' end of each reverse primer. Letters in bold represent additional base pair for optimal restriction digestion

secondary structures, self-complementary and hairpin. The chosen primer pairs were sent to Apical Scientific Sdn. Bhd. for synthesis. The stool of proboscis monkey was collected from Bako National Park, Sarawak, Malaysia (Permit number: NCCD.907.4.4 (JLD.11), park permit number: 537/2014 and 538/2014) and was frozen at -80°C before DNA extraction.

Stool DNA extraction was conducted using QIAamp DNA Stool Mini Kit (Qiagen, Germany) to isolate the genomic DNA of the proboscis monkey according to the manufacturer's protocol. Species verification was done via the polymerase chain reaction (PCR) isolation of cytochrome *b* gene using universal primers by Irwin et al. (1991) [forward primer L14724 (5′-GACTTGAAAAACCACCGTTG-3′) and reverse primer H15915R (5′-GGAATTCATCTCTCCGGTTTACAAGAC-3′)], sequenced and blasted to a 99% similarity to the entry KM889667.1 in the GenBank database. Gradient PCR was conducted using LA *Taq* (Takara Bio, USA) to achieve temperature optimization of each primer pair for enhancer isolation. The PCR products (enhancers) were purified using QIAquick PCR Purification Kit (Qiagen, Germany) before subjected to cloning. The enhancers were then inserted into pGL4.23 [*luc2*/minP] vector (Promega, USA) via restriction digestions of both *Bam*H I and *Sal* I, and this was followed by the cloning process. Colony selection was done via colony PCR and plasmid re-digestion using the same restriction enzymes before these plasmids each containing the enhancer insert were isolated using QIA Spin Miniprep Kit (Qiagen, Germany).

Luciferase Activity Detection

The cell lines used in this study were HepG2 (human liver carcinoma) and A549 (human lung carcinoma) (Riken, Japan). Both the cell lines were grown in MEM medium (Gibco, USA) supplemented with 10% fetal bovine serum (FBS) and 1% Pen-Strep (100 U/mL Penicilium and 100 µg/mL Streptomycin) under the 5% CO_2, pH 7.0 to 7.6 and 37°C conditions. Cell passaging was conducted when the cell reached 80-90% confluency before cell counting was conducted to allocate around 7500 to 10000 cells per well in the each of the 24-well plate. Transient transfection was performed using Lipofectamine 2000 (Thermo Fisher Scientific, USA). For each experimental plasmid, a total of nine replicates (three technical replicates for each of the three biological replicates), positive (pGL3-control plasmid) (Promega, USA) and negative control (empty pGL4.23 [*luc2*/minP] plasmid) (Promega, USA) sets were co-transfected alongside with pRL-TK plasmid (Promega, USA) according to manufacturer's protocol. The luciferase activity was detected using Dual-Luciferase® Reporter Assay (Promega, USA) and measured using Infinite M200 Pro luminometer (Tecan, Switzerland). The ratios for firefly luciferase: *Renilla* luciferase were determined and normalized to that of the non-insert control. Statistical analysis was carried out using one-way ANOVA ($p<0.05$) and post-hoc Tukey's test.

Deletion Analysis

The 13 enhancer candidates were subjected to transcription factor binding site (TFBS) analysis using MATCH liver-specific profile (Kel et al., 2003) to determine the relationship between TFBS and the resultant enhancer activities. A total of three aspects were investigated across all enhancers, namely overall TFBS abundance, TFBS composition and TFBS distribution pattern. A total of three enhancers were selected for their unique TFBS distribution pattern and luciferase activity levels: enhancer pme001, pme101 and pme103 (enhancer pme001 being the highest in activity, pme103 being the second highest and pme101 being one of the lowest in activity). Several internal primers (Table 2) were designed to exclude some key TFBS-enriched regions from the wildtype enhancer fragments to test for the effects of the deletions implied. The PCR reactions were conducted using the original forward or reverse primer pairing with the newly designed internal forward or reverse primer with flanking RE site (for new primers with suffix of "a", "b" and "c"). For new primers with suffix "D", the newly designed forward primer was

Table 2
The list of internal primers used for modified enhancer fragments amplification via PCR

Modified Enhancer Fragment	Sequences (5'-3')	Length (bp)	GC Content (%)
pme001a	F: **AA**GGATCCTGGGCCTCGTAGTTAAA	25	48
	R: **AA**GTCGACAGCGCAGTGCTTCTTCG	25	56
pme001b	F: **AA**GGATCCTGGGCCTCGTAGTTAAA	25	48
	R: **AA**GTCGACCAGCACAATGTCGCGAA	25	52
pme001c	F: **AA**GGATCCTGGGCCTCGTAGTTAAA	25	48
	R: **AA**GTCGACCATCCTGGCTGATTTTT	25	44
pme001D	F: CGCGACATTGTGCTGTGCTGTATCTATGCT	30	50
	R: AGCATAGATACAGCACAGCACAATGTCGCG	30	50
pme101D1	F: AGCGAGACTTTGACTATCTAGAGCGTTGTG	30	47
	R: CACAACGCTCTAGATAGTCAAAGTCTCGCT	30	47
pme101D2	F: ATGCTCTAACTGTAGGAACTGGCACACTGC	30	50
	R: GCAGTGTGCCAGTTCCTACAGTTAGAGCAT	30	50
pme101D3	F1: AGCGAGACTTTGACTATCTAGAGCGTTGTG	30	47
	R1: CACAACGCTCTAGATAGTCAAAGTCTCGCT	30	47
	F2: ATGCTCTAACTGTAGGAACTGGCACACTGC	30	50
	R2: GCAGTGTGCCAGTTCCTACAGTTAGAGCAT	30	50
pme103a	F: **AA**GGATCCTCACACCTACCCTTTCG	25	52
	R: **AT**GTCGACTACCCTGAAACTTTGCTGA	27	44
pme103D1	F: GGCTTCCCTTCCTACGTTACCAGCTATGCT	30	53
	R: AGCATAGCTGGTAACGTAGGAAGGGAAGCC	30	53
pme103D2	F: AATCGTAAATCCTAAAAAGTGTCTTTTAGT	30	27
	R: ACTAAAAGACACTTTTTAGGATTTACGATT	30	27

Restriction enzyme sites (underlined) where BamH I site was added to 5' end of each forward primer whereas Sal I site was added to 5' end of each reverse primer. Letters in bold represent additional base pair for optimal restriction digestion

matched with the original reverse primer and in a separate tube, while the newly designed reverse primer was matched with the original forward primer in another tube, and both tubes went through the first round of PCR reaction. Both the PCR products from the first round of PCR were pooled into one tube and diluted with dilution factor of 100 using double distilled water. This PCR product mixture was then used as template for the second round of PCR with the use of the original forward and reverse primer designed for the enhancer. Next, these modified enhancer fragments were inserted into pGL4.23 [*luc2*/minP] plasmid before subjected to cloning. Cloning and luciferase reporter assays were carried out according to the abovementioned protocols.

RESULTS AND DISCUSSION

Computational Enhancer Prediction Data Collection

The computational enhancer prediction data collection was carried out via three major approaches: functional conservation, sequence conservation and combinatorial approach. The functional conservation approach was performed on chromosome 18 of proboscis monkey using iEnhancer-2L (Liu et al., 2015) because this software not only has the capability to distinguish strong enhancers from non- and weak enhancers, also its benchmark datasets are constructed based on chromatin states of nine different cell lines (NHEK, HSMM, NHLF, HepG2, K562, HMEC, H1ES, GM12878 and HUVEC) (Liu et al., 2015). In this study, a total of 9,960,241 strong enhancers, 11,855,707 weak enhancers and 50,983,878 non-enhancers were identified from the chromosome 18 of the proboscis monkey genome.

The sequence conservation approach utilized the human liver membrane transporter gene enhancers identified by Kim et al. (2011) to search for the proboscis monkey orthologue based on sequence conservation across other non-human primate genomes available such as chimpanzee and rhesus macaque. Applying the methodology by Kim et al., (2011), a total of 105 sequence conserved enhancers surrounding the nine major liver membrane transporter genes (SLC47A1, SLC22A7, SLCO1B3, SLC22A1, SLCO2B1, ABCB11, SLCO1B1, ABCC2 and SLC10A1) were identified.

The third approach, which is the combinatorial approach of five different enhancer predictor tools (namely LS-GKM, GMFR-CNN, CSI-ANN, DeepBind and iEnhancer-2L) (Alipanahi et al., 2015; Firpi et al., 2010; Ghandi et al., 2014; Liu et al., 2015; Wong et al., 2016) utilized by Omar et al. (2017) had yielded a sum of 3861 enhancers. The enhancers obtained from all three approaches were subjected to MATCH liver-specific profile screening to enumerate TFBS per nucleotide frequency based on the six major liver-specific TFBS: AP-1, HNF-3β, C/EBP-β, GATA-3, NF-1 and HNF-1. The top 13 ranked proboscis monkey enhancers (pme) were chosen for enhancer assays: four from the first approach (pme001-pme004), five from the second approach (pme101-pme105), and four from the third approach (pme201-pme204).

Association of TFBS with Enhancer Activity

The activities of all 13 candidate enhancers were measured via the luciferase reporter assay and were displayed in normalized relative luciferase unit (RLU) bar graph as shown in Figure 1. The enhancer pme001 scored the highest activity level among all in HepG2 cell line, rocketing even above that of the positive control set. While all the others have moderate enhancer activities, the bottom three with the lowest luciferase score level are the enhancer pme003, pme101 and pme104 (all three of them are not significantly different from one another) in HepG2 cell line. No observable trends can be detected in the enhancer activities across three different approaches (no significant differences across groups).

The association of TFBS with enhancer activity was investigated based on three aspects: overall TFBS frequency per nucleotide, TFBS composition and TFBS distribution pattern. Based on Figure 2, the enhancer pme105 has the highest liver-specific TFBS per nucleotide frequency. The enhancer pme101 and pme203 are second and third in place whereas the enhancer pme001 is ranked among the bottom five in terms of TFBS frequency per nucleotide. Clearly, the correlation between the overall TFBS frequency per nucleotide and enhancer activity is weak, and in some cases, a negative correlation in enhancer pme001 and pme101 particularly. It is believed that other factors such as the TFBS composition and distribution pattern dictating the enhancer activities where compact TFBS within the enhancers is related to active enhancers (Bery et al., 2014). Therefore, TFBS abundance itself is inadequate to influence enhancer activity (Gotea et al., 2010; Hu et al., 2007; Lusk & Eisen, 2010).

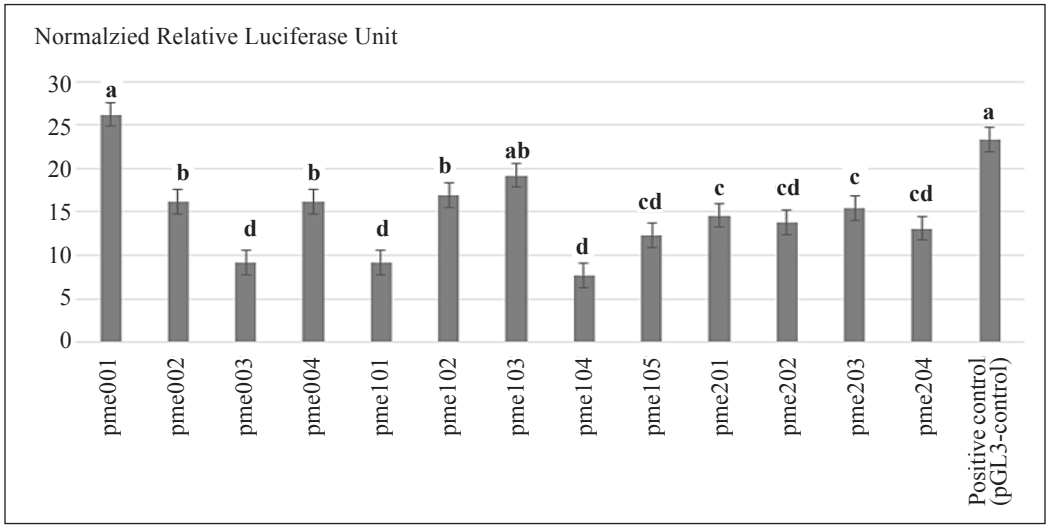


Figure 1. The measured normalized relative enhancer luciferase activity (relative to NIC=1) across 13 enhancer candidates. Statistical analysis was conducted using one-way ANOVA and post-hoc Tukey's Test with $p<0.05$. Significantly different data are represented by different alphabets

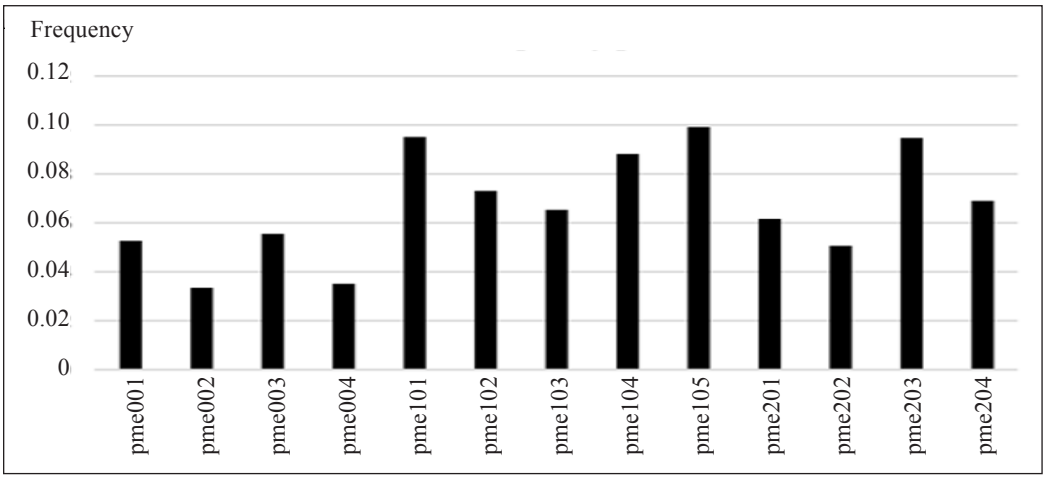


Figure 2. The predicted TFBS frequency per nucleotide across 13 enhancers

The TFBS composition of all 13 candidate enhancers were also examined for its association with the resultant enhancer activity (Figure 3). The enhancer pme001 has the greatest number of activator protein-1 (AP-1) TFBS within its sequences and this may be the reason for its high enhancer activity among all. In other study, the mutation of AP-1 and ATF (activating transcription protein) has resulted in at least two-fold reduction in expression of 87% of the enhancers tested (Grossman et al., 2017). The CCAAT enhancer binding protein beta (C/EBP-β) TFBS was most

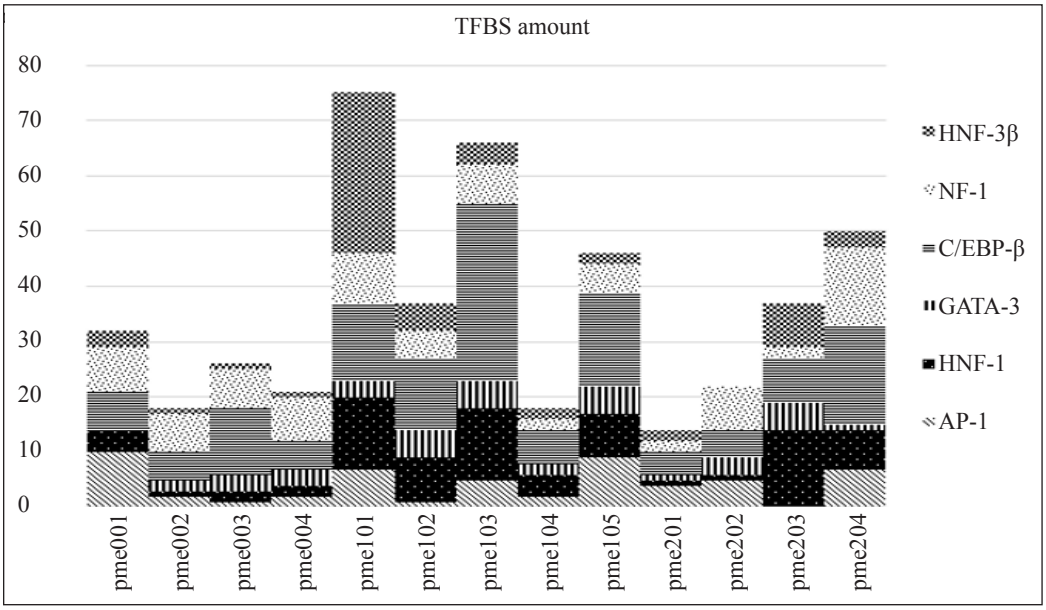


Figure 3. The predicted TFBS composition (based on six most abundant liver-specific transcription factors) across 13 enhancers

abundantly found in enhancer pme103 and it was shown in this study to be the main driver in placing this enhancer in the second place in terms of luciferase activity level among the 13 enhancers. As one the of vital players in the liver, the C/EBP-β is indispensable in various liver-specific processes such as adipose tissue differentiation, inflammation and metabolism (van der Krieken et al., 2015). The enhancer pme101 is one of the lowest activity enhancers but it was found to have high amount of hepatocyte nuclear factor 3 beta (HNF-3β) TFBS within its sequence. This phenomenon may suggest the supressing role of HNF-3β TFBS. However, in other literatures, the HNF-3β was found to contribute substantially towards liver-specific enhancers and human prothrombin gene enhancers (Ceelie et al., 2003; Kim et al., 2011). According to Chen et al. (2015), the contribution of TFBS-clustered regions towards cellular functions and enhancer activities exceeds that of individually scattered TFBS within the enhancer. Hence, this leads us to further investigate the TFBS distribution pattern of these proboscis monkey enhancers.

The TFBS distribution pattern of the 13 selected enhancers was inspected at every 100 bp interval to discover more on their associations with enhancer activity (Figure 4). Apart from having the most abundant AP-1 TFBS within its sequence, the enhancer pme001 had also four AP-1 adjacent pairs located at four different 100 bp intervals. This may explain for the two-fold difference in enhancer activity between enhancer pm001 and pme105 (containing two adjacent AP-1 pairs). According to Kerppola and Curran (1993), pairs of AP-1 TFBS which are capable of bending DNA can improve enhancer functioning more than that of individual one. The enhancer pme103 was found to encompass five adjacent C/EBP-β TFBS as well as two triplets of HNF-1TFBS which had unleashed the expression

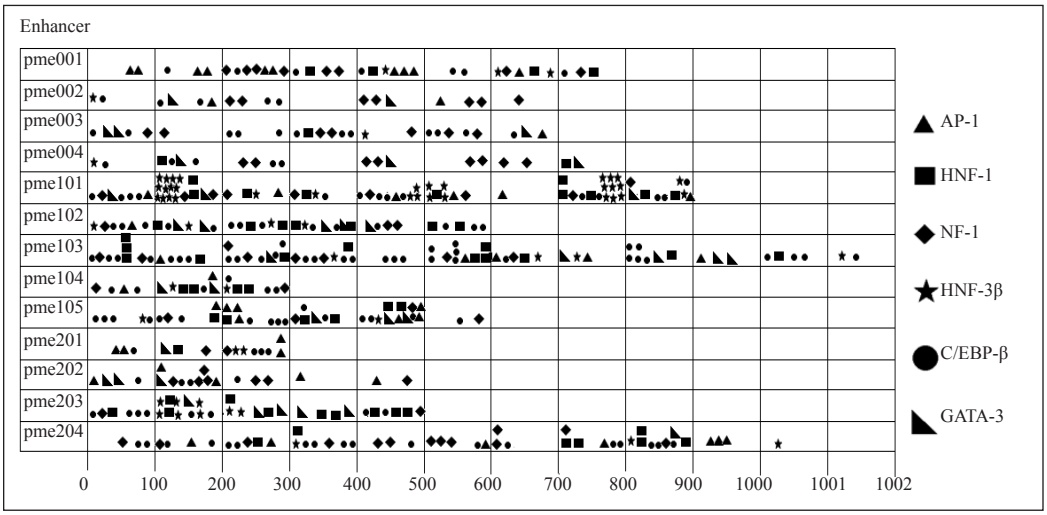


Figure 4. The TFBS distribution pattern per 100 bp interval of 13 enhancers. Stacking TFBS represents overlapping TFBS sites

capability of this enhancer in HepG2 cell line. The C/EBP-β has the ability to recruit other key transcription factors (including its counterparts) to orchestrate the proliferation, antiapoptotic responses as well as survivability of multiple myeloma (Pal et al., 2009) whereas the HNF-1 is essential regulator of various liver specific genes such as albumin, CYP2E1, HBV, Apolipoprotein B and Apolipoprotein AII gene (Maire et al., 1989). Both of these liver-specific TFBSs composite the strong enhancer activity in pme103. There are two huge clusters of HNF-3β TFBS found exclusively in the enhancer pme101 and we have proven in this study that this accounts for the low enhancer activity outcome in HepG2 cell line. Despite many lines of evidence pointing out on the enhancing effects of HNF-3β onto regulative activities in the liver (Ceelie et al., 2003; Rausa et al., 1997; Schrem et al., 2002), the consequence of deletion of large clusters of HNF-3β TFBS in proboscis monkey enhancers remains unproven until this study, to the best of our knowledge. Thus, deletion analysis was designed to elucidate the deletional effects of these TFBS regions in the selected proboscis monkey enhancers.

Deletion Analysis

The deletion analysis was performed onto selected enhancers (pme001, pme101 and pme103) to determine the contributions of the TFBS-enriched regions towards enhancer activities (Figures 5, 6 and 7). The internal primer pairs used for deletion analysis was listed in Table 2 and the deletion descriptions are shown in Table 3. In general, the deletion of each pair of AP-1 or the triple AP-1 region had resulted in reduction of enhancer activity as compared to the wildtype enhancer. However, the enhancer activity level of modified enhancer fragment pme001a and pme001D were the same even though the pme001a only contains the first two adjacent AP-1 pairs whereas the pme001D has all three adjacent AP-1 pairs with the tailing AP-1 TFBS. On the other hand, modified enhancer fragment pme001b has all three adjacent AP-1 TFBS pair sets (similar to pme001D but without the tailing AP-1 TFBS) and had scored a much higher expression level in HepG2 cell line compared to pme001D. On the side note, no new TFBS had been introduced following the deletion done onto the modified enhancer fragment pme001D. Fonseca et al. (2018) found that collaborative interactions between TFs could affect the binding affinity of all AP-1 monomers greatly in macrophage. Therefore, it is postulated that in the absence of the triple AP-1 region, the tailing AP-1 TFBS located nearest to the 3' end has disrupted DNA bending patterns and subsequently the collaborative interactions of all TFs within enhancer pme001.

The deletion analysis conducted onto enhancer pme101 had yielded modified enhancer fragments with activity level higher than that of the wildtype enhancer. The deletion of the first HNF-3β-enriched regions containing 12 HNF-3β TFBS from enhancer pme101 had elevated the enhancer activity to 1.7-fold (pme101D1) whereas the deletion of the second HNF-3β-enriched region containing 8 HNF-3β TFBS from the

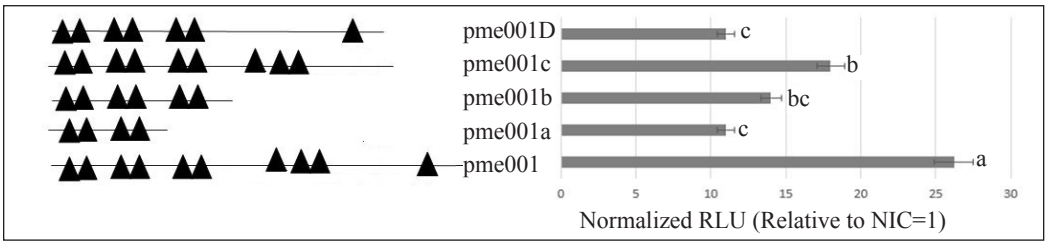


Figure 5. Overview of selected TFBS distribution and deletion analysis of enhancer pme001 in pGL4.23 [*luc*/minP], activities as measured in HepG2 cell line. Deletion constructs are shown on the left. AP-1 TFBS is represented by black triangle. Enhancer activities are shown on the right with values indicating normalized firefly luciferase: *Renilla* luciferase activities relative to no-insert control (NIC). Results are mean ± S.E. (n=3). Mean values were subjected to One-Way ANOVA and Tukey's test, $p<0.05$. Mean values represented by different alphabets are significantly different

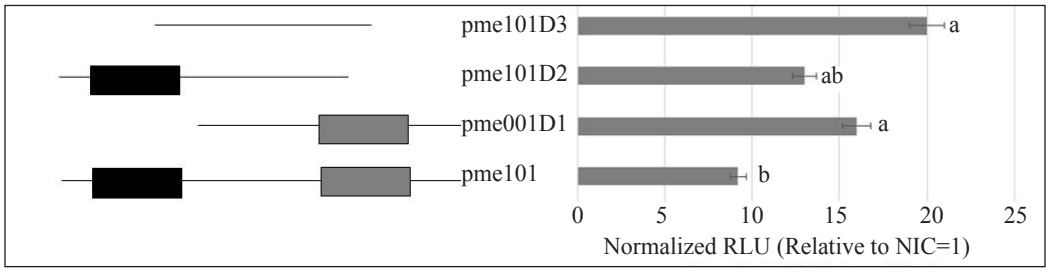


Figure 6. Overview of selected TFBS distribution and deletion analysis of enhancer pme101 in pGL4.23 [*luc*/minP], activities as measured in HepG2 cell line. Deletion constructs are shown on the left. The first HNF-3β TFBS-enriched region (12 HNF-3β TFBSs) is represented by black rectangle; The second HNF-3β TFBS-enriched region (8 HNF-3β TFBSs) is represented by grey rectangle. Enhancer activities are shown on the right with values indicating normalized firefly luciferase: *Renilla* luciferase activities relative to no-insert control (NIC). Results are mean ± S.E. (n=3). Mean values were subjected to One-Way ANOVA and Tukey's test, $p<0.05$. Mean values represented by different alphabets are significantly different

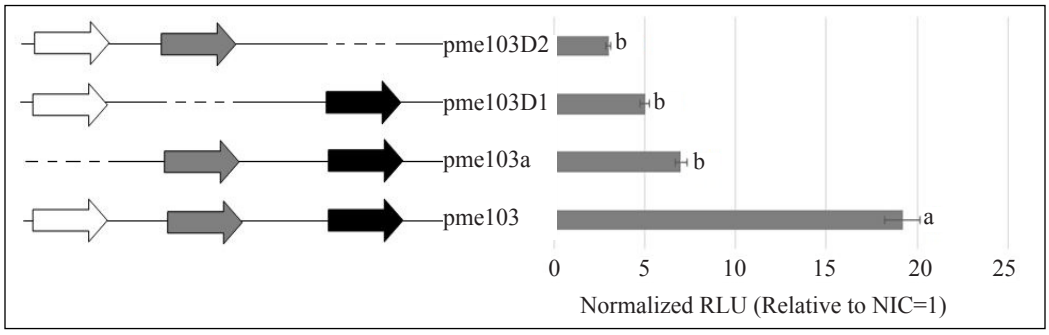


Figure 7. Overview of selected TFBS distribution and deletion analysis of enhancer pme103 in pGL4.23 [*luc*/minP], activities as measured in HepG2 cell line. Deletion constructs are shown on the left. The 1-100 bp region is represented by white arrow; The 500-599 bp are represented by grey arrow; The 800-899 bp are represented by black arrow; Deleted regions are represented by dotted lines. Enhancer activities are shown on the right with values indicating normalized firefly luciferase: *Renilla* luciferase activities relative to no-insert control (NIC). Results are mean ± S.E. (n=3). Mean values were subjected to One-Way ANOVA and Tukey's test, $p<0.05$. Mean values represented by different alphabets are significantly different

enhancer pme101 had yielded a 1.4-fold increase in activity (pme101D2). In the case of modified enhancer fragment pme101D3, the deletion of the both abovementioned HNF-3β-enriched regions had produced enhancer activity 2.2-fold higher than that of the wildtype enhancer pme101. Despite the repressing activity of the large HNF-3β clusters has never been reported previously. Li et al. (2002) however discovered that NKX2.1 could cooperate with HNF-3β to induce repression effects onto the transcription of surfactant protein B (*Sp-B*) gene. While the underlying mechanism of this phenomenon remains unknown, this study revealed the suppressing effects of huge clusters of the HNF-3β TFBS onto proboscis monkey enhancer.

The major players in regulating the enhancer activity of pme103 have been revealed following the deletion analysis: C/EBP-β and HNF-1 transcription factors. The deletion of the first C/EBP-β-HNF-1-enriched region (three C/EBP-β and three HNF-1 TFBSs overlapping in 1-100 bp region) had resulted in 2.7-fold activity reduction. A much higher decline (3.8-fold) in enhancer activity was observed in modified enhancer fragment pme103D1 with the second C/EBP-β-HNF-1-enriched region (five C/EBP-β and three HNF-1 TFBSs overlapping in 500-599 bp region) being deleted. The most significant decrease (6.4-fold) in enhancer activity was observed in modified enhancer fragment pme103D2 with the adjacently located five C/EBP-β-enriched region (800-899 bp region) being eliminated. The result from this study is consistent with that conducted by Plachetka et al. (2008) where they deduced that the opening of chromatin at *mim-1* enhancer could be activated by C/EBP-β when several binding sites were present within a close distance. Thus, it can be deduced that adjacently located five C/EBP-β-enriched region has the most significant influence towards pme103 enhancer activity in HepG2 cell line.

Comparison of Enhancer Activity across Two Cell Lines

In this study, the selected 13 candidate enhancers were tested for their activity in two cell lines, namely A549 (human lung carcinoma) and HepG2 (human liver carcinoma) and the comparison graphs are depicted in Figure 8. Interestingly, all studied enhancers have shown a significantly (unpaired student's t-test, $p<0.05$) lower luciferase activity level in A549 cell line compared to HepG2 cell line. This is a strong indication of high cell-type functional specificities of the liver-specific enhancers identified in this study. Similarly, Duan and Simpson-Haidaris (2006) had elucidated the liver tissue-specific role of IL-6-receptor-gp100-Stat3 signalling in Interleukin-6 (IL-6) initiation of γFBG promoter following the discovery that the promoter activity in HepG2 was 15-fold higher than that in A549.

As for the modified enhancer fragments generated from the deletion analysis, most of the modified enhancer fragments exhibited higher activity level than their wildtype enhancer in A549 cell line, with the

exception for modified enhancer fragments pme001a, pme001b and pme001c. The deletion of the TFBS-enriched regions in enhancer pme101 and pme103 had resulted in a significant (unpaired student's t-test, $p<0.05$) increment in enhancer activity in A549 cell line. This suggests that these TFBS-enriched regions are responsible for the maintenance of the cell-type functional specificities in the liver. In the case of the enhancer pme001, the modified enhancer fragments pme001a, pme001b and pme001c did not show higher enhancer activity in A549 cell line compared to their wildtype enhancer. This condition suggests that these three AP-1 TFBS pairs are not involved in the regulation of liver enhancer cell-specificity, probably due to their miniature length.

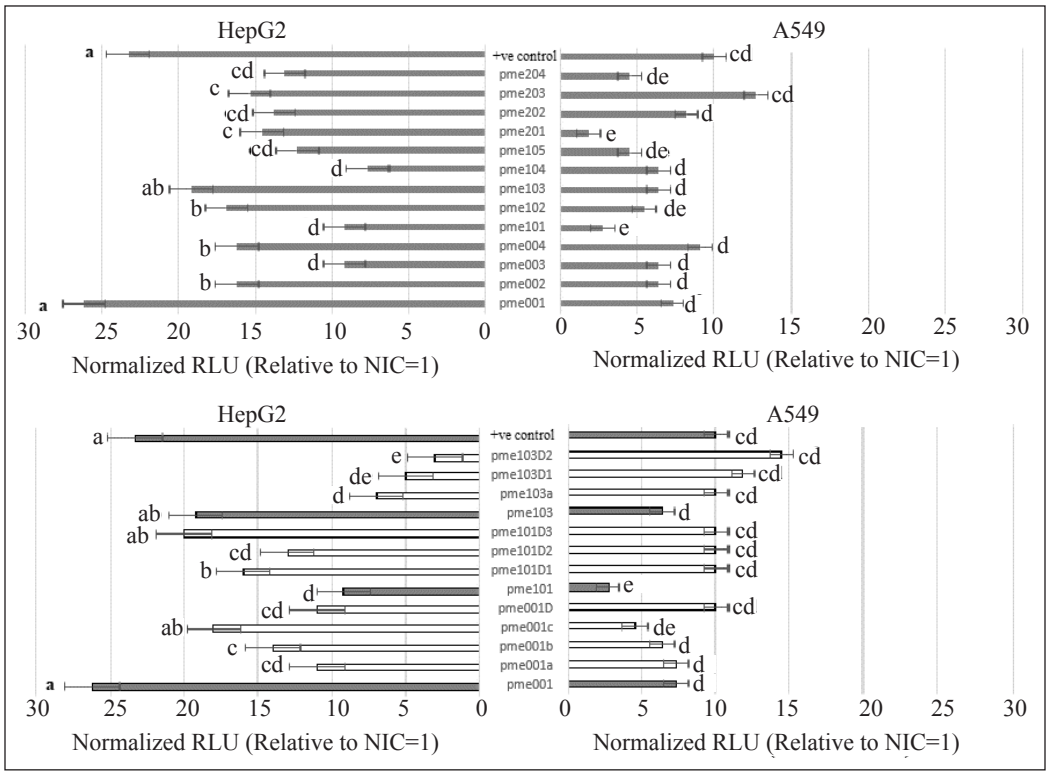


Figure 8. The normalized relative luciferase unit (relative to NIC=1) of wildtype enhancer and modified enhancer fragments in both HepG2 and A549 cell lines. Statistical analysis was conducted using one-way ANOVA and post-hoc Tukey's Test with $p<0.05$. Significantly different data are represented by different alphabets

CONCLUSION

In this study, a total of 13 liver-specific enhancers were isolated from the genome of the proboscis monkey, forming a population of enhancers identified via three different criterions: functional conservation, sequence conservation as well as combinatorial features. The luciferase reporter assay had

provided the information on the enhancer activities and the associations of these activities with TFBS are briefly investigated. The TFBS distribution pattern at every 100 bp interval had revealed a clearer picture on the possible contributors towards enhancer activities compared to the overall TFBS abundance and TFBS composition.

The deletion analysis was performed to further verify the influence of the TFBS-enriched regions exclusively identified in selected enhancers that accounted for their activities. Pairs of AP-1 TFBS and clusters of C/EBP-β TFBS were found to possess enhancing effects onto liver-specific enhancers of proboscis monkey whereas huge clusters of HNF-3β TFBS were found to have suppressing influences on enhancer activities in HepG2 cell line. Besides, these TFBS-enriched regions (except for AP-1 TFBS pairs) had substantial function in maintaining cell-type specificities in liver-specific enhancers of the proboscis monkey. This study serves as a preliminary exploration into the enhancer landscape of the proboscis monkey and as stepping stone for future conservation research for this endangered non-human primate. In future, *in vivo* studies using mice as model organism are essential for the verification of enhancer activities in the living organism context.

ACKNOWLEDGEMENT

This study was fully funded by the Ministry of Higher Education Malaysia, Fundamental Research Grant Scheme: FRGS/SG03(01)/1134/2014(01). We would like to particularly thank Mr Ng Kar Hon for the proboscis monkey faecal sample collection from Bako National Park, Sarawak, Malaysia.

REFERENCES

Abdullah, M. T., Mat Daud, M. H. R., Nur Aida, M. T., Idris, A., Croft, L., Saidin, A., … Hercus, R. (2014). *Draft genome sequence of* Nasalis larvatus *enables comparative analysis of 10 simian genomes* (Unpublished manuscript), Universiti Malaysia Sarawak, Malaysia.

Alipanahi, B., Delong, A., Weirauch, M., & Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology*, *33*(8), 831-839.

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*(2), 299-308.

Benoist, C., & Chambon, P. (1981). *In vivo* sequence requirements of the SV40 early promoter region. *Nature*, *290*(1), 304-310.

Bery, A., Martynoga, B., Guillemot, F., Joly, J. S., & Rétaux, S. (2014). Characterization of enhancers active in the mouse embryonic cerebral cortex suggests Sox/Pou *cis*-regulatory logics and heterogeneity of cortical progenitors. *Cerebral Cortex*, *24*(11), 2822-2834.

Bigoni, F., Stanyon, R., Wimmer, R., & Schempp, W. (2003). Chromosome painting shows that the proboscis monkey (*Nasalis larvatus*) has a derived karyotype and is phylogenetically nested within Asian Colobines. *American Journal of Primatology*, *60*(3), 85-93.

Blackwood, E. M., & Kadonaga, J. T. (1998). Going the distance: A current view of enhancer action. *Science's Compass*, *281*(5373), 60-63.

Boyd, J. L., Skove, S. L., Rouanet, J. P., Pilaz, L. J., Bepler, T., Gordân, R., … Silver, D. L.

(2015). Human-chimpanzee differences in a *FZD8* enhancer alter cell cycle dynamics in the developing neocortex. *Current Biology*, *25*(6), 772-779.

Cao, Q., & Yip, K. Y. (2016). A survey on computational methods for enhancer and enhancer target predictions. In K. Wong (Ed.), *Computational biology and bioinformatics: Gene regulation* (pp. 3-27). New York, NY: CRC Press.

Ceelie, H., Spaargaren-van Riel, C. C., De Jong, M., Bertina, R. M., & Vos, H. L. (2003). Functional characterization of transcription factor binding sites for HNF1-alpja, HNF3-beta (FOXA2), HNF4-alpha, Sp1 and Sp3 in the human prothrombin gene enhancer. *Journal of Thrombosis and Haemostasis*, *1*(8), 1688-1698.

Chen, H., Li, H., Liu, F., Zheng, X., Wang, S., Bo, X., & Shu, W. (2015). An integrative analysis of TFBS-clustered regions reveals new transcriptional regulation models on the accessible chromatin landscape. *Scientific Reports*, *5*(8465), 1-13.

de Kok, Y. J., Vossenaar, E. R., Cremers, C. W., Dahl, N., Laporte, J., Hu, L. J., … Cremers, F. P. (1996). Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Human Molecular Genetics*, *5*(9), 1229-1235.

Duan, H. Q., & Simpson-Haidaris, J. (2006). Cell type-specific differential induction of the human γ-fibrinogen promoter by interleukin-6. *The Journal of Biological Chemistry*, *281*(18), 12451-12457.

Emison, E. S., McCallion, A. S., Kashuk, C. S., Bush, R. T., Grice, E., Lin, S., … Chakravarti, A. (2005). A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature*, *434*(7035), 857-863.

Firpi, H., Ucar, D., & Tan, K. (2010). Discover regulatory DNA elements using chromatin signatures and artificial neural network. *Bioinformatics*, *26*(13), 1579-1586.

Fonseca, G., Tao, J., Westin, E. M., Duttke, S. H., Spann, N. J., Strid, T., … Glass, C. K. (2018). *Diverse motif ensembles specify non-redundant DNA binding activities of AP-1 family members in macrophages*. Retrieved October 26, 2018, from https://doi.org/10.1101/345835.

Ghandi, M., Lee, D., Mohammad-Noori, M., & Beer, M. A. (2014). Enhanced regulatory sequence prediction using gapped *k*-mer features. *PLoS Computational Biology*, *10*(7), e1003711.

Gotea, V., Visel, A., Westland, J. M., Nobrega, M. A., Pennacchio, L. A., & Ovcharenko, I. (2010). Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Research*, *20*(5), 565-577.

Grice, E. A., Rochelle, E. S., Green, E. D., Chakravarti, A., & McCallion, A. S. (2005). Evaluation of the RET regulatory landscape reveals the biological relevance of a HSCR-implicated enhancer. *Human Molecular Genetics*, *14*(24), 3837-3845.

Grossman, S. R., Zhang, X., Wang, L., Engreitz, J., Melnikov, A., Rogov, P. … Lander, E. S. (2017). Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proceedings of the National Academy of Sciences*, *114*(7), E1291-E1300. doi:10.1073/pnas.1621150114

Groves, C. P. (2001). *Primate taxonomy*. Washington, USA: Smithsonian Institution Press.

Gruss, P., Dhar, R., & Khoury, G. (1981). Simian virus 40 tandem repeated sequences as an element of the early promoter. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(2), 943-947.

He, W., & Jia, C. (2016). EnhancerPred: A predictor for discovering enhancers based on

the combination and selection of multiple features. *Scientific Reports*, *6*(38741), 1-7.

He, W., & Jia, C. (2017). EnhancerPred2.0: Predicting enhancers and their strength based on position-specific trinucleotide propensity and electron-ion interaction potential feature selection. *Molecular Biosystems*, *13*(4), 767–774.

Hu, Z., Hu, B., & Collins, J. F. (2007). Prediction of synergistic transcription factors by function conservation. *Genome Biology*, *8*(12), R257.

Irwin, D. M., Kocher, T. D., & Wilson, A. C. (1991). Evolution of the cytochrome *b* gene in mammals. *Journal of Molecular Evolution*, *32*(2), 128-144.

Kel, A. E., Gössling, E., Reuter, I., Cheremuskhin, E., Kel-Morgoulis, O. V., & Wingender, E. (2003). MATCH: A tool for searching transcription factor biding sites in DNA sequences. *Nucleic Acids Research*, *31*(13), 3576-3579.

Kerppola, T. K., & Curran, T. (1993). Selective DNA bending by a variety of bZIP proteins. *Molecular and Cell Biology*, *13*(9), 5479-5489.

Khoury, G., & Gruss, P. (1983). Enhancer elements. *Cell*, *33*(2), 313-314.

Kibbe, W. A. (2007). OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Research*, *35*(suppl_2), W43-46.

Kim, M. J., Skewes-Cox, P., Fukushima, H., Hesselson, S., Yee, S. W., Ramsey, L. B., … Ahituv, N. (2011). Functional characterization of liver enhancers that regulate drug-associated transporters. *Clinical Pharmacology and Therapeutics*, *89*(4), 571-578.

Kim, S. G., Harwani, M., Grama, A., & Chaterji, S. (2016). EP-DNN: A deep neural network-based global enhancer prediction algorithm. *Scientific Reports*, *6*(1), 38433.

Kleftogiannis, D., Kalnis, P., & Bajic, V. B. (2015). Progress and challenges in bioinformatics approaches for enhancer identification. *Briefings in Bioinformatics*, *17*(6), 967-979. doi:10.1093/bib/bbv101, 1-13.

Kochagul, V., Srivorakul, S., Boonsri, K., Somgrid, C., Sthitmatee, N., Thitaram, C., & Pringproa, K. (2018). Production of antibody against elephant endotheliotropic herpesvirus (EEHV) unveils tissue tropisms and routes of viral transmission in EEHV-infected Asian elephants. *Scientific Reports*, *8*(4675), 1-12.

Kvon, E. (2015). Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics*, *106*(3), 185-192.

Lettice, L. A., Horikoshi, T., Heaney, S. J., van Baren, M. J., van der Linde, H. C., Breedveld, G. J., … Noji, S. (2002). Disruption of a long-range cis-acting regulator for *Shh* causes preaxial polydactyly. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(11), 7548-7553.

Li, C., Zhu, N. L., Tan, R. C., Ballard, P. L., Derynck, R., & Monoo, P. (2002). Transforming growth factor-β inhibits pulmonary surfactant protein B gene transcription through SMAD3 interactions with NKX2.1 and HNF-3 transcription factors. *The Journal of Biological Chemistry*, *277*(41), 38399-38408.

Lim, L. W. K., Chung, H. H., Chong, Y. L., & Lee, N. K. (2018a). A survey of recently emerged genome-wide computational enhancer predictor tools. *Computational Biology and Chemistry*, *74*(1), 132-141.

Lim, L. W. K., Chung, H. H., Chong, Y. L., & Lee, N. K. (2019). Enhancers in proboscis monkey: A primer. *Pertanika Journal of Tropical Agricultural Science*, *42*(1), 261-276.

Lim, L. W. K., Tan, H. Y., Aminan, A. W., Jumaan, A. Q., Moktar, M. Z., Tan, S. Y., … Sulaiman, B. (2018b). Phylogenetic and expression of ATP-binding cassette transporter genes in *Rasbora*

sarawakenesis. *Pertanika Journal of Tropical Agricultural Science*, *41*(3), 1341-1354.

Liu, B., Fang, L., Long, R., Lan, X., & Chou, K. (2015). iEnhancer-2L: A two layer predictor for identifying enhancers and their strength by pseudo *k*-tuple nucleotide composition. *Bioinformatics*, *32*(3), 362-369.

Liu, F., Li, H., Ren, C., Bo, X., & Shu, W. (2016). PEDLA: Predicting enhancers with a deep learning-based algorithmic framework. *Scientific Reports*, *6*(28517), 1-14.

Lusk, R. W., & Eisen, M. B. (2010). Evolutionary mirages: Selection on binding site composition creates the illusion of conserved grammars in *Drosophila* enhancers. *PLoS Genetics*, *6*(1), e1000829.

Maire, P., Wuarin, J., & Schibler, U. (1989). The role of cis-acting promoter elements in tissue-specific albumin gene expression. *Science*, *244*(4902), 343-346.

Matsuda, I., Murai, T., Clauss, M., Yamada, T., Tuuga, A., Bernard, H., & Higashi, S. (2011). Regurgitation and remastication in the foregut-fermenting proboscis monkey (*Nasalis larvatus*). *Biology Letters*, *7*(5), 786-789.

McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R., … Cohen, J. C. (2007). A common allele on chromosome 9 associated with coronary heart disease. *Science*, *316*(5830), 1488-1491.

Meijaard, E., Nijman, V., & Supriatna, J. (2008). Nasalis larvatus. *The IUCN red list of threatened species*. Retrieved September 11, 2018, from https://www.iucnredlist.org/details/14352/0

Omar, N., Wong, Y. S., Chong, Y. L., Abdullah, M. T., & Lee, N. K. (2017). Enhancer prediction in proboscis monkey genome: A comparative study. *Journal of Telecommunication, Electronic and Computer Engineering*, *9*(1), 175-179.

Ovcharenko, I., Nobrega, M. A., Loots, G. G., Stubbs, L. (2004). ECR Browser: A tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Research*, *32*(suppl_2), W280-W286.

Pal, R., Janz, M., Galson, D. L., Gries, M., Li, S., Jöhrens, K., … Lentzsch, S. (2009). C/EBPβ regulates transcription factors critical for proliferation and survival of multiple myeloma cells. *Blood*, *114*(18), 3890-3898.

Palmer, J. M., Drees, K. P., Foster, J. T., & Lindner, D. L. (2018). Extreme sensitivity to ultraviolet light in the fungal pathogen causing white-nose syndrome of bats. *Nature Communications*, *9*(35), 1-10.

Peng, Y. Z., Pan, R. L., & Jablonski, N. G. (1993). Classification and evolution of Asian colobines. *Folia Primatologica*, *60*(1-2), 106-117.

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bajerano, G. (2015). Enhancers: Five essential questions. *Nature Review Genetics*, *14*(4), 288-295.

Plachetka, A., Chayka, O., Wilczek, C., Melnik, S., Bonifer, C., & Klempnauer, K. H. (2008). C/EBPβ induces chromatin opening at a cell-type-specific enhancer. *Molecular and Cellular Biology*, *28*(6), 2102-2112.

Prescott, S. L., Srinivasan, R., Marchetto, M. C., Grishina, I., Narvaiza, I., Selleri, L., … Wysocka, J. (2015). Enhancer divergence and *cis*-regulatory evolution in the human and chimp neural crest. *Cell*, *163*(1), 68-83.

Rausa, F., Samadani, U., Ye, H., Fletcher, C. F., Jenkins, N. A., Copeland, N. G., & Costa R. H. (1997). The cut-homeodomain transcriptional activator HNF-6 is coexpressed with its target gene HNF-3β in the developing murine liver and pancreas. *Developmental Biology*, *192*(2), 228-246.

Schrem, H., Klempnauer, J., & Borlak, J. (2002). Liver-enriched transcription factor in liver function and development. Part I: The hepatocyte nuclear factor network and liver-specific gene expression. *Pharmacological Reviews*, *54*(1), 129-158.

Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: From properties to genome-wide predictions. *Nature Reviews Genetics*, *15*(4), 272-286.

Stewart, C. B., Schilling, J. W., & Wilson, A. C. (1987). Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature*, *330*(6146), 401-404.

Tamrin, N. A. (2016). *Whole genome sequencing and assembly of proboscis monkey (*Nasalis larvatus*: Cercopithecidae) and evolutionary history of sweet taste receptor* TAS1R2 *gene within Simiiformes (Primates: Haplorhini)* (Doctoral dissertation), Universiti Malaysia Sarawak, Malaysia.

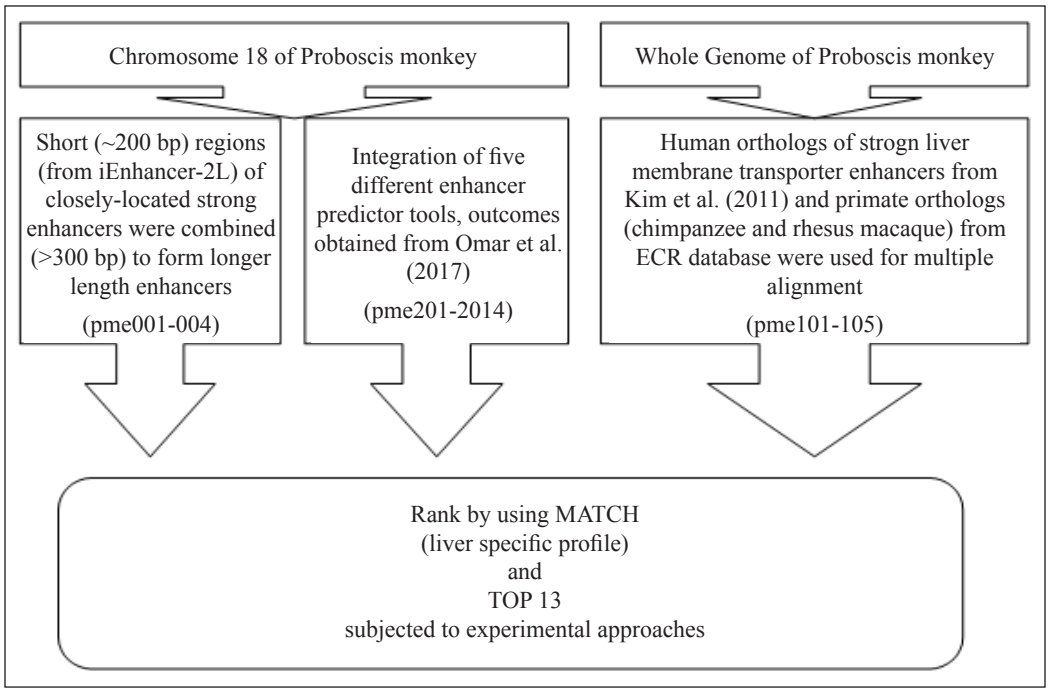
Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3: New capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115.

van der Krieken, S. E., Popeijus, H. E., Mensink, R. P., & Plat, J. (2015). CAAT/enhancer binding protein β in relation to ER stress, inflammation, and metabolic disturbances. *BioMed Research International*, *2015*, 1-13.

Wong, Y. S., Lee, N. K., & Omar, N. (2016). GMFR-CNN: An integration of gaped motif feature representation and deep learning approach for enhancer prediction. *Proceedings of the 7th International Conference on Computational Systems-Biology and Bioinformatics* (pp. 41-47). Beijing, China: Association for Computing Machinery.

Yegnasubramanian, S., Wu, Z., Haffner, M. C., Esopi, D., Aryee, M. J., Badrinath, R., … Nelson, W. G. (2011). Chromosome-wide mapping of DNA methylation patterns in normal and malignant prostate cells reveals pervasive methylation of gene-associated and conserved intergenic sequences. *BMC Genomics*, *12*(313), 1-19.

APPENDIX



Supplementary Figure 1. Flow chart of computational enhancer predictions