MEDWIN PUBLISHERS
Committed to Create Value for Researchers

International Journal of Zoology and Animal Biology
ISSN: 2639-216X

Recently Emerged Genome Wide Computational Enhancer Target Prediction Tools: A Brief Survey

Lim LWK*

Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Malaysia

***Corresponding author:** Leonard Whye Kit Lim, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Malaysia, Email: limwhyekitleonard@gmail.com

Review Article

Volume 6 Issue 1

Received Date: January 04, 2023

Published Date: January 31, 2023

DOI: 10.23880/izab-16000440

Abstract

Enhancers are non-coding genomic regulatory elements capable of elevating gene transcription in various biological as well as developmental stages in the host organism. Discovered since 1981, the enhancers play major roles in genetic disease onset and development, orchestrating gene regulation patterns even across the same species via the sequence variations. To date, predicting enhancers and their targets remain a daunting task as universal enhancer markers are yet to be discovered. Computational enhancer target prediction involves three major approaches: supervised, unsupervised and semi-supervised machine learning methods which work on enhancer target features such as enhancer-promoter distance, closest promoter, co-conservation and correlation of molecular signals. In this review, we introduced some recently emerged enhancer target prediction tools as well as their modus operandi, in hope that we can provide future directions towards the development of a more robust tool to aid in the advancement of enhancer targeted treatment researches.

Keywords: Enhancer Target Prediction; Supervised Learning; Unsupervised Learning; Semi-Supervised Learning

Abbreviations: KB: Kilo Basepair; TSS: Transcription Start Site; AUC: Area Under Curve; DHS: DNase I Hypersensitive site; IMM: Interpolated Markov Chain Model; EM: Expectation Maximization.

Introduction

The central dogma forms the foundation of molecular biology and it is widely deemed as the source of all life form on earth [1-10]. The non-coding regions within the genome are the ones responsible for various biological and developmental processes, especially the major regulatory elements such as the promoters and enhancers [11]. Promoters are comparably easier to be identified from the gene it regulates in the genome, with length ranging within a few kilo basepair (kb), located in proximity with the transcription start site (TSS) of coding or non-coding genes. As for enhancers, it is a different story on the other side of the spectrum. Enhancers are *cis*-regulatory elements that can initiate gene transcription elevation, even from a distance and regardless of orientation, contributing to various disease-related biological progressions known up to now [12-14].

The term 'enhancer' was first coined by De Villiers, et al. [15] in an attempt to describe a short (72 bp) DNA sequence repeat with the capability of triggering and elevating the gene transcription of β-globin gene of rabbit. Then, Banerji, et al. [16] further identified the SV40 enhancer found to heighten the expression of beta-globin gene in HeLa cell line. Generally, enhancers are short DNA elements (50-1500 bp) that functions as platforms for the binding of transcription

factors [12]. These transcription factors then work in tandem to collectively contribute to the increase in gene transcription ultimately. Enhancers are more versatile than promoters in terms of their readability (both forward and backwards) as well as their mode of action coverage (up to 1 Mbp from upstream or downstream of genes they regulate in one-to-many or many-to-many manner) [12]. Several enhancers can even function in the form of enhancer-originating RNAs (eRNAs) in which the enhancer brings together both the RNA polymerase II as well as the general transcription factors for the eRNAs to be transcribed [17,18].

Thus far, genome-wide enhancer target prediction remains a big predicament in the field of genomics as there is no universal enhancer feature identified, to add on to that, their numerous cell and tissue specificities as well as the lack of enhancer primate model species beside human as reference [14,19-21]. The enhancer target prediction involves fewer types when compared to genome-wide enhancer prediction [22]. While the genome-wide enhancer prediction involves features like sequences, epigenomic modifications and eRNAs, the enhancer target prediction encompasses features such as enhancer-promoter distance, closest promoter, co-conservation and correlation of molecular signals utilizing supervised, semi-supervised as well as unsupervised machine learning approaches. Studying the whole genome gives us a complete view of all valuable information available for the subject species [11,23-36]. In this review, we described in brief the recently emerged genome wide enhancer target prediction tools and further compared their prediction methods. Towards the end of this review, we provided some insights on improvements and development of a more robust enhancer target prediction tool in future.

Gene Regulation by Enhancer

The expression of gene in a given cell or tissue context at a specific spatial and temporal manner is orchestrated by a group of DNA elements named the regulatory components at various biological developmental life stages [37,38]. The gene expressions are usually regulated via a very strict *modus operandi*, at every cellular component the gene product passes to eventually become protein through processes like chromatin remodeling, transcription activation, modifications of transcripts, mRNA degradation, translation, posttranslational modifications as well as protein transport and degradation. This is to ensure the host organism can equipped itself with the required gene products to survive and strive in harsh environments. The gene regulation in eukaryotic organisms are much more complicated than that of the prokaryotes as multi-layered networks and cross-acting regulatory elements are actively involved [39].

Gene regulations in eukaryotes are mostly governed by regulatory elements such as the promoter and enhancers via transcription factor binding. The core promoter can be found in all eukaryotic genes and the TATA box (TATAAAAAA) is one of the most commonly discovered examples [39]. Generally, the core promoter is highly conserved across all protein-coding genes in terms of its structure and binding factor as compared to other upstream promoters [40]. One major feature that the enhancers differ from promoter is that the enhancers are located in the non-coding regions and can be either functionally or sequence-conserved across diverse species. Transcription activation requires the RNA Polymerase II to be recruited to the transcription start site following the transcription initiation signals emitted via the interactions between enhancers and promoter as well as the general transcription factors (TFIIA, -B, -D, -E, -F and -H) and chromatin remodeling complexes (ACF, PBAF, SWI/SNF and RSF) [39].

Back in the year 1981, De Villiers, et al. [15] first define the term 'enhancer' as DNA elements that can significantly elevate the beta-globin gene expression in rabbit. And thus, the first proposed action of the enhancer is the element that could modify the superhelical density of DNA, improve the accessibility of RNA Polymerase II as well as tolerate nuclear matrix binding [15]. Then, Banerji, et al. [16] further identified the SV40 enhancer that are capable of expression elevation of beta-globin gene in HeLa cell line. They cloned the hemoglobin beta I gene harvested from rabbit and further inserted it into a recombinant expression plasmid with pre-inserted SV40 enhancer, as a result, the expression was found to be 200-fold more than that of the negative control. It is from this study also, Banerji, et al. [16] discovered that this SV40 enhancer can work in both orientations and at any distances from the beta-globin gene. Since then, the enhancer discovery in the human genome progresses exponentially with the discovery of enhancers like the sensory vibrissae enhancer, penile spine enhancer, HACNS1 and forebrain subventricular zone enhancer as well as establishment of various enhancer database like FANTOM5 and VISTA [41-43].

Besides, enhancers can also recruit transcription factors and serve as binding dock for transcription activation to take place and their sizes between 50 and 1500 base pair generally [12]. They are mostly cis-acting but can sometimes be trans-acting, with the ability to modulate genes as far as 1 Mbp away regardless of upstream or downstream. Enhancers can also work in the form of enhancer-originating RNAs (eRNAs) where eRNAs can enhance the efficiency of enhancers [17,18]. In this case, RNA Polymerase II is recruited solely by enhancer itself and in turn work in tandem with general transcription factors to transcribe the eRNAs. The vital roles

of enhancers in shaping phenotypes evolutionarily as well as being pivotal in essential biological developmental processes such as anatomy progress and morphogenesis are deemed indispensable in proper cell and tissue functioning. It was not until recently that, the various links between enhancer and genetic diseases emerged through experimental discoveries [41,44], thus created pressing needs for rapid enhancer target identification, especially computationally.

Supervised Machine Learning

The supervised machine learning involves machine learning that is built with strong reliance on high-confidence negative and positively labelled datasets (known enhancer target and non-enhancer target set). This training model is aimed towards the maximum differentiation across the cases from the control sets [45].

One of the supervised machine learning tools for enhancer target predictions that utilizes chromosome conformation data in defining known experimental determined enhancer targets is the IM-PET [46]. IM-PET employs Random Forest classifier to train on ChIA-PET connected enhancer-promoter pairs in both MCF7 as well as K562 cells as positive example datasets. The negative sample dataset used covers the random enhancer-promoter pairs with distances that obeys the background distribution of non-associating genomic loci in a chromatin fiber [47-49]. The four features employed in this enhancer target prediction which includes correlation of enhancer-promoter activity, enhancer-promoter genomic distance, co-conservation of enhancer and promoter sequences as well as transcription factor expression levels that bind to the enhancer. This tool had achieved 94% AUC, which outperformed PresTIGE, nearest-promoter as well as methods utilized by Ernst, et al. [50] and Thurman, et al. [51].

Another sequence feature based tool utilizing supervised machine learning to uncover enhancer targets is the PETModule which is motif module based [50]. The PETModule sourced from P300 ChIP-seq peaks from seven human cell lines, namely HepG2, SK-N-SH, MCF7. H1-hESC, GM12878, K562 and HeLa-S3, as well as known active enhancers from IMR90 cell line. The positive training set was defined by the random selection from a gigantic pool consisting of 1000 ChIA-PET enhancer-promoter pairs in K562, 1000 Hi-C enhancer-promoter pairs in IMR90 as well as 500 ChIA-PET enhancer-promoter pairs in MCF7 [51,52]. The negative training datasets were randomly selected genes within 2 Mb around the enhancers. A fair comparison of PETModule across other previously available tools revealed that this tool had outpaced its predecessors (IM-PET and PresTIGE) in terms of AUC (94.9%), precision (0.205) and F1 score (0.286) across all datasets trained.

The CISMAPPER is one powerful tool that can employs the correlation between gene expression and histone mark of transcription factor binding sites across diverse tissue types to predict regulatory targets of a transcription factor, a state-of-art tool that differs from other tools which uses genomic distance [53]. This tool was designed to generate a total of four ranked lists of predictions based on set of scored (TSS, peak) links: two on genes and TSSs as latent targets of the ChIP-ed TF whereas two on TF ChIP-seq peaks as latent regulators of genes and TSSs [53]. The link score was applied to sort the data entries in ascending order in each group respectively. When compared with other tools that predicts based on genomic distances, the CISMAPPER excelled in terms of TSS target predictions and accuracies across six out of the eight diverse tissues: NHEK, HepG2, HeLa-S3, HUVEC, Ag04450, H1-Hesc, K562 and GM12878.

Unsupervised Machine Learning

The unsupervised machine learning is most feasible for the unearthing of hidden and unprecedented configurations straightforwardly from the data. No known and validated information or datasets was required as input for this type of training, thus diminishing the occurrence of false positives in the output but also narrowing the enhancer target output concurrently due to the scarcity in terms of known enhancer target markers.

Ernst, et al. [50] employed the histone modification profile (namely H3K27ac, H3K4me1 and H3K4me2) correlation across enhancer-promoter pair within 125-kbp range to predict enhancer targets from the nine human cell types: GM12878, HUVEC, HMEC, HSMM, NHEK, H1 ES, NHLF, K562 and HepG2. First, the HMM was trained using chromatin states with 10 data tracks for each cell types. Next, a 15 state model was placed in focus to improve the resolution of biologically-meaningful patterns that has reproducibility across different cell types when independently processed. Locations associated to strong promoter state 1 (or strong enhancer state 4) in at least one cell type was subjected to multi-cell type clustering using the *k*-means algorithm. Utilizing mark intensity-expression correlation data, the logistic regression classifiers were trained to discriminate control pairs from real instances of couples of gene expression data and enhancer states, based on expression data that are arbitrarily re-assigned to dissimilar genes. This approach had achieved a satisfactory result of 67% area under curve (AUC).

The most direct method of predicting enhancer targets is the closest promoter approach applied by Andersson, et al. [45]. They utilized the bidirectional capped RNAs (as measured by CAGE) as predictor of cell enhancer activities.

This method is simple in nature but imperfect in terms of the prediction coverage as the population of enhancers that regulates the nearest promoter covers only 40% of the total genomic enhancer population and multiple promoters can be regulated by one enhancer alone. One way to improve the predictions is to expand the distance range of enhancer target predictions to extend the population of nearest promoter prediction dataset pool [43].

PreSTIGE is another enhancer target prediction tool that pairs the H3K4me1 signals that are cell type-specific with significantly expressed cell type-specific genes to perform its enhancer-promoter pair predictions across 13 different cell types [54]. The multiple linear domain models were utilized to associate cell type-specific enhancers to their target genes. After several evaluation, their finalized domain model were trained to produce the maximum number of predictions, and at the same time, keeping the FDR at its lowest, maintaining the distance boundary at 100 kb as well as considering subset of CTCF sites. For PreSTIGE to function as an enhancer target predictor in a desired cell line, the normalized H3K4me1-enhancer signal must be greater than the background (>10) with the addition circumstances that both the enhancer and the gene must be cell line specific. This approach was found to have identified more enriched enhancer-gene interactions as compared to experimental methods such as ChIA-PET , 3C [55], 5C [56] and eQTL [57-61].

The DNase I hypersensitive site (DHS) correlation among all candidate pairs situated inside the range of 500-kbp gap was applied by Thurman, et al. [51] to accurately predict enhancer-promoter pairs from the genome. Firstly, emulating the protocols from Farazi, et al. [62] and John, et al. [63], the DNaseI hypersensitivity mapping was completed on 125 cell-types [49]. Then, the datasets were sequenced and mapped, allowing maximum two mismatches. Only the sequence reads that map uniquely to the genome and the data were sorted using the algorithm by Boyle, et al. [64] to achieve DNaseI hypersensitive sites localization. In this study, they had successfully discovered 578,905 DHSs with high correlations (R>0.7) to at least one promoter from the 1,454,901 distal (>2.5kb from transcription start site) DHSs pool from 79 diverse types of cells, with the impressive outcome of extensive map of candidate enhancers with the specific genes they regulate.

Semi-Supervised Machine Learning (Co-Training)

The semi-supervised machine learning (or co-training) encompasses both labelled and unlabeled datasets in its algorithm. This type of machine learning is a powerful approach to enlarge the labelled dataset via the inclusion of its own accurate predictions and therefore explains for is suitability in times when labelled dataset is lacking [63].

The McEnhancer is the one tool that utilizes semi-supervised machine learning to predict enhancer targets together with the other co-regulated genes using third-order interpolated Markov chain model (IMM) via expectation maximization (EM) algorithm in *Drosophila melanogaster* [64]. First, the McEnhancer was trained on common sequences from labelled (known) DHS-gene pairs before predicting, at one expression at a time, the unlabeled DHSs with analogous subsequences. Then, The McEnhancer-determined enhancer sets were used to train the sparse *k*-mer-based logistic regression classifiers for expression patterns predictions. One big advantage McEnhancer have over other tools is that it is able to assign multiple target genes to a single enhancer which in turn had improved the resolution of expression classification in a spatial and temporal manner [65,66].

The Future of Enhancer Target Prediction Tool

The current progress in the field of enhancer target predictions had evolved a big step since the discovery of the first enhancer back in the year 1981. Although the available tools are scarce in amount to date, these tools presented in this paper have provided insights as essential milestone towards the complete mapping of enhancers and its targets within the gnome in the future. The CISMAPPER had proven that its one-of-its-kind approach had achieved greater accuracies as compared to the commonly used feature (genomic distance). Moreover, the McEnhancer had achieved yet another breakthrough with its prediction powers extending towards one-to-many enhancer-promoter pairs prediction, and further provide greater resolution for huge *in vivo* regulatory datasets that does not have complete alignment with one another. Hariprakash, et al. [67] have listed all the computational biology solutions pertaining to enhancers-target gene pairs identification in their mini review. Moore, et al. [68] further described a curated benchmark of enhancer-gene interactions with regards to enhancer-target gene prediction methods. These guidelines and benchmark will greatly aid in curating all enhancer-gene outcomes for future comparisons and improvements.

As the outcome of these research were to ultimately be being able to accurately predict enhancer targets in any cell context, the enhancer target prediction tools will have to improve in terms of their robustness where multiple enhancer target features are included for predictions. For instance, the PETModule can be further improved with the addition of features such as CTCF locations. Besides, the dynamic regulation of enhancers under specific conditions is also deemed essential to make this tool more robust in terms of wider application spectrum [69-73]. All in all, it is both

as important to include more enhancer target features into the prediction and also to conduct experimental validations following computational predictions.

References

1. Jee MS, Lim LWK, Dirum MA, Hashim SIC, Masri MS, et al. (2017) Isolation and Characterization of Avirulence Genes in Magnaporthe oryzae. Borneo Journal of Resource Science and Technology 7(1): 31-42.

2. Aminan AW, Lim LWK, Chung HH, Sulaiman B (2020) Morphometric Analysis and Genetic Relationship of *Rasbora* spp. in Sarawak, Malaysia. Tropical Life Sciences Research 31(2): 33-49.

3. Lim LWK, Chung HH (2020) Salt tolerance research in sago palm (Metroxylon sagu Rottb.): past, present and future perspectives. Pertan. J Trop Agric Sci 43(2): 91-105.

4. Yusni NZ, Lim LWK, Chung HH (2020) Mutagenesis Analysis of ABCG2 Gene Promoter of Zebrafish (Danio rerio). Trends in Undergraduate Research 3(2): 53-59.

5. Yeaw ZX, Lim LWK, Chung HH (2020) Mutagenesis Analysis of ABCB4 Gene Promoter of Danio rerio. Trends in Undergraduate Research 3(2): 44-52.

6. Lai PN, Lim LWK, Chung HH (2021) Mutagenesis Analysis of ABCB8 Gene Promoter of Danio rerio. Trends in Undergraduate Research 4(1): 1-8.

7. Lim LWK (2022a) Eco-Economically Indispensable Borneo-Endemic Flora and Fauna: Proboscis Monkey (Nasalis larvatus), Malaysian Mahseer (Tor tambroides), Engkabang (Shorea macrophylla), Sarawak Rasbora (Rasbora sarawakensis) and Sago Palm (Metroxylon sagu). International Journal of Zoology and Animal Biology 5(3): 000381.

8. Lim LWK (2022b) Comparative genomic analysis reveals the origin and global distribution of melon necrotic virus isolates. Gene Reports 29: 101685.

9. Lim LWK, Chung HH, Ishak SD, Waiho K (2021c) Zebrafish (Danio rerio) ecotoxicological ABCB4, ABCC1 and ABCG2a gene promoters depict spatiotemporal xenobiotic multidrug resistance properties against environmental pollutants. Gene Reports 23: 101110.

10. Lim LWK, Chung HH, Lau MML, Aziz F, Gan HM (2021d) Improving the phylogenetic resolution of Malaysian and Javan mahseer (Cyprinidae), Tor tambroides and Tor tambra: whole mitogenomes sequencing, phylogeny and potential mitogenome markers. Gene 791: 145708.

11. Lim LWK, Tan HY, Aminan AW, Jumaan AQ, Moktar MZ, et al. (2018b) Phylogenetic and Expression of Atp-Binding Cassette Transporter Genes in Rasbora sarawakensis. Pertanika Journal of Tropical Agricultural Science 41(3): 1341-1354.

12. Blackwood EM, Kadonaga JT (1998) Going the distance: A current view of enhancer action. Science 281(5373): 60-63.

13. Bulgar M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell 144: 327-339.

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

15. De Villiers J, Schaffner W (1981) A small segment of polyoma virus DNA enhances the expression of a cloned β-globin gene over a distance of 1400 base pairs. Nucleic Acids Research 9(23): 6251-6264.

16. Banerji J, Rusconi S, Schaffner W (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 27(2 Pt1): 299-308.

17. Natoli G, Andrau JC (2012) Noncoding transcription at enhancers: General principles and functional models. Annual Review of Genetics 46(1): 1-19.

18. Melo CA, Drost J, Wijchers PJ, Van de Werken H, de Wit E, et al. (2013) eRNAs are required for p53-dependent enhancer activity and gene transcription. Molecular Cell 49(3): 524-535.

19. Boyd JL, Skove SL, Rouanet JP, Pilaz LJ, Bepler T, et al. (2015) Human-chimpanzee differences in a *FZD8* enhancer alter cell cycle dynamics in the developing neocortex. Current Biology 25(6): 772-779.

20. Lim LWK, Chung HH, Chong YL, Lee NK (2019a) Enhancers in proboscis monkey: A primer. *Pertanika* Journal of Tropical Agricultural Science 42(1): 261-276.

21. Lim LWK, Chung HH, Chong YL, Lee NK (2019b) Isolation and characterization of putative liver-specific enhancers in proboscis monkey (*Nasalis larvatus*). Pertanika Journal of Tropical Agricultural Science 42(2): 627-647.

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

23. Lim LWK, Chung HH, Hussain H, Bujang K (2019c.) Sago palm (Metroxylon sagu Rottb.): now and beyond. Pertan.

J Trop Agric Sci 42(2): 435-451.

24. Lim LWK, Roja JS, Kamar CKA, Chung HH, Liao TTY, et a. (2019d) Sequencing and characterization of complete mitogenome DNA for Rasbora myersi (Cypriniformes: Cyprinidae: Rasbora) and its evolutionary significance. Gene Reports 17: 100499.

25. Lim LWK, Chung HH, Hussain H (2020a) Complete chloroplast genome sequencing of sago palm (Metroxylon sagu Rottb.): molecular structures, comparative analysis and evolutionary significance. Gene Rep 19: 100662.

26. Lim LWK, Chung HH, Hussain H (2020b) Organellar genome copy number variations and integrity across different organs, growth stages, phenotypes and main localities of sago palm (Metroxylon sagu Rottboll) in Sarawak, Malaysia. Gene Reports 21: 100808.

27. Lim LWK, Roja JS, Kamar CKA, Chung HH, Liao TTY, et al. (2020c) Sequencing and characterisation of complete mitogenome DNA for Rasbora sarawakensis (Cypriniformes: Cyprinidae: Rasbora) with phylogenetic consideration. Computational Biology and Chemistry 89: 107403.

28. Chung HH, Kamar CKA, Lim LWK, Liao Y, Lam TT, et al. (2020a) Sequencing and Characterisation of Complete Mitogenome DNA for Rasbora hobelmani (Cyprinidae) with Phylogenetic Consideration J Ichthyol 60: 90-98.

29. Chung HH, Kamar CKA, Lim LWK, Roja JS, Liao Y, et al. (2020b) Sequencing and characterization of complete mitogenome DNA of Rasbora tornieri (Cypriniformes: Cyprinidae: Rasbora) and its evolutionary significance. Genet 99: 67.

30. Chung HH, Lim LWK, Liao Y, Lam TTY, Chong YL (2020c) Sequencing and Characterisation of Complete Mitochondrial DNA Genome for Trigonopoma pauciperforatum (Cypriniformes: Cyprinidae: Danioninae) with Phylogenetic Consideration. Trop Life Sci Res 31(1): 107-121.

31. Lau MML, Lim LWK, Chung HH, Gan HM (2021a) The first transcriptome sequencing and data analysis of the Javan mahseer (Tor tambra). Data Brief 39: 107481.

32. Lau MML, Lim LWK, Ishak SD, Abol-Munafi A, Chung HH (2021b) A Review on the Emerging Asian Aquaculture Fish, the Malaysian Mahseer (Tor tambroides): Current Status and the Way Forward Proc Zool Soc 74: 227-237.

33. Chew IYY, Chung HH, Lim LWK, Lau MML, Gan HM, et al. (2022) Complete chloroplast genome of Shorea macrophylla (engkabang): Structural features, comparative and phylogenetic analysis.

34. Lau MML, Kho CJY, Lim LWK, Sia SC, Chung HH, et al. (2022) Microbiome Analysis of Gut Bacterial Communities of Healthy and Diseased Malaysian Mahseer (Tor tambroides). Malaysian Society for Microbiology 18(2): 170-191.

35. Lim LWK, Hung IM, Chung HH (2022a) Cucumber mosaic virus: global genome comparison and beyond. Malays J Microbiol 18 (1): 79-92.

36. Lim LWK, Liew JX, Chung HH (2022b) Piper yellow mottle virus: A deep dive into the genome. Gene Reports 29: 101680.

37. Scott M (2000) Development: The natural history of genes. Cell 100(1): 1127-1140.

38. Sikora-Wohlfeld W, Ackermann M, Christodoulou EG, Singaravelu K, Beyer A (2013) Assessing computational methods for transcription factor target gene identification based ChIP-seq data. PLoS Comput Biol 9: e1003342.

39. Laybourn P (2001) Gene regulation. Encyclopedia of Genetics 1(1): 803-813.

40. Levine M, Tjian R (2003) Transcription regulation and animal diversity. Nature 424(6945): 147-151.

41. Watson JD, Baker TA, Bell SP, Gann A, Levine M, et al. (2014) Molecular biology of the gene. In: 7th (Edn.), London, United Kingdom: Pearson.

42. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, et al. (2003) The evolution of transcriptional regulation in eukaryotes. Molecular Biology and Evolution 20(9): 1377-1419.

43. Visel A, Minovitsky S, Dubchak I, Pennacchio LA (2007) VISTA enhancer browser: A database of tissue-specific human enhancers. Nucleic Acids Research 35*:* D88-D92.

44. McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, et al. (2011) Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature 471(7337): 216-219.

45. Andersson R, Gebhard C, Miquel-Escalada I, Hoof I, Bornholdt J, et al. (2014) An atlas of active enhancers across human cell types and tissues. Nature 507(7493): 455-461.

46. Kim MJ, Skewes-Cox P, Fukushima H, Hesselson S, Yee SW, et al. (2011) Functional characterization of liver enhancers that regulate drug-associated transporters.

Clinical Pharmacology and Therapeutics 89(4): 571-578.

47. Greene CS, Tan J, Ung M, Moore JH, Cheng C (2014) Big data bioinformatics. Journal of Cellular Physiology 229(12): 1896-1900.

48. He B, Chen C, Teng L, Tan K (2014) Global view of enhancer-promoter interactome in human cells. PNAS 111(21): E2191-2199.

49. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558): 1306-1311.

50. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, et al. (2011) Systematic analysis of chromatin state dynamics in nine human cell types. Nature 473(7345): 43-49.

51. Thurman RE, Rynes E, Humbert R, Viersta J, Maurano MT, et al. (2012) The accessible chromatin landscape of the human genome. Nature 489(7414): 75-82.

52. Zhao C, Li X, Hu H (2016) PETModule: A motif module based approach for enhancer target gene prediction. Scientific Reports 6: 30043.

53. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, et al. (2013) A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503(7475): 290-294.

54. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148: 84-98.

55. O'Connor T, Bodén M, Bailey TL (2017) Cismapper: Predicting regulatory interactions from transcription factor ChIP-seq data. Nucleic Acids Research 45(4): e19.

56. Corradin O, Saiakhova A, Akhtar-Zaidi B, Myeroff L, Willis J, et al. (2014) Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Research 24(1): 1-13.

57. Vernimmen D, Marques-Kranc F, Sharpe JA, Sloane-Stanley JA, Wood WG, et al. (2009) Chromosome looping at the human α-globin locus is mediated via the major upstream regulatory element (HS-40). Blood 114: 4253-4260.

58. Sanyal A, Lajoie BR, Jain G, Dekker J (2012) The long-range interaction landscape of gene promoters. Nature 489: 109-113.

59. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. Nat Genet 39(10): 1217-1224.

60. Schadt EE, Molony C, Chudin E, Hao K, Yang X, et al. (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol *6:* 1020-1032.

61. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, et al. (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464: 773-777.

62. Farazi TA, Spitzer JI, Morozov P, Tuschi T (2011) miRNAs in human cancer. The Journal of Pathology 223(2): 102-115.

63. John S, Sabo PJ, Thurman RE, Sung MH, Biddie SC, et al. (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nature Genetics 43(3): 264-268.

64. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, et al. (2008) High-resoltuion mapping and characterization of open chromatin across the genome. Cell 132(2): 311-322.

65. Huska MR, Ramisch A, Vingron M, Marsico A (2016) Predicting enhancers using a small subset of high confidence examples and co-training. German Conference on Bioinformatics e24071: 1-10.

66. Hafez D, Karabacak A, Krueger S, Hwang YC, Wang LS, et al. (2017) McEnhancer: Predicting gene expression via semi-supervised assignment of enhancers to target genes. Genome Biology 18: 199.

67. Hariprakash JM, Ferrari F (2019) Computational biology solutions to identify enhancers-target gene pairs. Computational and Structural Biotechnology Journal 17: 821-831.

68. Moore JE, Pratt HE, Purcaro MJ, Weng Z (2020) A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. Genome Biology 21: 17.

69. Lim LWK, Lau MML, Chung HH, Hussain H, Gan HM (2021a) First high-quality genome assembly data of sago palm (Metroxylon sagu Rottboll). Data Brief 40: 107800.

70. Lim LWK, Chung HH, Hussain H, Gan HM (2021b) Genome survey of sago palm (Metroxylon sagu Rottboll). Plant Gene 28: 100341.

71. Lim LWK, Chung HH, Gan HM (2022c) Genome-wide identification, characterization and phylogenetic analysis of 52 striped catfish (Pangasianodon hypophthalmus)

ATP-binding cassette (ABC) transporter genes. Tropical Life Sciences Research 33 (2): 257-293.

72. Lim LWK, Chung HH, Chong YL, Lee NK (2018a) A survey of recently emerged computational enhancer predictor tools. Computational Biology and Chemistry 74(1): 132-141.

73. Visel A, Akiyama JA, Shoukry M, Afzal V, Rubin EM, et al. (2009) Functional autonomy of distant-acting human enhancers. Genomics 93(6): 509-513.