(Revised 07/09)

RGC Ref.: N_HKU752/10 NSFC Ref. : 31061160497
(please insert ref. above)

NSFC/RGC Joint Research Scheme
Joint Completion Report

(Please attach a copy of the completion report submitted to the NSFC by the Mainland researcher)

Part A: The Project and Investigator(s)

1. Project Title

Understanding genetic and epigenetic features of active promoters in human/mouse embryonic stem cells

2. Investigator(s) and Academic Department/Units Involved

	Hong Kong Team	Mainland Team
Name of Principal Investigator *(with title)*	Dr. Junwen Wang	Prof. Michael Q. Zhang
Post	Assistant Professor	Professor
Unit / Department / Institution	Department of Biochemistry, The University of Hong Kong	Tsinghua University
Co-investigator(s) *(with title)*		Dr. Xiaowo Wang

3. Project Duration

	Original	Revised	Date of RGC/ Institution Approval *(must be quoted)*
Project Start date	2010.1.1		2010.1.1
Project Completion date	2012.12.31		2012.12.31
Duration *(in month)*	24		24

Part B: The Completion Report

5. Project Objectives

5.1 Objectives as per original application

We aim to compare the transcription modules, evolution, and epigenetic features of different subclasses of active promoters in human and embryonic stem cells, based on following classifications:
1) protein coding, microRNA, and long noncoding RNA genes
2) unidirectional, bidirectional, and divergent transcribed genes
3) genes expressed in embryonic stem cells and in differentiated cells

With better understanding of different subclass of promoters, we will incorporate these features to
4) develop novel algorithms for effective identification of active promoters for a specific cell type

5.2 Revised Objectives

No.

6. Research Outcome

Major findings and research outcome
(maximum 1 page; please make reference to Part C where necessary)

We did a comprehensive study on human promoter, covering the genetic and epigenetic features. By dividing promoters according to three different standards, the connection between promoter feature and functional specificity is shown more clearly. On one hand, CpG island content, conservation value, frequency of TFBS motif and peaks detected by ChIP-seq and long-range chromatin interaction of promoters could be used to classify promoters based on gene expression product and transcriptional direction. On the other hand, DNA methylation and histone modification profiles in promoter would help to decide tissue specificity of gene expression. Through the summary and comparison of these features on grouped promoter, we gain deeper knowledge on promoter characteristics and mechanisms of their functioning. These findings would provide other researchers a reference of a summary of promoter features and potential ideas on the study of promoter function.
Besides, focusing on epigenetic features we found the close relationship between genes' tissue specific expression and histone modification pattern. Histone modification profile in genome wide largely decides gene expression profile in a cell, and thus contributes to cell lineage specification. We trace histone modification profile from ESC to differentiated cells in promoter regions of grouped genes. By classifying genes based on relative expression values, we

summarize common rules concerning transcriptional regulation role of each histone modification. Generally ESC bears more active histone modification marks and during differentiation they are lost gradually. But these active marks, including promoter marks and gene body ones, are gained for tissue specific genes to be activated with differentiation progresses. On the contrary, as to the repressive mark, the profile of H3K27me3 for non-expressed genes shows by default it would increase the signal with differentiation. However, for expressed genes its signal in promoter regions are eliminated during differentiation, but this does not necessarily lead to tissue specific gene expression. Lastly, H4K20me1 shows no common pattern concerning its function of transcriptional regulation. The results could provide some preliminary ideas on how histone modification profile in pluripotent cells is founded and how it is changed during cell differentiation.

Potential for further development of the research and the proposed course of action *(maximum half a page)*

Having found the unique features of specific promoters, we could apply these features to promoter prediction and classification. Besides, further analysis is necessary to understand the mechanism why a type of promoter presents special features. For example, while studying promoter core motif enrichment, we found two subtypes of BRE motif, BREu and BREd, are enriched in protein coding gene promoter and lncRNA gene promoter respectively. Meanwhile, some promoters to both types of gene harbor both subtypes of BRE. The finding reminds us that the promoter core motif type may be different concerning the type of gene. And it is interesting if a promoter could lead to different type of transcript depending on the core motif used in transcription initiation. In addition to further analysis based on present research, we also plan to study other cis-regulatory elements, including enhancer, and compare the results. Enhancer detection is an important field and many sources are employed in prediction, such as histone modification, p300 binding, TFBS enrichment. Similar to the study on promoter, we could study enhancer from all perspectives and use the findings in this project to differentiate promoter and enhancer.

7. The Layman's Summary

(describe in layman's language the nature, significance and value of the research project, in no more than 200 words)

Gene expression is complicated and well organized in mammals. For a gene to be expressed, the DNA sequence surrounding transcription start site, called promoter, is interacted with other parts of the genome and many proteins to decide when and to what level this gene is to be expressed. Therefore, promoter is of great importance to gene expression regulation. To study gene expression regulation, it is necessary we understand the features of promoter and special characters of different promoters. In this study, we summarized and compared these features and applied them to new promoter prediction and promoter classification. With the knowledge of promoter features and how a feature leads to a specific event of gene expression, we understand the mechanism of gene expression. More importantly, it is possible we

would control gene expression or modify gene expression in patients if the disease is caused by abnormal state of a promoter.

Part C: Research Output

8. Peer-reviewed journal publication(s) arising directly from this research project
(Please attach a copy of each publication and/or the letter of acceptance if not yet submitted in the previous progress report(s). All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.))

The Latest Status of Publications				Author(s) *(**bold** the authors belonging to the project teams and denote the corresponding author with an asterisk*)*	Title and Journal/Book *(with the volume, pages and other necessary publishing details specified)*	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this report *(Yes or No)*	Acknowledged the support of this Joint Research Scheme *(Yes or No)*
Year of publication	Year of Acceptance *(For paper accepted but not yet published)*	Under Review	Under Preparation *(optional)*					
2011				Weixin Wang, Zhi Wei, Tak-Wah Lam & Junwen Wang*	Next generation sequencing has lower sequence coverage and poorer SNP-detecti on capability in the regulatory regions, Scientific Reports, 2011, 1- 55		yes	yes

2011				Jing Qin, Mulin Jun Li, Panwen Wang, Michael Q. Zhang and Junwen Wang*	ChIP-Array: combinatory analysis of ChIP-seq/chip and microarray gene expression data to discover direct/indirect targets of a transcription factor, Nucleic Acids Research, 2011, v. 39, suppl. 2, p. W430-W436		yes	yes
2011				Wang LY, Wang PW, Li MJ, Qin J, Wang XO, Zhang MQ and Wang JW*	EpiRegNet: constructing epigenetic regulatory networks from high throughput gene expression data for human.Epigenetics, 6(12):1505-12.		yes	yes
2011				Yang S, Yalamanchili HK, Li X, Yao KM, Sham PC, Zhang MQ, and Wang JW*	Correlated evolution of transcription factors and their binding sites. Bioinformatics, 27(21):2972-2978.		yes	yes

2012				Li MJ, Sham PC, and Wang JW*	Genetic variants representation, annotation and prioritization in the post-GWAS era. Cell Research, 22(10):1505-1508.		yes	yes
2012				Yalamanchili HK, Xiao QW, and Wang JW*	A Neural Response Algorithm for Protein Function Prediction. BMC Systems Biology, 6(S1):S19.		yes	yes
2012				Xu F, Wang W, Wang P, Li MJ, Sham PC, and Wang JW*	A fast and accurate SNP detection algorithm for next-generation sequencing data. Nat. Commun. 3:1258.		yes	yes

2012				Li MJ, Wang P, Liu X, Lim EL, Wang Z, Yeager M, Wong MP, Sham PC, Chanock S, and Wang JW*	GWASdb: a database for human genetic variants identified by genome wide association studies. ***Nucleic Acids Research***, 40(1):D1047-54.		Yes	Yes
2011				Zhang G, Chen X, Chan L, Zhang M, Zhu B, Wang L, Zhu X, Zhang J, Zhou B, and Wang JW*	An SNP selection strategy identified IL-22 associating with susceptibility to tuberculosis in Chinese. ***Scientific Reports***, 1:20.		Yes	Yes

9. Recognized International conference(s) in which paper(s) related to this research project was/were delivered *(Please attach a copy of each delivered paper)*

Month/Year/ Place	Title	Conference Name	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this report *(Yes or No)*	Acknowledged the support of this Joint Research Scheme *(Yes or No)*
2011	NRProF: Neural Response Based Protein Function Prediction Algorithm	IEEE International Conference on Systems Biology, 33-40.		yes	yes

10. Student(s) trained *(Please attach a copy of the title page of the thesis.)*

Name	Degree registered for	Date of registration	Date of thesis submission/ graduation
Yan Wang	PhD	1/8/2010	In progress
Jing Qin	PhD	1/8/2010	In progress

11. Other impact *(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, etc.)*

The PI, Dr. Junwen Wang was awarded outstanding young research award of HKU, 2012.