國 立 交 通 大 學

生物資訊及系統生物研究所

博士論文

探討人類可轉錄假基因的調控及功能分析

Exploring Regulations and Functions of Transcribed Pseudogenes in *Homo Sapiens*

研 究 生：詹雯玲

指導教授：黃憲達 教授

張建國 教授

中 華 民 國 一 百 零 一 年 十 一 月

探討人類可轉錄假基因的調控及功能分析

Exploring Regulations and Functions of Transcribed Pseudogenes in *Homo Sapiens*

研 究 生：詹雯玲　　Student : Wen-Ling Chan

指導教授：黃憲達 博士　　Advisor : Dr. Hsien-Da Huang

張建國 博士　　Dr. Jan-Gowth Chang

國立交通大學

生物資訊及系統生物研究所

博士論文

A Thesis

Submitted to Institute of Bioinformatics and Systems Biology

College of Biological Science and Technology

National Chiao Tung University

In partial Fulfillment of the Requirements

for the Degree of

Ph.D.

in

Bioinformatics and Systems Biology

November 2012

Hsinchu, Taiwan, Republic of China

中　華　民　國　一　百　零　一　年　十　一　月

探討人類可轉錄假基因的調控及功能分析

學生：詹雯玲　　指導教授：黃憲達 博士

張建國 博士

國立交通大學　生物資訊及系統生物研究所

摘要

假基因 (pseudogene)主要被認為是演化過程中所產生的垃圾 DNA 序列。然而，近年來研究發現假基因-特別是可轉錄的假基因，可能經由產生內生性小的干擾 RNA、反股 RNA 或抑制 RNA，以扮演調控基因的角色。在老鼠及果蠅已被証實轉錄假基因可以產生內生性小的干擾 RNA 進而調控蛋白質基因的表現，但是在人類此機制未知。因此，本研究主要目的是系統化探討人類可轉錄假基因的二個機制：產生內生性小的干擾 RNA 以及 miRNA 誘餌的功能。為了系統化的處理及更新這些結果及相關的資訊，我們也建立一個新穎整合性的資料庫-pseudoMap，做為研究人類可轉錄假基因的研究平台。由生物資訊所導出的一個假基因 *ψPPM1K*-經由 *PPM1K* 部分反轉錄所產生的轉錄假基因，因反向重覆序列建構成髮夾式結構而產生二個內生性小的干擾 RNA 進而調控許多細胞相關基因，也包含了 *NEK8*。41 對肝癌及其周邊非癌化組織研究顯示，由 *ψPPM1K* 所產生的二個內生性小的干擾 RNA 及其對應的正常蛋白質基因，在癌組織的表現均低於非癌化組織的表現，而此現象相反於所預測的 esiRNA1 的調控基因 *NEK8* 的表現。除此之外，*NEK8* 及 *PPM1K* 在過度表現 *ψPPM1K* 的載體比將 esiRNA1 刪除的載體表現還低。甚至於若將 *NEK8* 表現在已轉染 *ψPPM1K* 的載體，顯示 *NEK8* 可抵消 *ψPPM1K* 對細胞成長的抑制。就我們所知，這是第一個實驗証實人類可轉錄假基因所產生的內生性小的干擾RNA 調控肝癌細胞的表現。也同時驗証生物資訊預測的可行性。經由此研究，可以對人類可轉錄假基因的功能、調控，以及與其他蛋白質基因的關係有更完整、深入的認識。

Exploring Regulations and Functions of Transcribed Pseudogenes in *Homo Sapiens*

Student: Wen-Ling Chan Advisor : Dr. Hsien-Da Huang

Dr. Jan-Gowth Chang

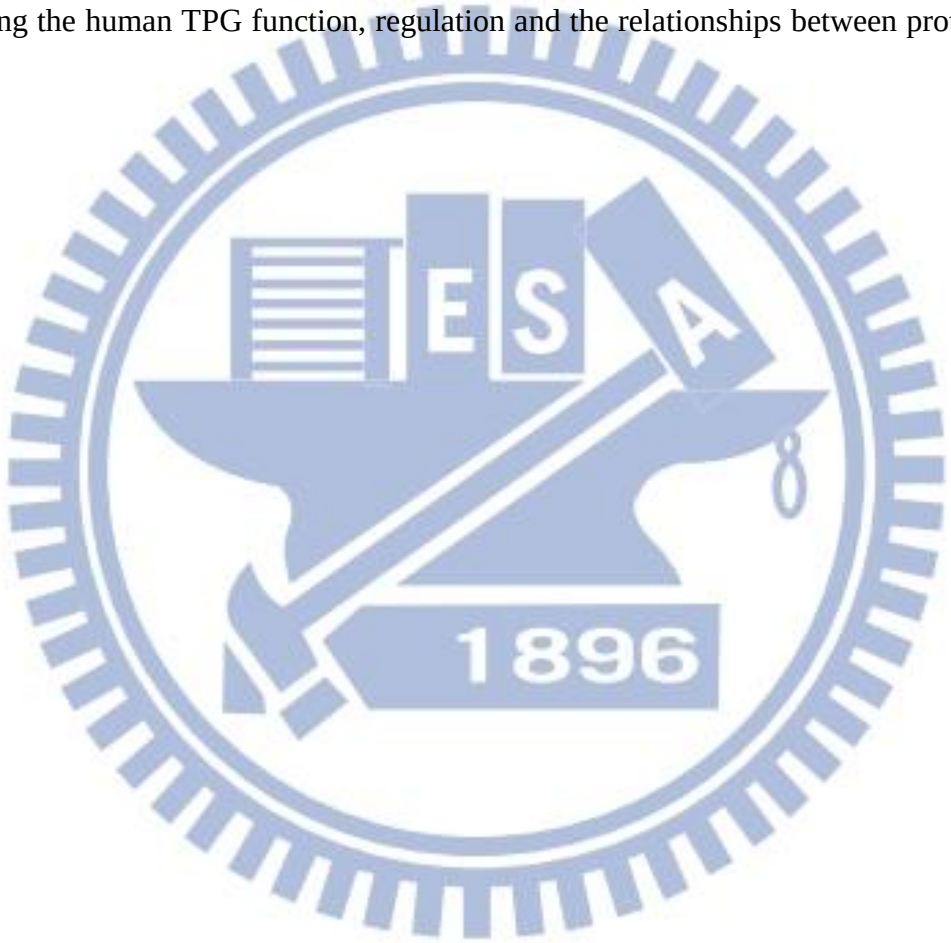
Institute of Bioinformatics and Systems Biology,

National Chiao Tung University

Abstract

Pseudogenes have been mainly considered as “junk” DNA, failed copies of genes that arise during the evolution of genomes. However, recent studies indicate that pseudogenes, especially the transcribed ones, may function as gene regulators through generation of endogenous small interfering RNAs (esiRNAs), antisense RNAs, or RNA decoys. In mice and flies, pseudogene transcripts can be processed into esiRNAs that regulate protein-coding genes, but this mechanism in human remains unknown. Therefore, the aim of this work is systematically to demonstrate two mechanisms of human transcribed pseudogenes (TPGs): encoding esiRNAs and decoying miRNAs that target the parental gene. To enable the systematic compilation and updating of these results and additional information, we have developed a database, pseudoMap, an innovative and comprehensive resource to study human TPGs. A genome-wide survey revealed a TPG *ψPPM1K*, a partial retrotranscript from *PPM1K* (protein phosphatase, Mg^{2+}/Mn^{2+} dependent, 1K), containing inverted repeats capable of folding into hairpin structures that can be processed into two esiRNAs; these esiRNAs potentially target many cellular genes, including *NEK8*. In 41 paired surgical specimens, we found significantly reduced expression of two predicted *ψPPM1K*-specific esiRNAs, and the cognate gene *PPM1K*, in hepatocellular carcinoma (HCC) compared to matched non-tumor tissues, whereas the expression of target gene *NEK8* was increased in

tumors. Additionally, *NEK8* and *PPM1K* were down-regulated in stably transfected *ψPPM1K*-overexpressing cells, but not in cells transfected with an esiRNA1-deletion mutant of *ψPPM1K*. Furthermore, expression of *NEK8* in *ψPPM1K*-transfected cells demonstrated that *NEK8* can counteract the growth inhibitory effects of *ψPPM1K*. These findings indicate that a transcribed pseudogene can exert tumor suppressor activity independent of its parental gene by generation of esiRNAs that regulate human cell growth. To our knowledge, this is the first investigation of an esiRNA-mediated role of human pseudogenes in HCC as well as verification of computational prediction. This study provides further information for elucidating the human TPG function, regulation and the relationships between protein-coding genes.



誌謝

求學生涯終於畫下休止符！

一開始的雄心萬丈，在博二即取得博士候選人資格，歷經撞牆期，研究停滯不前，到最後柳暗花明又一村，終於開花結果。

很幸運的一路上，獲得許多貴人的幫忙與扶持。特別感謝中國醫藥大學的張建國教授，啟發我對生物研究的興趣進而激發對研究的熱忱，可謂是我的啟蒙老師！進入博士生涯，有幸獲得黃憲達教授的指導，其對資訊及生物並重的態度及活躍的思考，讓我得以整合二個不同的領域，也體恤我台中、新竹二地奔波。感謝中國醫藥大學的楊文光教授及已故的楊師母，帶領我進入 cell culture 這個聖殿，他們二位將細胞治療應用於治療癌症，不但對病患提供一道曙光且有顯著的成效。亦感謝高雄醫學大學的鐘育志教授及交通大學的廖光文教授撥冗指導我的博士論文，提供許多寶貴的建議，俾使本論文更加充實完整。此外，感謝宗夷、熙淵、博凱、勝達、豐茂及交大、中國實驗室的室友們，對研究及生活上的互相扶持，是我的良師益友！

今日，若是我有些許的成就，歸功於家人強力的支援！感謝我的公婆、大姑子及其他家人體恤我長久的求學生涯，在必要時予以協助，讓我無後顧之憂。最後，幕後大功臣－我的老公，給予最大的包容與支持，以及二個小孩在我疲憊不堪時，適時幫媽媽加油打氣。在此致上深深的謝意，您們是我強而有力的後盾！

僅以此論文，獻給來不及看到我畢業的母親！

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

The human genome comprises more numbers of pseudogenes than corresponding functional genes [1]**.** Generally, pseudogenes are defined as nonfunctional copies of gene fragments incorporated into genome by either gene duplication of genomic DNA or retrotransposition of mRNA [2]. In apparent contradiction of the assumption that pseudogenes are genomic fossils, genome-wide investigations have recently provided evidences for actively transcribed pseudogenes (TPGs) with potential functional implications [3-10]. It implied that pseudogenes, especially those that are transcribed, may not be mere genomic fossils, but their biological significance remains unclear. This dissertation is concerning that what are the mechanisms involved in the human TPGs. To address this issue, we focus on two topics of that the relationships between TPG and its cognate gene with miRNA decoyed mechanisms and TPG-derived endogenous small interfering RNAs (esiRNAs)-target interactions. Additionally, to enable the systematic compilation and updating of these results and additional information, a database, pseudoMap, capturing various types of information, including sequence data, TPG and cognate annotation, deep sequencing data, RNA-folding structure, gene expression profiles, miRNA annotation and target prediction, will be constructed to study the human TPGs. Hepatocellular carcinoma (HCC) is one of the most common human cancers worldwide, particularly in Asia and Africa [11]. Therefore, we are interested in exploring the relationship between TPGs and HCC. Finally, consideration of prediction results and the functions of TPG, its cognate and target gene, a TPG-*ψPPM1K*, a partial retrotranscript from *PPM1K* (protein phosphatase, Mg^{2+}/Mn^{2+} dependent, 1K), is experimented in HCC to verify the *in silico* results.

1.1 Biological background

1.1.1 Central dogma

The biological central dogma (Figure 1.1) describes the flow of genetic information within a biological system. It was first stated by Francis Crick [12]. In briefly, four steps in this dogma: First RNA polymerase docks to the chromosome and slides along the gene, transcribing the sequence on one strand of DNA into a single strand of RNA. Next, all introns-noncoding parts of the initial RNA transcript-are spliced out, and the rests are joined together to make a messenger RNA. The RNA then moves out of the nucleus to the cytosol of the cell, where molecular machines translate it into chains of amino acids. Finally, each chain twists and folds into an intricate three-dimensional shape. Traditionally, the proteins are recognized as the main responsibility for biological function.

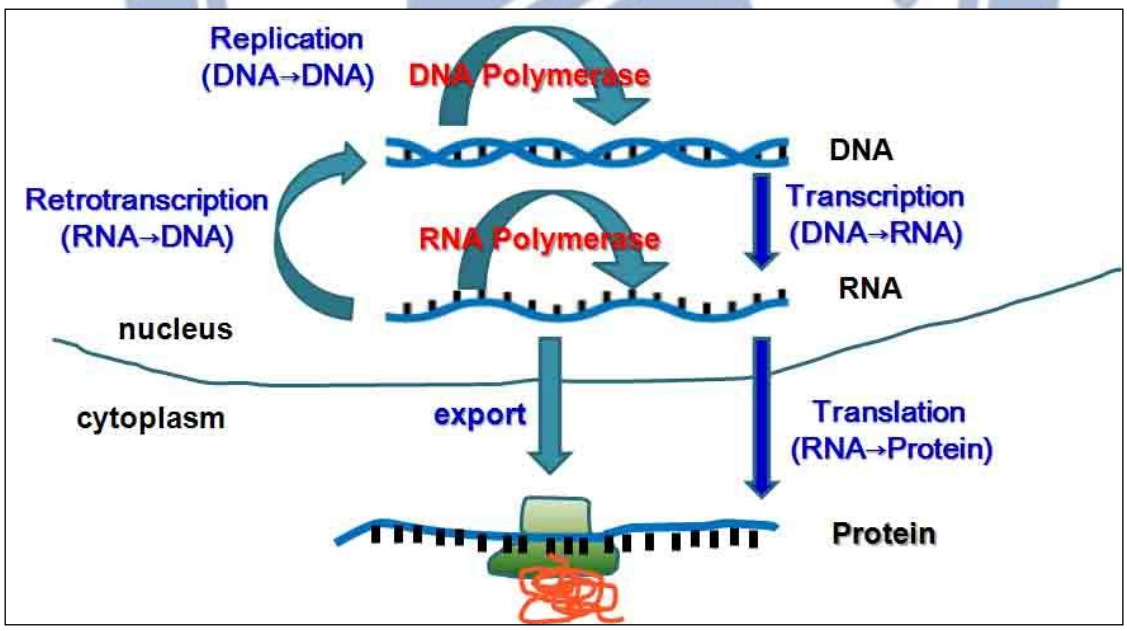


Figure 1.1 The central dogma of molecular biology.

1.1.2 Non-coding RNA

Classically, proteins are recognized as having the main responsibility for biological function, with RNA merely a messenger that transfers protein-coding information from DNA [13, 14]. This concept has changed in recent years, however, while less than 2% of the genome encodes protein, over 80% of the genome produces non-protein coding RNA transcripts (**Figure 1.2**) [13-17] and these ncRNAs have important biological functions including gene regulation [18, 19], imprinting [20-24], epigenetic regulation [25, 26], cell cycle control [27], regulation of transcription, translation and splicing [19, 28-32] and others. There are many studies discoveries of non-coding RNAs (ncRNAs) to regulate protein-coding gene expressions. ncRNA is any RNA molecule that is not translated into a protein, such as piwi interacting RNA (piRNA), microRNAs (miRNAs), short interfering RNAs (siRNA), long ncRNAs (lncRNAs) and transcribed pseudogenes (TPGs) (**Figure 1.3**). Following, we will introduce the functions of these ncRNAs.

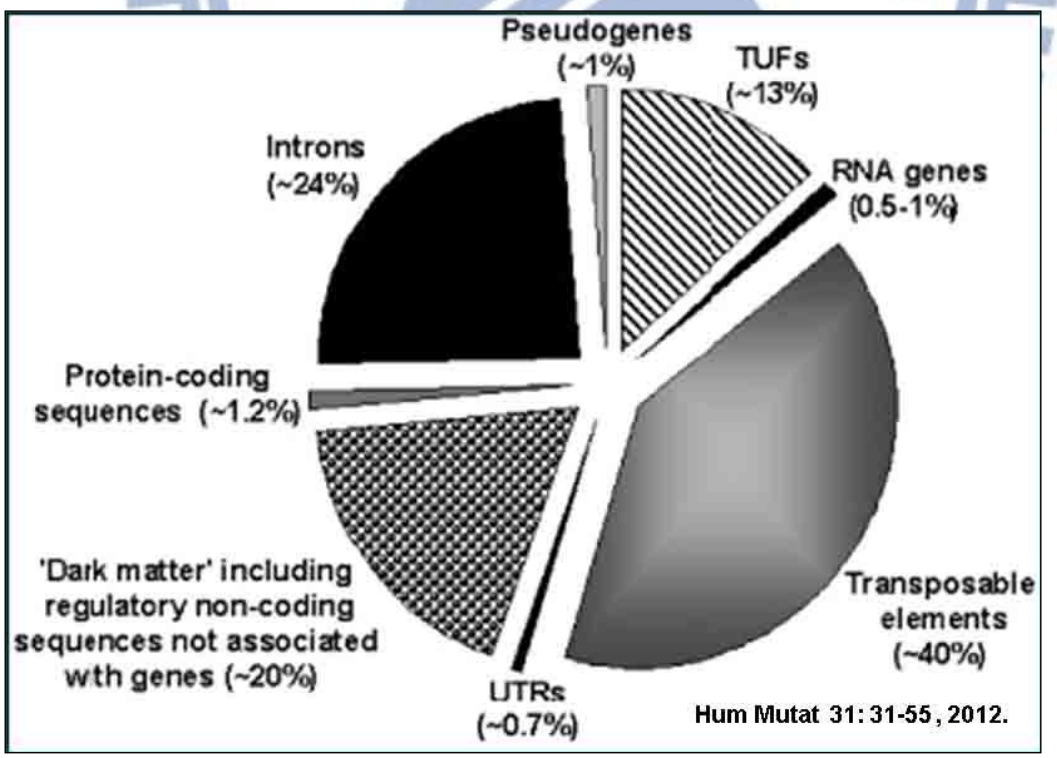


Figure 1.2 Human genome.

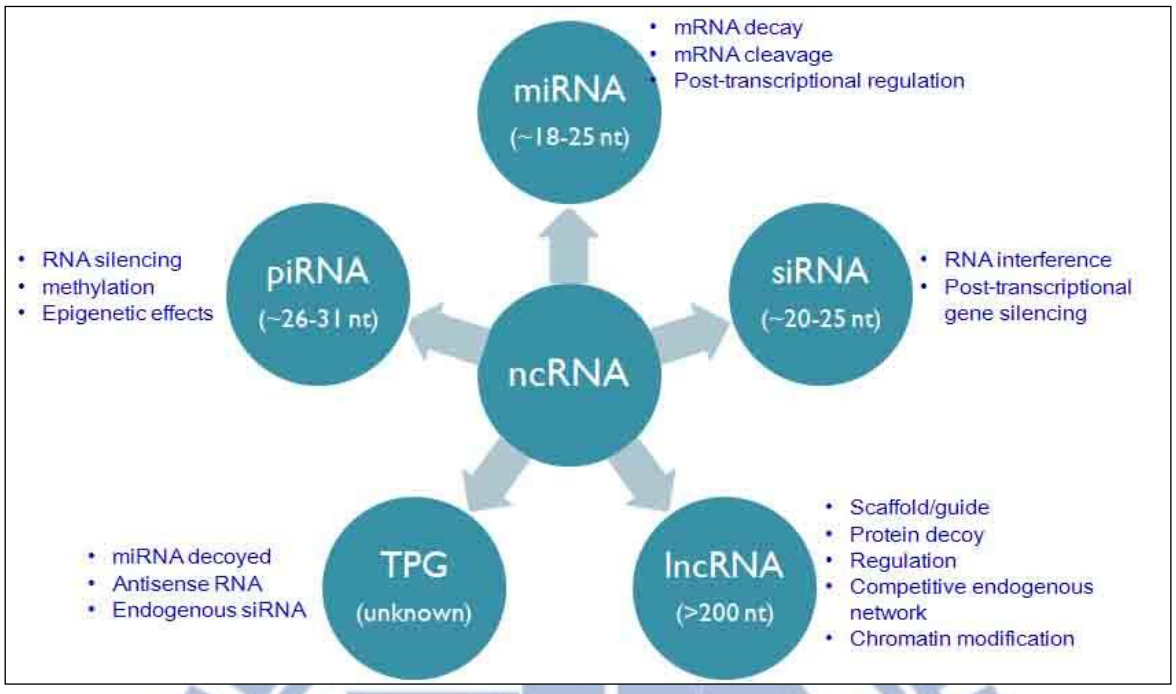


Figure 1.3 Schematic representation of the emerging ncRNA world.

1.1.3 Piwi interacting RNA (piRNA)

Piwi interacting RNA (piRNA), form RNA-protein complexes through interactions with piwi proteins, is the largest class of sRNA molecules that is expressed in animal cells [33]. The biogenesis of piRNAs is not yet fully understand, although possible mechanisms have been linked to both epigenetic and post-transcriptional gene silencing of retrotransposons and other genetic elements in germ line cells, particularly those in spermatogenesis [34].

1.1.4 MicroRNA (miRNA)

MicroRNAs (miRNAs) play important roles on development, oncogenesis and apoptosis by binding to mRNAs to regulate the post-transcriptional level of gene expression in mammals, plants and insects [35, 36]. The general biogenesis of the miRNA is shown in **Figure 1.4**. In briefly, microRNA is defined as single-stranded RNAs of ~22 nt in length generated from endogenous transcripts. It is transcribed by RNA polymerase II [8]、, and the primary miRNA

(pri-miRNA) is first processed by the nuclear RNase type III enzyme, Drosha, to release the hairpin-shaped intermediates, become precursor miRNA (pre-miRNAs) [37]. Pre-miRNA is typically 60-70 nt, is a hairpin structure, which contain an ~22 bp double-stranded stem and a ~10 nt terminal loop. The nuclear export factor, Exportin 5, export the pre-miRNA from the nucleus to the cytoplasm [38]. Then pre-miRNA is cleaved by another RNase III type enzyme, Dicer, to generate an ~22 nt RNA duplex that includes the mature miRNA which becomes part of the RNA-induced silencing complex (RISC) [39]. The mature miRNA then binds to complementary sites in the mRNA target to negatively regulate gene expression through two major mechanisms: one is mRNA degradation through perfect hybridization between miRNA and its target sites, another is translation repression with imperfect hybridization.

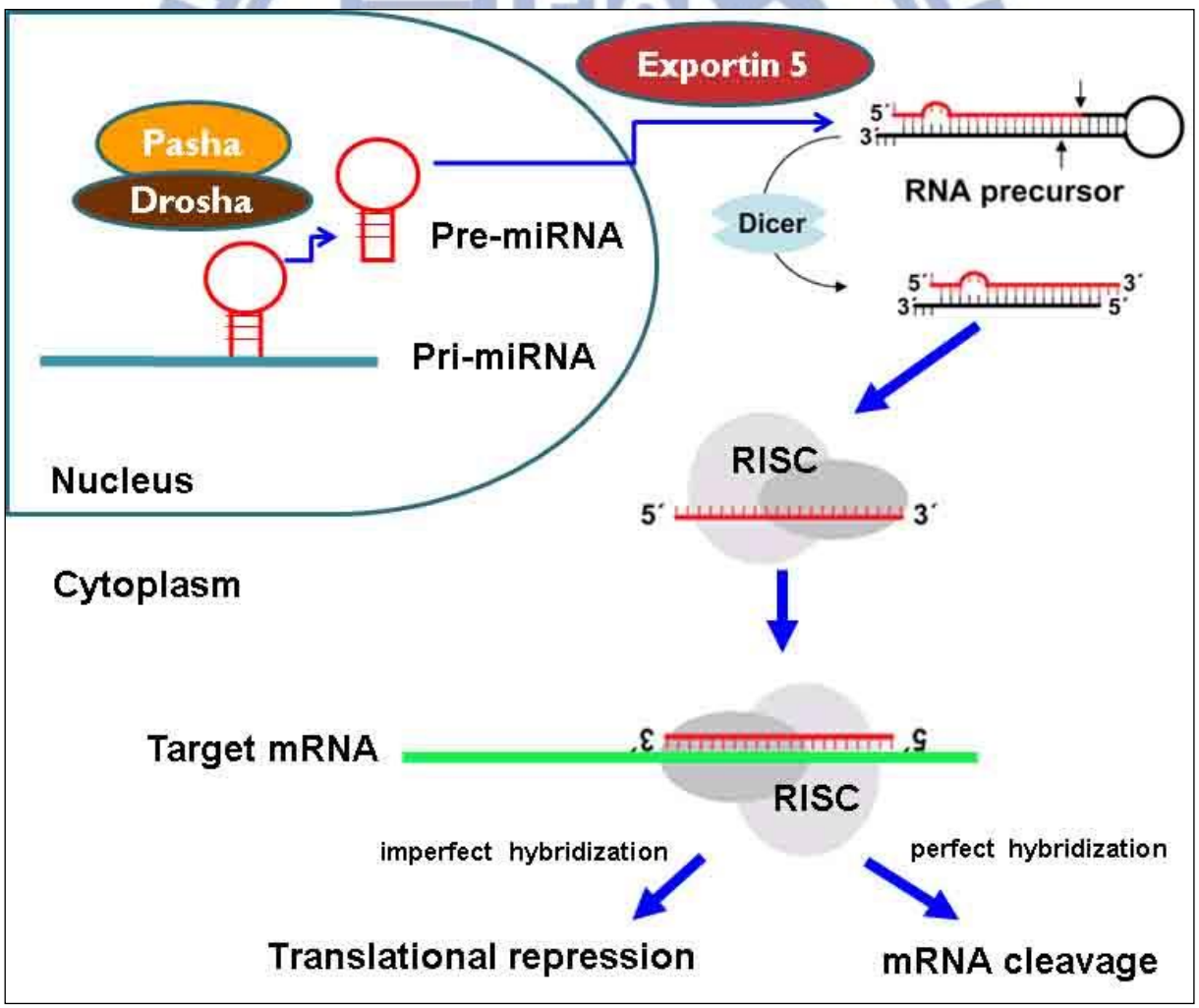


Figure 1.4 Micro-RNA biogenesis.

1.1.5 Small interfering RNA (siRNA)

Small interfering RNA (siRNA) is a class of double-stranded RNA molecules with 20-25 bp in length. siRNA plays many roles, but its most notable is in the RNA interference (RNAi) pathway, where it interference with the expression of specific genes with complementary nucleotide sequence [40].

1.1.6 Long non-coding RNA (lncRNA)

Long non-coding RNA (lncRNA) is in general considered as non-protein coding transcript with more than 200 bp in length. This limitation is due to practical considerations including the separation of RNAs in common experimental protocols. Large scale sequencing of cDNA libraries and more recently transcriptomic sequencing by next generation sequencing indicate that the number of lncRNAs is over than ten thousand in human genome [41]. The functions of lncRNA are showed in **Figure 1.5** [42]. In briefly, lncRNA transcribed from an upstream non-coding promoter can negative (**Figure 1.5** ①) or positively (**Figure 1.5** ②) affect expression of the downstream gene by inhibiting RNA polymerase II recruitment and/or inducing chromatin remodelling, respectively. lncRNA is able to hybridize to the pre-mRNA and block recognition of the splice sites by the spliceosome, thus resulting in an alternatively spliced transcript (**Figure 1.5** ③). Alternatively, hybridization of sense and antisense transcripts can allow Dicer to generate endo-siRNAs (**Figure 1.5** ④). The binding of lncRNA to miRNA results in the miRNA silencing (**Figure 1.5** ⑤). The complex of lncRNA and specific protein partners can modulate the protein activity (**Figure 1.5** ⑥), structure (**Figure 1.5** ⑦), localization (**Figure 1.5** ⑧) or epigenetic regulation (**Figure 1.5** ⑨). Finally, lncRNA also can produce sRNAs (**Figure 1.5** ⑩).

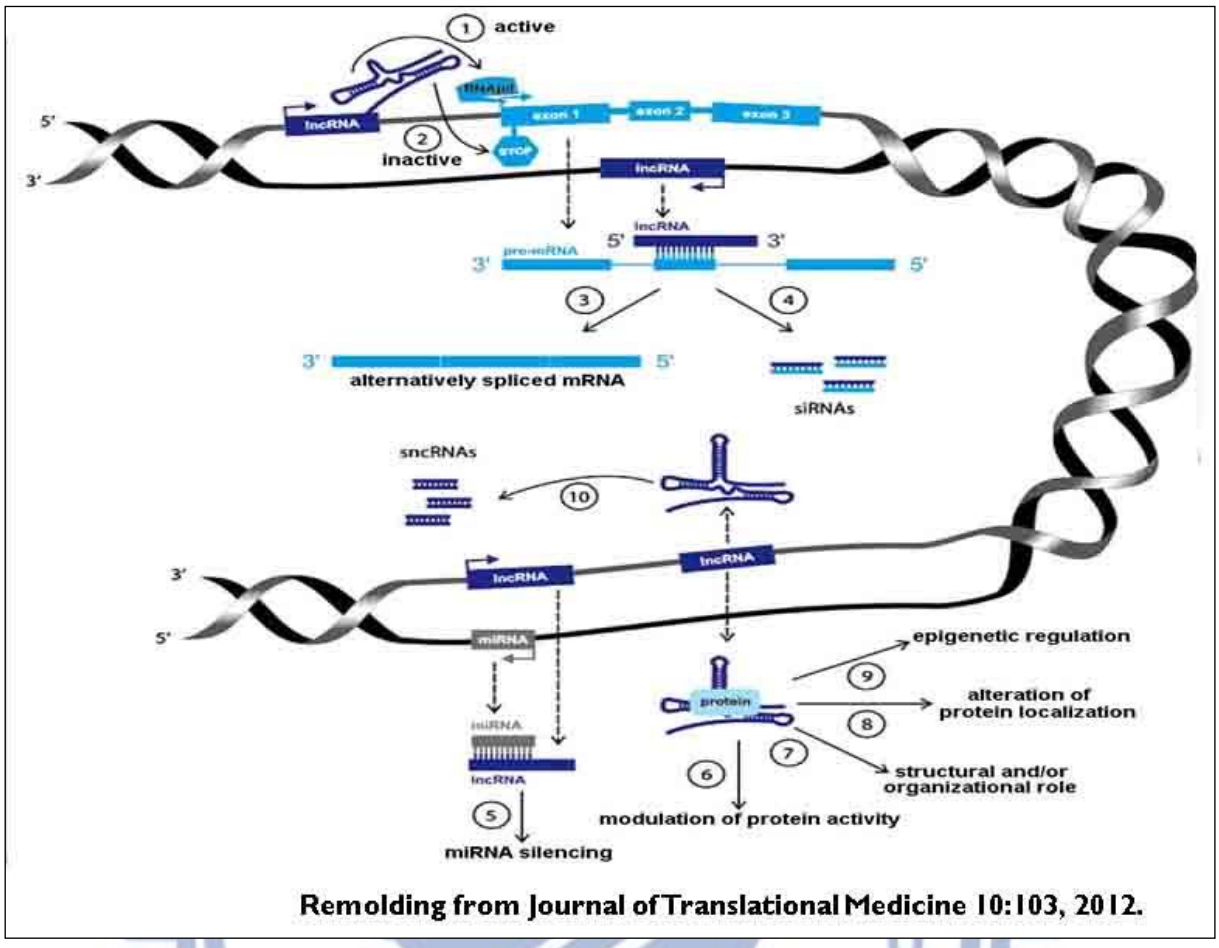


Figure 1.5 The functions of lncRNA.

1.1.7 Pseudogene

Pseudogenes are DNA sequences in the genome that similarity to specific protein-coding genes, but are unable to produce functional proteins due to existence of frameshifts, premature stop codons or other deleterious mutations [2]. Pseudogenes have been denoted in several ways including the prefixed Greek symbol ψ, for example *ψPPM1K*, or by a capital '*P*' suffix, for example *ZNF355P*. There are two major classes of pseudogene (**Figure 1.6**): one represents processed forms that contain poly-A tails, lack introns and arise through retrotransposition, while the other comprises nonprocessed pseudogenes resulting from gene duplication, which retain exon/intron structure, although occasionally incompletely [2]. Pseudogenes are usually considered to be junk DNA and genomic fossils, however, a number

of recently studies showed that pseudogenes, especially transcribed ones, may not mere genomic fossils, but function as gene regulators. The following section will introduce the functions of TPGs..

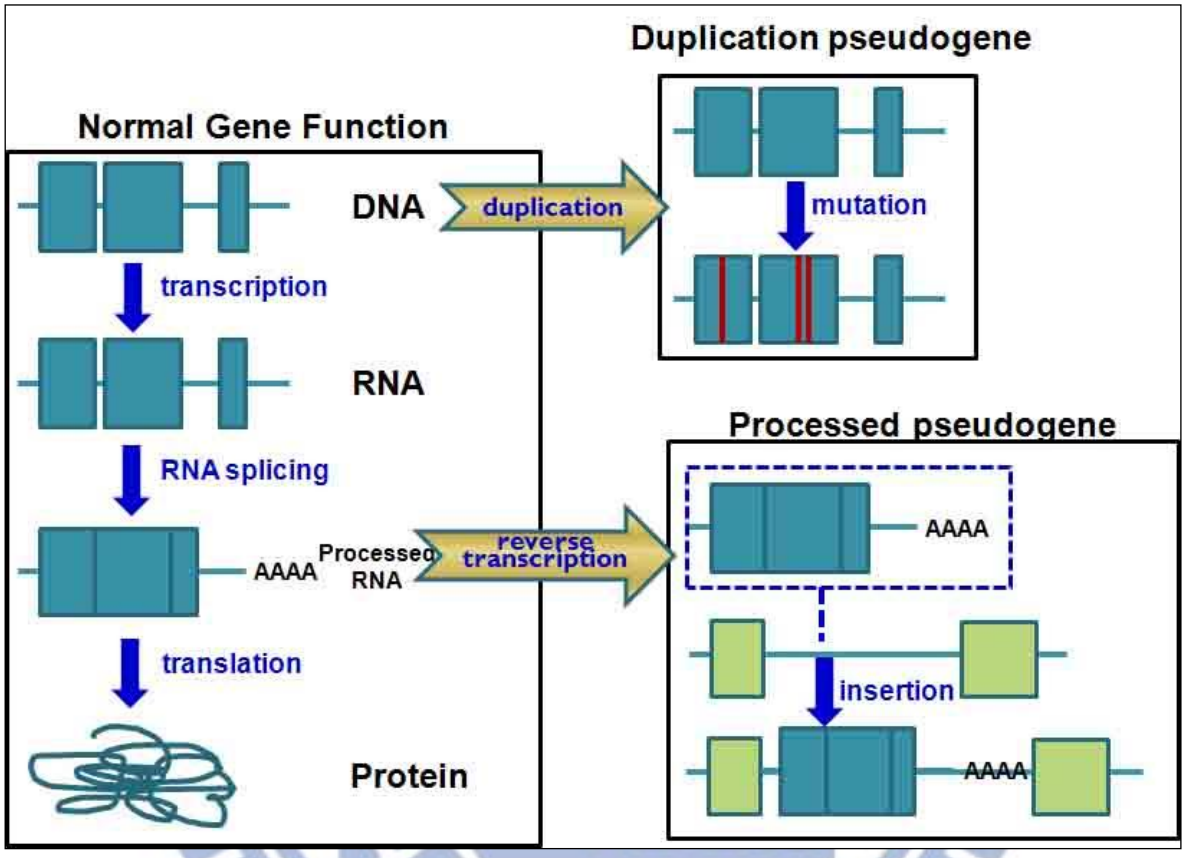


Figure 1.6 The mechanisms of pseudogene.

1.1.8 Transcribed pseudogene (TPG)

The transcribed pseudogenes (TPGs) are disabled but nonetheless transcribed. These TPGs may function as gene regulators through generation of endogenous siRNAs (esiRNAs), antisense RNAs, or RNA decoys. For instance, the ψNOS transcript acts as a natural antisense regulator of neuronal NOS protein synthesis in snails [44, 45]; and in mice, reduced expression of *ψmakorin1-p1* due to a transgene insertion caused mRNA instability of its parental gene *Mkrn1*, resulting in polycystic kidneys and bone deformity [10, 46], although

contradictory results were also reported [47]. Additionally, a transcript of *ψPTEN/PTENP1*, a highly homologous processed TPG of tumor suppressor gene *PTEN*, not only interacts with its cognate sequence but also exerts a growth suppressor role as a decoy by binding to *PTEN*-targeting miRNAs [48]. These findings clearly imply that TPGs may play active regulatory roles in cellular functions.

1.2 Related works

RNA interference (RNAi) is a natural cellular process that defends cells against viruses and transposons, and also regulates gene expression in a sequence-specific manner [40]. Three RNAi pathways can be distinguished on the basis of the biogenesis and functional roles of the classes of small RNA involved, two of which are siRNA, resulting from processing of dsRNA, and miRNA, which derive from shRNA, respectively [39, 49]. The third category is piRNA: these are ssRNA sequences that interact with piwi protein and seem to be involved in transcriptional gene silencing of retrotransposons and other genetic elements in germ line cells [50]. By binding to mRNAs and thereby repressing protein synthesis, miRNAs may regulate cellular development, oncogenesis and apoptosis [35, 36]. Previously studies showed that pseudogenes may produce esiRNAs to regulate protein-coding genes through RNAi mechanism (**Figure 1.7**). In mice and fruit flies, dsRNAs arising from the antisense/sense transcripts of processed TPGs and their cognates (showed in **Figure 1.7 b**), or hairpin structures resulting from inversion and duplication (showed in **Figure 1.7 a**), are cut by Dicer into 21 nt siRNA that can bind RNA-induced silencing complex (RISC) and regulate the expression of the parental gene [49, 51-55]. Such regulatory mechanism in human remains unclear.

Another study indicated pseudogene may act as a miRNA-decoyed (**Figure 1.8**). A transcript of *ψPTEN/PTENP1*, a highly homologous processed TPG of tumor suppressor gene

PTEN, not only interacts with its cognate sequence but also exerts a growth suppressor role as a decoy by binding to *PTEN*-targeting miRNAs [48].

In present, most of computational studies focus on evolution of pseudogenes. A number of public database and tools of studying pseudogenes are described as follows.

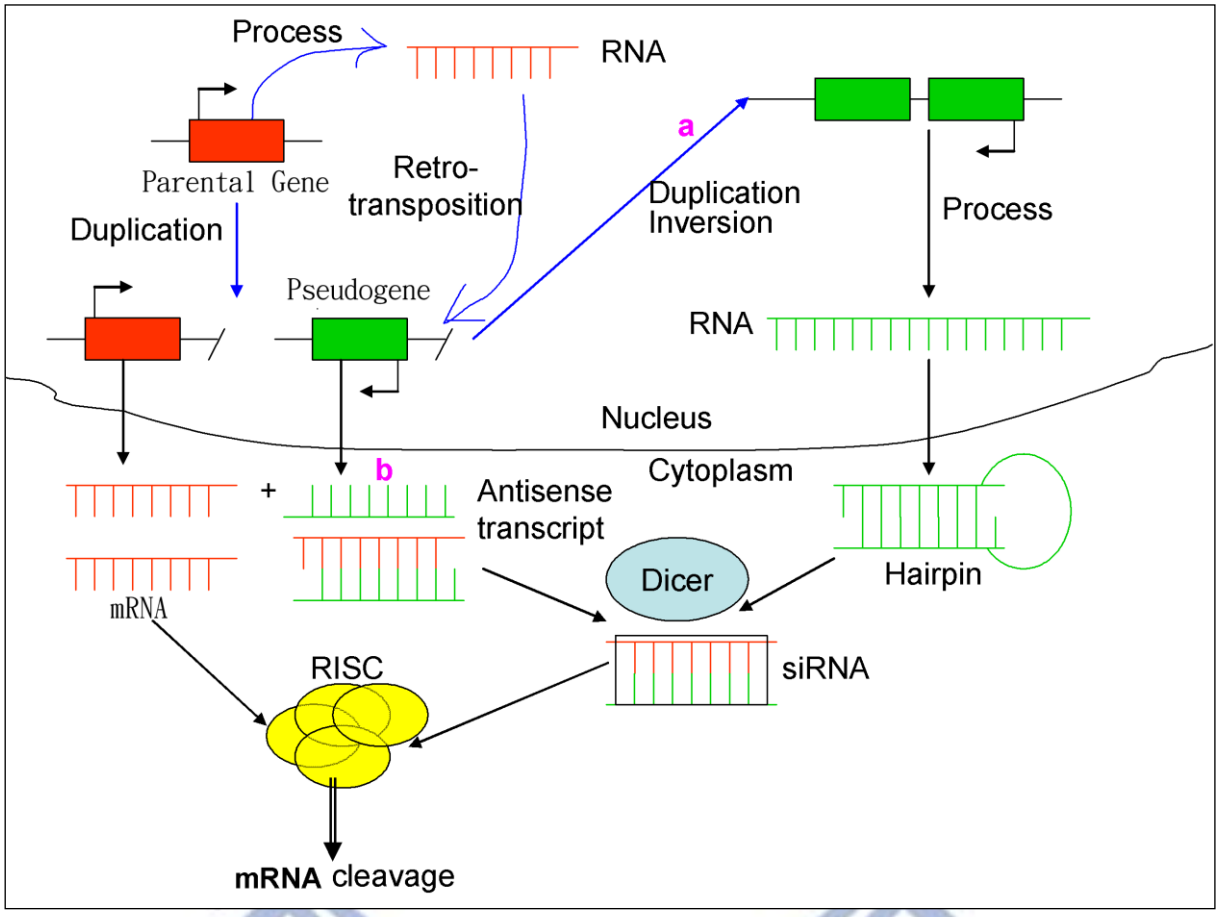


Figure 1.7 The mechanism of pseudogene-mediated production esiRNAs.

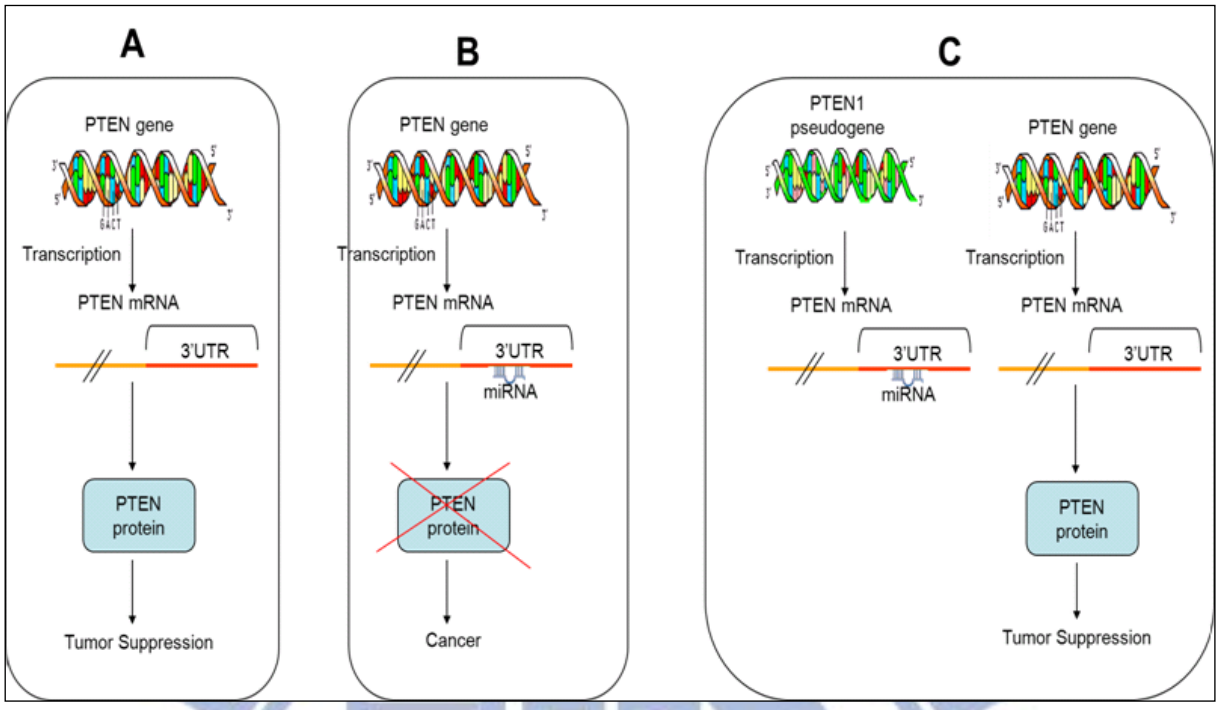


Figure 1.8 TPG-*PTENP1* may act as a miRNA decoyed.

1.2.1 Public biological databases

In present, the computational analyses focus on the evolution of pseudogenes. There is no resource indeed providing the functional information in pseudogenes. Following sections will describe the pseudogene-related database.

1.2.1.1 Pseudogene.org

Pseudogene.org [56] is developed and maintained by Yale Gerstein Group. This site contains a comprehensive database of identified pseudogenes, utilities used to find pseudogenes, various publication data sets and a pseudogene knowledgebase.

1.2.1.2 Hoppsigen DB

Hoppsigen [57] is a nucleic database of homologous processed pseudogenes. The database is developped at the PBIL (Pôle Bioinformatique Lyonnais). The authors have identified 5,823 human retroelements and 3,934 mouse retroelements. These retroelements were annotated and

stored in the database HOPPSIGEN (Homologous processed pseudogenes). Sequences were grouped in families considering their homologies. The database contains 3,168 families of exclusively human (1,966) or mouse retroelements (1,202) and 323 families containing human and mouse retroelements. 5,206 human retroelements were annotated as processed pseudogenes. The database contains functional genes from ENSEMBL homologous to Hoppsigen retroelements.

1.2.1.3 UI Pseudogenes

UI Pseudogenes website [58] serves as a repository for all pseudogenes in the human genome. They also provide a ranked list of human pseudogenes that have been identified as candidates for gene conversion.

1.2.1.4 Human Pseudogenes

Torrents, et al. [1] have published “A genome-wide survey of human pseudognes” in Genome Research. In this paper, the authors screened all intergenic regions in the human genome to identify pseudogenes with a combination of homology searches and a functionally test using the ratio of silent to replacement nucleotide substitutions. Finally, they detected 19,537 pseudogenes which include 17,759 processed pseudogenes and 1,778 non-processed pseudogenes.

1.2.1.5 miRBase

The miRBase database [59] aims to provide integrated interfaces to comprehensive microRNA sequence data, annotation and predicted gene targets. miRBase takes over functionality from the microRNA Registry and fulfils three main roles: the miRBase Registry acts as an independent arbiter of microRNA gene nomenclature, assigning names prior to

publication of novel miRNA sequences.

1.2.1.6 Ensembl

Ensembl [60] is a joint project between European Molecular Biology Laboratory (EMBL) and the Wellcome Trust Sanger Institute (WTSI) to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes..

1.2.1.7 UCSC Genome Browser Database

UCSC Genome Browser Database [61] contains the reference sequence and working draft assemblies for a large collection of genomes. This database is optimized to support fast interactive performance with web tools that provide powerful visualization and querying capabilities for mining the data. The Genome Browser displays a wide variety of annotations at all scales from single nucleotide level up to a full chromosome. The Table Browser provides direct access to the database tables and sequence data, enabling complex queries on genome-wide datasets. The Proteome Browser graphically displays protein properties. The Gene Sorter allows filtering and comparison of genes by several metrics including expression data and several gene properties. BLAT and In Silico PCR search for sequences in entire genomes in seconds.

1.2.1.8 fRNAdb

The Functional RNA Database (fRNAdb) [62], which hosts a large collection of known/predicted non-coding RNA sequences from public databases: H-invDB v5.0 [10], FANTOM3 [63], miRBase 17.0 [64], NONCODE v1.0 [65], Rfam v8.1 [66], RNAdb v2.0 [67] and snoRNA-LBME-db rel. 3 [68].

1.2.1.9 Gene Expression Omnibus (GEO)

The Gene Expression Omnibus (GEO) [69] at the National Center for Biotechnology Information (NCBI) is a gene expression/molecular abundance repository supporting MIAME [70] compliant data submissions, and an online resource for gene expression data browsing, query and retrieval.

1.2.2 Public analysis tools

1.2.2.1 BLAST

The basic local alignment search tool (BLAST) [71] finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

1.2.2.2 ClustalW

ClustalW [72] is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen.

1.2.2.3 RNAhybrid

RNAhybrid [73] is a variation of the classic RNA secondary structure prediction. It determines the most favorable hybridization site between two sequences. RNAhybrid does not use any RNA folding or pairwise sequence alignment code, but implements an algorithm that was specifically designed for RNA hybridization.

1.2.2.4 TargetScan

Lewis et al. [74], predict regulatory targets of mammalian microRNAs (miRNAs) by identifying mRNAs with conserved complementarity to the seed (nucleotides 2-8) of the miRNA. They developed an algorithm called TargetScan, which combines minimum free energy of miRNA/mRNA duplex with comparative sequence analysis to predict miRAN targets conserved across multiple genomes.

1.2.2.5 miranda

miRanda [75] is a open-source software for miRNA target prediction. It was used to scan all available miRNA sequences for a given genome against 3'UTR sequences of that genome derived from the Ensembl database and –tabulated separately-against all cDNA sequences and coding regions. This algorithm uses dynamic programming to search for maximal local complementarity alignments, corresponding to a double-stranded antiparallel duplex.

1.2.2.6 Mfold

Mfold [76] is a tool for predicting the secondary structure of RNA and DNA by free energy minimization. The 'm' simply refers to 'multiple'. A dynamic programming algorithm is used to predict a minimum free energy as well as minimum free energies for folding that must contain any particular base pair. Base-pairs within this free energy increment are chosen either automatically or else by the user. They also provided a web system [77] for prediction of the secondary structure of single stranded nucleic acids.

1.3 Motivation

In the last ten years, a variety of complex mechanisms such as gene silencing, gene transcription, DNA imprinting, DNA demethylation, chromatin structure dynamtics, RNA interference and others have already been connected to ncRNAs [78]. The TPGs are as like as lncRNAs, furthermore, some of them regulate the protein-coding genes. The previously study indicated TPGs are significantly over-represented near both the 5' and 3' ends of genes; this suggests that TPGs can be formed through gene-promoter co-option, or intrusion into untranslated regions [5]. However, roughly half of the TPGs are located away from genes in the intergenic DNA [5]. In accordance with the study of ψNOS transcript acts as a natural antisense regulator of neuronal NOS protein synthesis in snails [44, 45], in 2005, we have a hypothesis that human TPG may produce esiRNAs to regulate protein-coding genes (**Figure 1.9**). However, this project is broke down since we can't predict the exactly esiRNA derived from TPG. In 2008, the Next Generation Sequencing (NGS) studies showed that pseudogene-derived esiRNAs regulate transcripts in mice oocytes and drosophila somatic cells [49, 51-55]. These evidences supported our previously hypothesis. Recently years, the productions of NGS data from various sRNA libraries are quickly generated. This information supported us to predict the TPG-derived esiRNAs. Additionally, a transcript of *ψPTEN/PTENP1*, a highly homologous processed TPG of tumor suppressor gene *PTEN*, not only interacts with its cognate sequence but also exerts a growth suppressor role as a decoy by binding to *PTEN*-targeting miRNAs [48]. It implied that miRNA can co-regulate TPG and its parental gene, moreover, TPG may act as a miRNA-decoyed. Therefore, we also want to know how many miRNAs co-regulate TPG and its cognate gene.

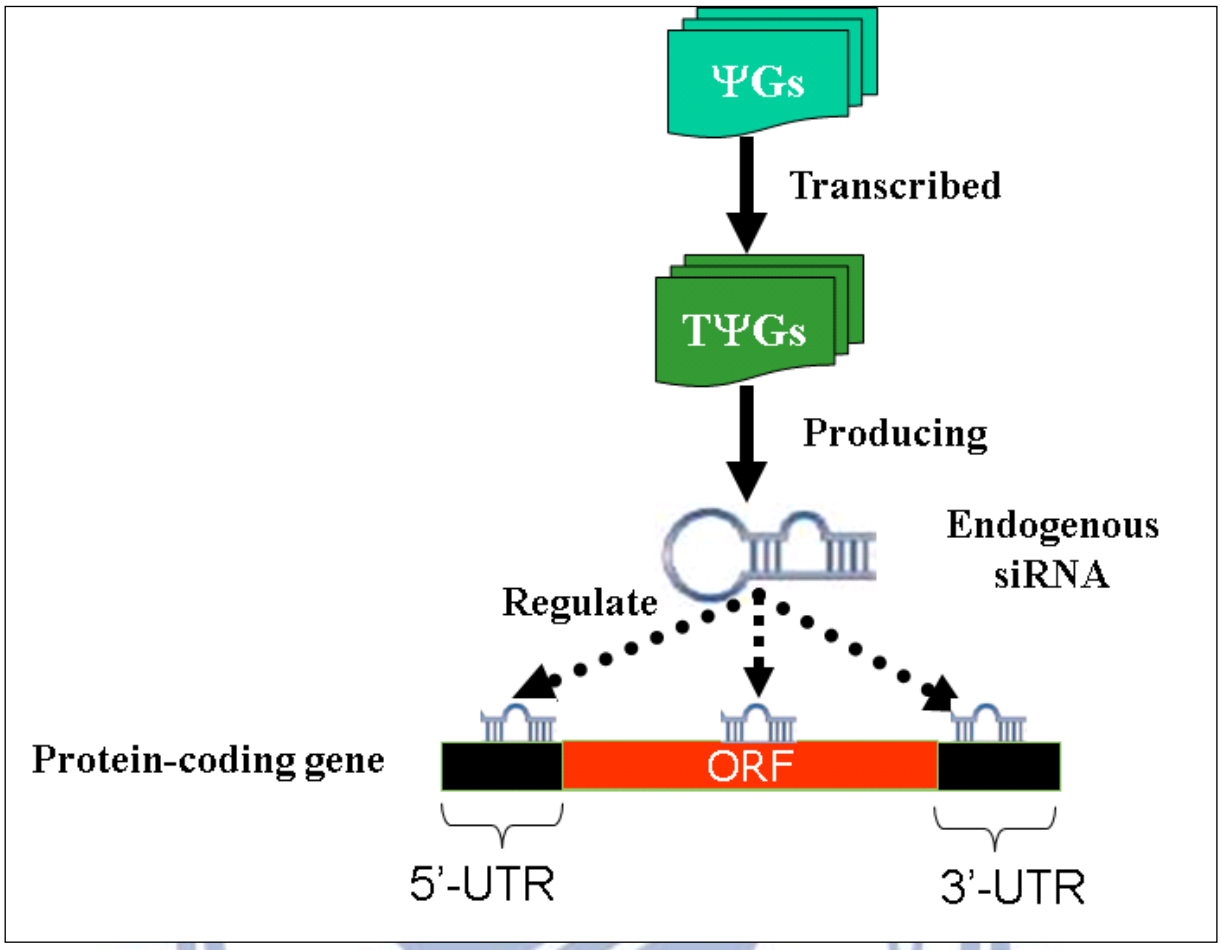


Figure 1.9 Our hypothesis of human TPGs production of esiRNAs to regulate protein-coding genes.

1.4 Research goals

The purpose of this project is systematically identifying the regulations and functions of TPGs in *Homo Sapiens*. **Figure 1.10** showed the system flow of this project. Four major works in this project, first, we want to prove our hypothesis that human TPGs may produce esiRNA/miRNA to regulate protein-coding genes. To address this issue, we develop a computational pipeline that including mapped the TPGs to small RNAs (sRNAs) that were supported by publicly deep sequencing data from various sRNA libraries, and constructed the TPG-derived esiRNA-target interactions. Second, we detect regulation of TPGs, specifically focus on miRNA co-regulated TPG and its cognate gene. Third, to enable the systematic

compilation and updating of these results and additional information, we develop a database, pseudoMap, capturing various types of information, including sequence data, TPG and cognate annotation, deep sequencing data, RNA-folding structure, gene expression profiles, miRNA annotation and target prediction. Finally, we selected *ψPPM1K*, a partial retrotranscript from *PPM1K* (protein phosphatase, Mg^{2+}/Mn^{2+} dependent, 1K), to verify our *in silico* results.

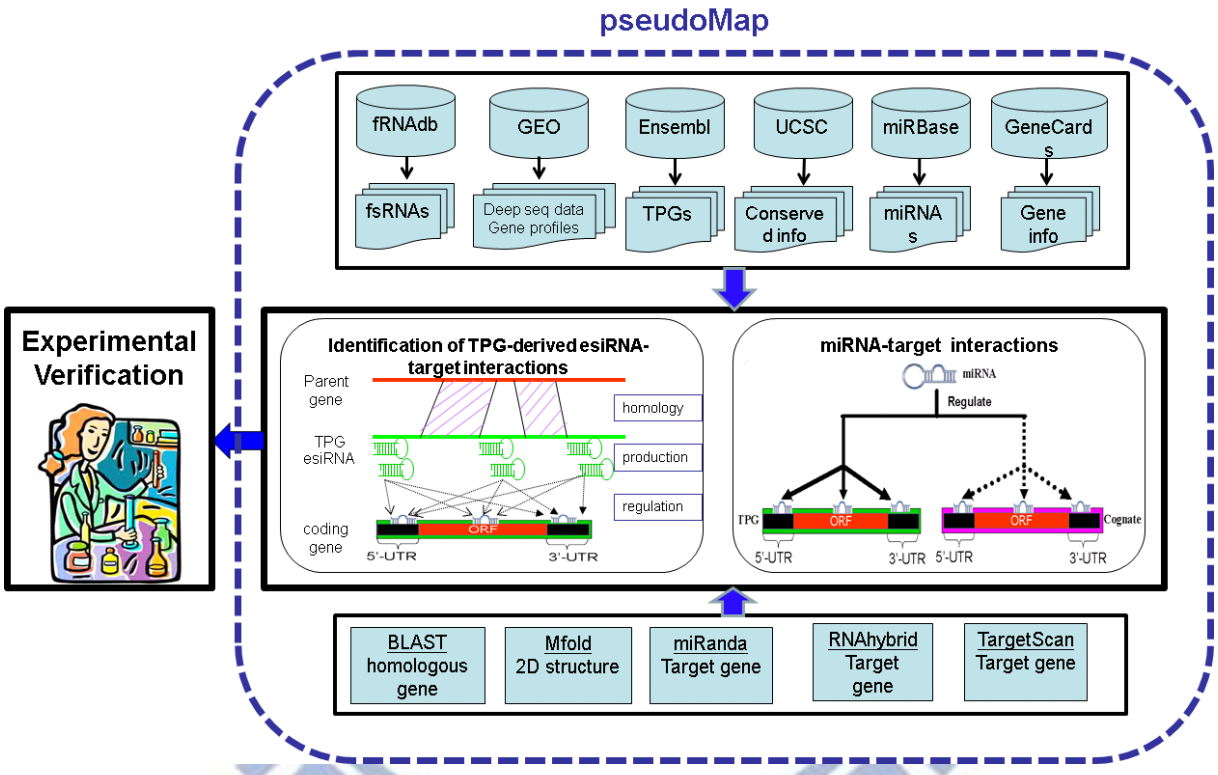
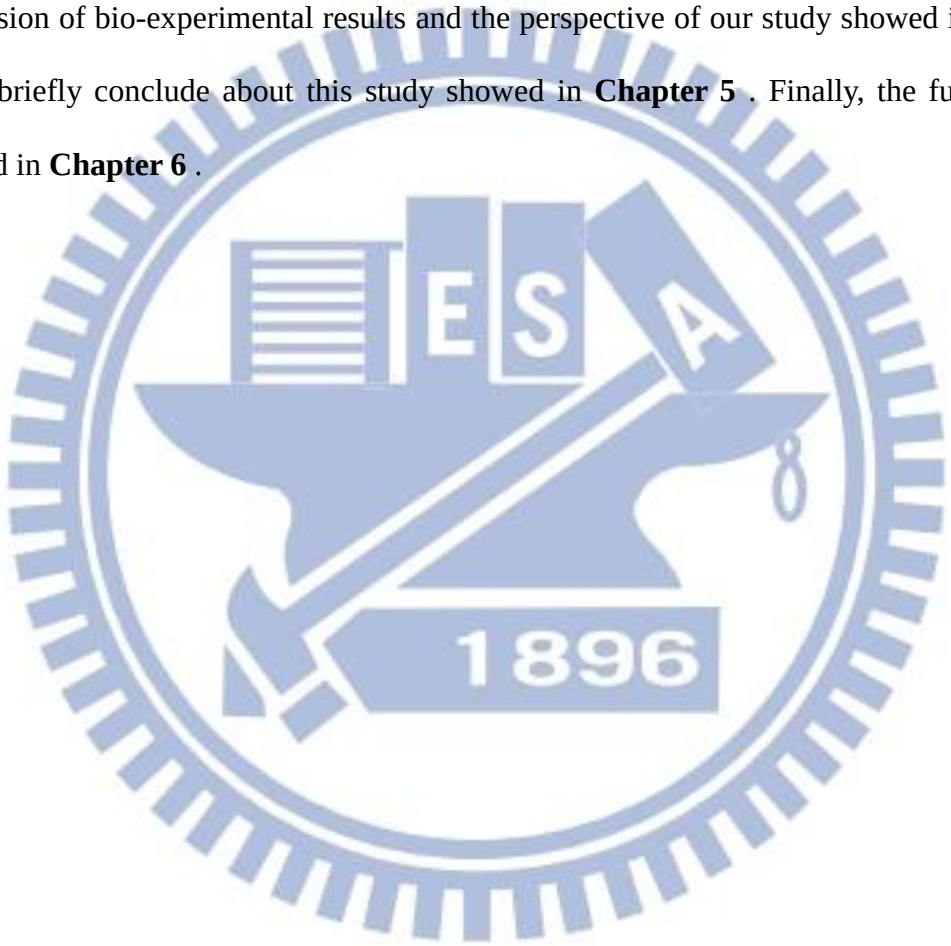


Figure 1.10 The system flow of this project.

1.5 Dissertation Organization

In this dissertation, we will analyze the regulations and functions of TPGs in *Homo Sapiens*. For these objectives to be achieved, the article is structured as follows. The first we describes some background information on the ongoing research within which the present study was carried out and a statement of the specific research questions (**Chapter 1**). Following, a computational pipeline and subsequent experimental tests were constructed for the

identification of TPG-derived esiRNAs-target interactions and miRNA-mediated mechanism of TPG and its cognate gene (**Chapter 2**). To enable the systematic compilation and updating of these results and additional information, we developed a human TPGs resource, pseudoMap (**Chapter 3**), In addition, TPG-*ψPPM1K*, a partial retrotranscript from *PPM1K* (protein phosphatase, Mg^{2+}/Mn^{2+} dependent, 1K), was first collected for detailed model studies to evaluate the performance and feasibility of this systematic approach (**Chapter 3**). A discussion of bio-experimental results and the perspective of our study showed in **Chapter 4** . The briefly conclude about this study showed in **Chapter 5** . Finally, the future works described in **Chapter 6** .



Chapter 2 Materials and Methods

In this section, we presented the approaches of computational and experimental analyses. The system flow of this study indicated in **Figure 1.10**. The first, we obtained a number of public databases to create dataset. Following, various public tools and in-house programs were integrated into our study. Third, to enable the systematic compilation and updating of our results and additional information, we developed a database, pseudoMap. Finally, processed a serial experiments to verify our *in silico* results.

2.1 Data generation

In total, more than 20,000 human pseudogenes and their cognate genes were obtained from the Ensembl database (Ensembl 63, GRCH37) using BioMart (http://www.ensembl.org/index.html). Affymetrix GeneChip® Human Genome U133A or U133Plus2 is one microarray comprised of oligonucleotide probes to measure the level of transcription of each sequence represented, which included transcribed pseudogenes. Total of 1, 404 pseudogenes have been detectable by this chip, thus considered being transcribed and named transcribed pseudogenes (TPGs). Functional small RNAs (fsRNAs) with sequence length between 18 to 40 nt were collected from the Functional RNA Database (fRNAdb) [62], which hosts a large collection of known/predicted non-coding RNA sequences from public databases: H-invDB v5.0 [10], FANTOM3 [63], miRBase 17.0 [64], NONCODE v1.0 [65], Rfam v8.1 [66], RNAdb v2.0 [67] and snoRNA-LBME-db rel. 3 [68]. The sequences of human miRNAs were obtained from miRBase 18 [79]. Genomic sequences and conservation data were collected from UCSC hg19 [80] (http://hgdownload.cse.ucsc.edu/downloads.html).

2.2 System flow of identifying pseudogene-derived esiRNA-target interactions

Figure 2.1 depicts the workflow for identifying pseudogene-derived esiRNA-target interactions (eSTIs). After collection of pseudogenes, protein-coding genes and fsRNAs, the pseudogene-specific esiRNAs were examined by aligning the pseudogenes with fsRNAs, excluding alignments with parental genes. Candidate pseudogene-specific esiRNAs were validated by reference to publicly-available deep sequencing data from various sRNA libraries. Additionally, eSTIs were analyzed by three target prediction tools and verified with gene expression profiles. Detailed procedures are described below.

2.2.1 Identification of pseudogene-derived esiRNAs

To predict candidate pseudogene-derived esiRNAs, we aligned the sequences of pseudogenes and fsRNAs with sequence length 18-40 bp which obtained from fRNADB. Deep sequencing data of sRNA libraries derived from human embryo stem cells or HCC/liver tissues were used to verify these candidates [81-83]. Then, the extended sequences of these candidate esiRNAs were used to predict hairpin structure by Mfold [77]. Details of publicly-available deep sequencing data are shown in **Table 2.1**.

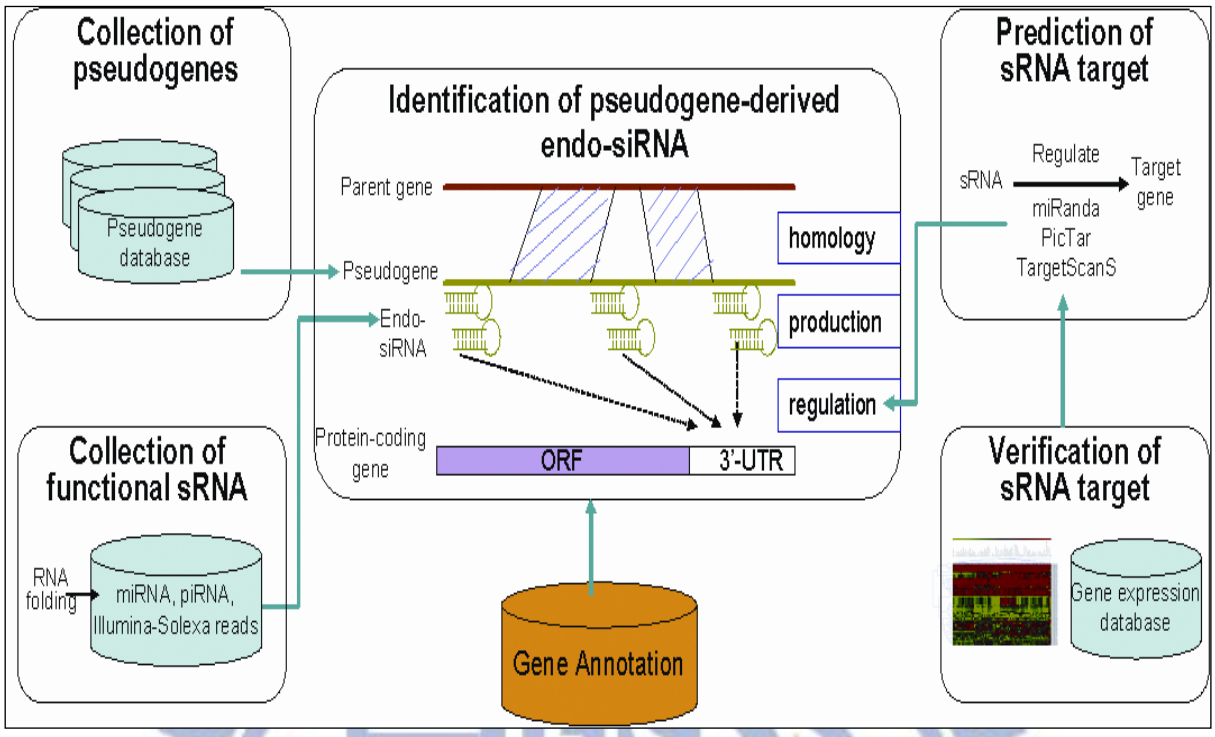


Figure 2.1 Workflow for identification of TPG-derived esiRNA-target interactions.

2.2.2 Identification of esiRNA-target interactions (eSTIs)

Based on experimentally supported data sets, Sethupathy et al. [11] and Baek et al. [64] have shown that the intersection of miRNA target prediction tools can yield improved specificity with only a marginal decrease in sensitivity relative to any individual algorithm. We modified our previous approach [84] for identifying pseudogene-derived esiRNA targets. Briefly, three previously developed computational approaches, TargetScan [85-87], miRanda [75] and RNAhybrid [73] were employed to identify esiRNA target sites within the conserved regions of the 3'-UTR of genes in 12 metazoan genomes. The minimum free energy (MFE) threshold was -20 kcal/mol with score ≥ 150 for miRanda; default parameters were used for TargetScan and RNAhybrid. The three criteria for identifying targets were: (1) potential target sites must be predicted by at least two tools; (2) hits with multiple target sites are prioritised; (3) target sites must be located in accessible regions. Finally, three gene expression profiles were obtained from NCBI GEO [88] to verify those eSTIs with pseudogene expression higher than

their target genes. Gene expression profiles included GDS596 [89], GSE5364 [90] and GSE6222 [91]; detailed experimental conditions are described in **Table 2.1**. The Pearson correlation coefficient was computed for pseudogenes and their target genes.

2.3 miRNA-target interactions (MTI)

In this process, we focus on miRNA-target interactions (MTIs) of TPG and its cognate gene (NG). The schema of identification of MITs is shown in **Figure 2.2**. To obtain the pairs of TPG-parental gene (TPG-NG), TPGs were mapped to all of human protein-coding genes with BLAST (*P*-value≤0.00001) [71]. We calculated the TPG-similarity and NG-similarity with alignment-length/TPG-length and alignment-length/NG-length, respectively. In accordance with the similarity score, we classified the TPG-NG pairs in **Table 2.2**. Following, to explore the relationships of TPG and its cognate gene with miRNA target, we obtained the MTIs from miRTarBase [91], a database curates experimentally validated miRNA-targtet interactions. If there is no experimental data supported to TPG and its cognate gene, we used the eSTIs approach to predict the MTI.

Table 2.1 Summary of public deep sequencing data from various sRNA libraries and gene expression profiles.

Experiment	GEO ID	Reads Count/Samples
Human embryo stem cell-hB*	-	5,026,203
Human embryo stem cell-hESC*	-	5,031,920
Human embryo stem cell -hues6	GSM339994	13,869
Human embryo stem cell -hues6NP	GSM339995	11,883
Human embryo stem cell - hues6Neuron	GSM339996	1,786
HBV(+) Adjacent Tissue Sample 1	GSM531980	358,994
HBV(+) Adjacent Tissue Sample 2	GSM531983	652,641
HBV(+) Distal Tissue Sample 1	GSM531979	372,454
HBV(+) HCC Tissue Sample 1	GSM531982	372,454
HBV(+) HCC Tissue Sample 2	GSM531984	423,338
HBV-infected Liver Tissue	GSM531977	354,524
HBV(+) Side Tissue Sample 1	GSM531981	315,838
HCV(+) Adjacent Tissue Sample	GSM531985	2,503,533
HCV(+) HCC Tissue Sample	GSM531986	2,543,980
HBV(-) HCV(-) Adjacent Tissue Sample	GSM531987	369,926
HBV(-) HCV(-) HCC Tissue Sample	GSM531988	618,689
Human Normal Liver Tissue Sample 1	GSM531974	328,915
Human Normal Liver Tissue Sample 2	GSM531975	282,476
Human Normal Liver Tissue Sample 3	GSM531976	238,522
Severe Chronic Hepatitis B Liver Tissue	GSM531978	259,727
Gene expression profiles of 79 human physiologically normal tissues	GDS596	158
Gene expression profiles of tumor and adjacent non-tumor tissues from colon, liver, lung, oesophagal and thyroid cancers	GSE5364	341
Gene expression profiles of Huh-7 cells, normal liver and cancer tissues	GSE6222	13

* Dataset obtained from ref.36 and ref.37.

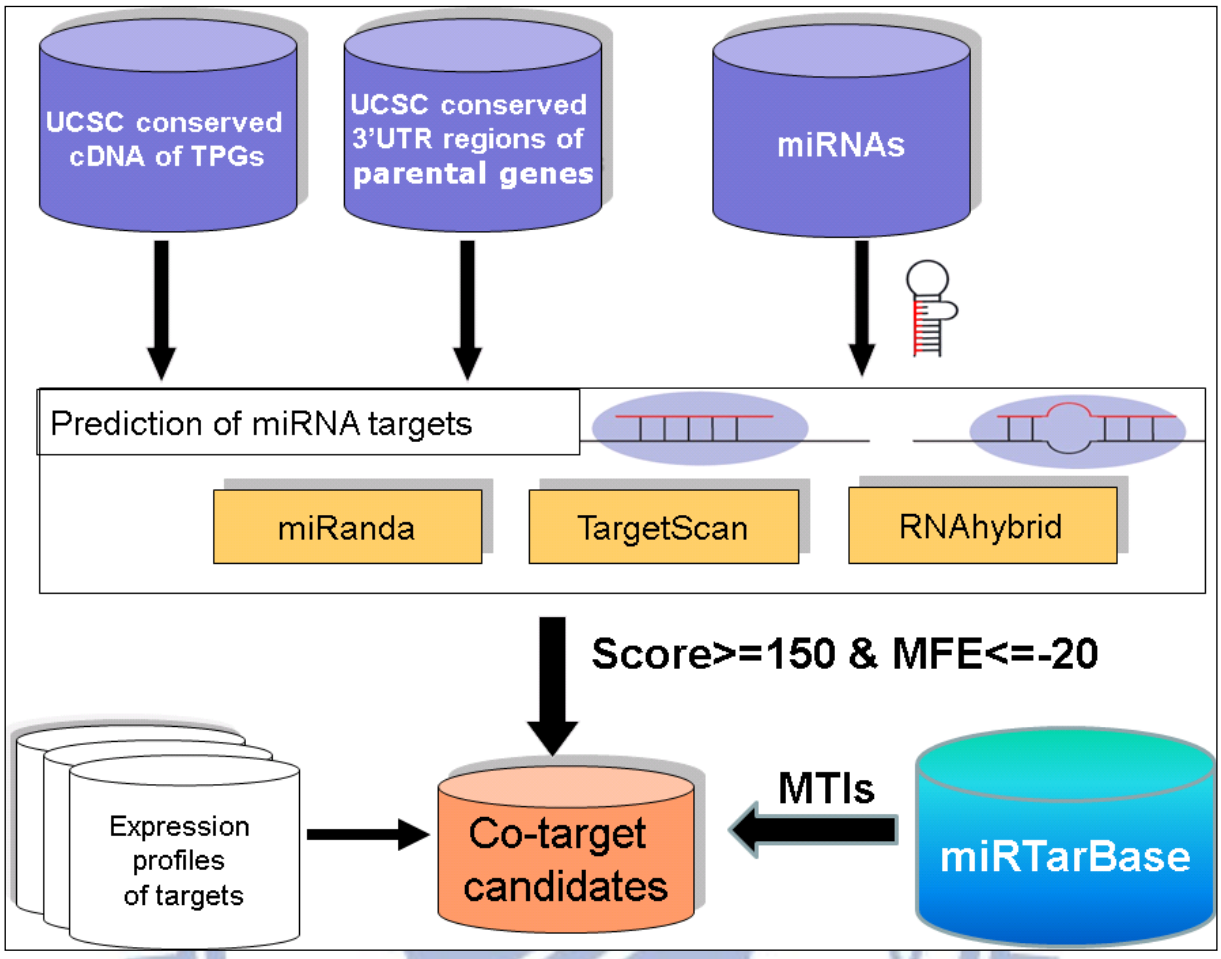


Figure 2.2 Identification of miRNA-target interactions of TPG and its cognate gene.

Table 2.2 Classification of TPG-NG pairs.

Class	Description	Definition
1	TPG highly similarity with NG	TPG-similairty≥0.8 & NG-similairty≥0.8
2	TPG similarity with NG	0.5≤TPG-similarity<0.8 & 0.5≤NG-similarity<0.8
3	TPG highly similarity with partial NG	TPG-similairty≥0.8 & NG-similarity<0.5
4	NG highly similarity with partial TPG	TPG-similarity<0.5 & NG-similairty≥0.8
5	TPG low similarity with NG	TPG-similarity<0.5 & NG-similarity<0.5

2.4 Gene expression analysis

The mRNA abundances of TPGs and protein-coding genes were obtained from Gene Expression Omnibus **[88]**, such as GDS596 examined from 79 human physiologically normal tissues **[89]**, GSE2109 examined from 2158 samples with 61 tumor tissues, GSE3526 examined from 353 samples with 65 normal tissues **[35]** and GSE5364 examined from primary human tumors and adjacent non-tumor tissues, which include 270 tumors and 71 normal-cancer pairs from patients with breast, colon, liver, lung, oesophagal and thyroid cancers **[90]**. Moreover, the Pearson correlation coefficient was computed from TPGs and protein-coding genes.

2.5 GO and KEGG enrichment analyses

The function of target genes was examined by performing GO terms and KEGG pathway enrichment annotation [92] using the DAVID gene annotation scheme [93]. DAVID is a biological knowledgebase which integrates up to 60 published resources for functional annotations of large gene/protein lists. To utilize DAVID for GO terms and KEGG pathway enrichment analysis of esiRNA targets, the Fisher's exact P-value were calculated to ensure that the enrichments possess the property of statistical significance rather than random occurrence. The associated biological meanings of esiRNA-mediated genes allow investigators comprehensively understanding the molecular functions of these esiRNAs.

2.6 Construction of pseudoMap

To enable the systematic compilation and updating of these results and additional information, a database, pseudoMap, capturing various types of information, including sequence data, TPG and cognate annotation, deep sequencing data, RNA-folding structure, gene expression

profiles, miRNA annotation and target prediction, will be developed. In pseudoMap, various databases and tools (**Table 2.3**) for mining potential regulators and functions of human TPGs are integrated and maintained with MySQL (http://www.mysql.com/) relational database management system. While operating on an Apache HTTP server (http://www.apache.org/) and PHP (http://www.php.net/) on a Linux operation system (http://www.linux.com/), pseudoMap was constructed using the Smarty template engine (http://www.smary.net/). Based on PHP, JavaScript (http://www.javascriptsource.com/), CSS (http://www.w3schools.com/css/) and HTML languages (http://www.w3schools.com/html/), the web interface enables dynamic MySQL queries with user-friendly graphics.

2.7 Bio-experiments

In accordant with predicted distinct esiRNAs, RNA structure folding, multiple target sites, gene expression profiles as well as the functions of both the cognate and target gene, the *ψPPM1K*, a partial retrotranscript from *PPM1K* (protein phosphatase, Mg^{2+}/Mn^{2+} dependent, 1K), was first collected for detailed model studies. The detailed experiments described following sections.

2.7.1 Samples

Resected primary HCC and nearby non-cancerous tissue samples (n=41) were obtained from 41 patients at the Changhua Christian Hospital. The tumor tissues were composed of 90-100% tumor cells and frozen immediately after surgical resection, then stored in liquid nitrogen until extraction of either RNA or DNA. All studies were approved by the Institutional Review Board of Changhua Christian Hospital.

Table 2.3 Supported databases and tools in pseudoMap.

Integrated database or tools	Dataset	Description
miRBase **[79, 94]**	miRNA annotation	This database not only provides published miRNA sequences and annotations but also supplies known/predict targets.
fRNAdb **[62]**	sRNA annotation	A database to support mining and annotation of functional RNAs.
Ensembl Genome Browser **[95]**	Pseudogene, protein-coding gene	It produces genome databases for vertebrates and other eukaryotic species.
UCSC Genome Browser **[80]**	Conserved region Genomic view of genes	This browser provides a rapid and reliable display of any requested portion of genomes at any scale, together with dozens of aligned annotation tracks.
GeneCards **[96]**	Gene annotation	GeneCards is a searchable, integrated, database of human genes that provides concise genomic related information, on all known and predicted human genes
Mfold **[77]**	RNA folding tool	Folding RNA structure
GEO **[88]**	Gene expression profiles and deep sequencing data	A public functional genomics data
BLAST **[97]**	Sequence alignment tool	BLAST finds regions of similarity between biological sequences.

2.7.2 Cell culture

Human hepatoma Huh-7 and HepG2 cells were grown using standard procedures for all experiments. Cells were maintained in DMEM supplemented with 10% FBS, 2 mM glutamine, and antibiotics (100 units/ml penicillin and 100 μg/ml streptomycin) at 37℃ in a humidified atmosphere of 5% CO_2 incubator.

2.7.3 RNA isolation, reverse transcription and real-time quantitative PCR (RT-qPCR) analysis

RNA isolation from specimens or cultured cells and reverse transcription were performed as described [98, 99]. RT-qPCR analysis of *PPM1K* and *PPMIK* in HepG2 and Huh-7 cells, and of *PPM1K*, *NEK8*, *TBRG1* and *BMPR2* in *ψPPM1K*-expressing Huh-7/HepG2 stable cell lines, was performed using SYBR Green with the ABI 7500 Real-Time PCR System (Applied Biosystems). RT-qPCR of precursor esiRNA1 (24-144 nt), precursor esiRNA2 (170-273 nt), *PPM1K*, and *NEK8* in paired HCC tumor and non-tumor tissues was performed using a LightCycle 480 (Roche, Mannheim, Germany) with a primer/probe system. The specific primer/probe sets are shown in Table 2.3. All RNA expression levels were normalized to *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) RNA with the △Ct method according to Liu et al [100].

RT-PCR of mature esiRNA1 levels in Huh-7 stable cell lines was performed using a TaqMan MicroRNA Assay designed for esiRNA1 according to the manufacturer's instructions (Applied Biosystems) following isolation of small RNA with the mirVana miRNA Isolation Kit. U6 small nuclear RNA was used as an internal control.

2.7.1 Northern blot of pseudogene-derived esiRNAs

Northern blotting was performed according to a previous study [101] with minor modifications. Briefly, 10 μg of total RNA from human hepatoma cell line Huh-7 were dissolved in loading buffer (50 mM EDTA, 8 M urea, 20% formamide, xylene cyanol), loaded onto a 2% agarose gel, then run for 1.5 h at 120 V at room temperature. The biotin-labeled esiRNA probes (5′- GTGGCACGCGCCTGTAGTCCCAGC-3′ for esiRNA1 and 5'-GAGGCAGGAGAATGGCGTGAACC-3' for esiRNA2, Genomics BioSci & Tech Co.,

Taipei, Taiwan) were used as the positive control for the avidin-biotin reaction and the size control for esiRNAs. The agarose gel was incubated sequentially in 0.05 M NaOH/NaCl, 0.05 M Tris/NaCl and 2x sodium citrate. Then RNA was transferred to a nitrocellulose membrane (Pall Corporation, East Hills, NY, USA) followed by cross-linking with 254-nm UV radiation. The membrane was hybridized with the biotinylated esiRNAs overnight, then membranes were washed sequentially with 2x SSC/0.1% SDS, 1x SSC/0.1% SDS and 0.5x SSC/0.1% SDS at 42℃. The membrane was incubated with horseradish peroxidase (HRP)-conjugated avidin (Biolegend, SanDiego, CA, USA) and probe detected by chemiluminescence with the WesternBrightTM-ECL kit (Advansta, Menlo Park, CA, USA).

Table 2.4. The sequences of the probes and primers used in RQ-PCR/RT-PCR.

Gene	Probe seq.	Primer (5'→3')
siRNA1 (24-144 nt)	CTCTGCCT	F: GGAGTACAGTGGTGCGGTCT
		R: GCTGAGGCAGAAGAATCGTT
siRNA2 (170-273 nt)	GGCTGGAG	F: GAGAAGGAGTCTCTCTCTGTCACC
		R: CAGTGAGCCGAGATCGTG
PPM1K	TGGGGCAG	F: TGACCATTGACCATACTCCAGA
		R: CAAGCCTGCCATTTACGTG
NEK8 (for tissues)	CTGGGGCC	F: TGTCCACTGAGCGAGAACTATT
		R: TCTGACCCCGATCCGTGG
GAPDH	TGGGGAAG	F: AGCCACATCGCTCAGACAC
		R: GCCCAATACGACCAAATCC
TPG_PPM1K	-	F: GGAATTCCTCCATCAGCTGTTCGTTTG
		R: GCTCTAGATGGCAAAACCCCATCTCTAC
NEK8 (for cell)	-	F: TCCACTGAGCGAGAACTATTTGC
		R: GGATCATGGAGGAATCGATACC
TBRG1	-	F: CCGTGGGCTATTGCAGTACTC
		R: AAGAGCTGACAATGGCATTCTG
BMPR2	-	F: GGCCATCAAAGCCCAGAAG
		R: CTGATCCTGATTTGCCATCTTG

2.7.2 Fluorescent in situ hybridization (FISH)

Nocodazole and colchicine were added to cell lines before FISH was performed. Interphase and metaphase spreads were prepared for FISH using standard methods [102]. DNA probes (S1: GTGGCACGCGCCTGTAGTCCCAGC, antisense of esiRNA1; S2: GAGGCAGGAGAATGGCGTGAACC, antisense of esiRNA2; scramble 1: GTGGCTCATGCCTGTAATCCCAGCACTTTG; and scramble 2: TTAAGACATACAAAGATCTGGCCAGGTGCG) were mixed with hybridization buffer, centrifuged, and heated to 73℃ for 5 min in a water bath. Slides were immersed in 70% formamide/2× standard saline citrate for 5 min at 73℃, followed by dehydration, dried, and hybridized with probe mix in a 42℃ incubator for 30 min. Slides were then washed in 0.4× standard saline citrate/0.3% NP-40 for 2 min, air dried in the dark, and counterstained with DAPI (4, 6-diamidino-2-phenylindole) (1 μg/ml, Abbott, Illinois, USA). Imaging was performed on a Nikon E600 microscope with cytovision software.

2.7.3 Transduction of the pseudogene transcript in Huh-7 and HepG2 stable cell lines

TPG-expressing (and vector control) Huh-7/HepG2 stable cell lines were established by G418 selection after transfection with the *ψPPM1K*-expression plasmid or blank vector. Total RNA was isolated from cells and subjected to RT-PCR analysis to amplify *ψPPM1K* mRNA (primers showed in **Table 2.4**). The PCR was performed with a denaturing step at 95℃ for 2 minutes, then 30 cycles of 30 s at 95℃, 1 min at 60℃ and 1 min at 72℃, followed by a final 7 min at 72℃.

2.7.4 Cell proliferation assay

To investigate the proliferation of *ψPPM1K*-expressing Huh-7 stable cell lines, 2.5×10^4 cells were plated in each well of a 12-well plate. Cells were trypsinized and counted with a hematocytometer every day until day 4. Each experiment was repeated twice in triplicate wells, separately. Huh-7 TPG7 cells were transfected with *NEK8*-overexpressing plasmid or empty vector using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA). The *NEK8*-overexpressing plasmid, which contains the human *NEK8* ORF without 3'-UTR in the pCMV6-Entry vector, was obtained from Origene (Rockville, MD, USA). Twenty-four hours after transfection, 2.5×10^4 cells were plated in each well of a 12-well plate. Cells were trypsinized and counted with a hematocytometer every day until day 4.

2.7.5 Clonogenic activity

For determination of clonogenic activity, we plated 1000 cells of mock2, TPG1, TPG2 and TPG7 in 10 ml growth medium in 100 mm dishes. Soft agar culture was also performed by inoculating 500 cells/ml in 0.3% agar-growth medium over 0.5% agar-growth medium in 6-well culture dishes. The dishes were incubated under normoxic 19% O_2 and hypoxic 3% O_2 in 5% CO_2 incubators for 12 days and fixed/stained for counting colony formation. We also took serial photographs of the same colonies at day 5, day 7 and day 9 to visualize the two-dimensional growth of mock2 and three transfected clones.

2.7.6 Transfection of synthetic siRNA1 into Huh-7 cells

siRNA1 was chemically synthesized by Invitrogen (Carlsbad, CA, USA). Oligonucleotides were annealed before use in annealing buffer containing 100 mM potassium acetate, 30 mM Hepes-KOH (pH 7.4), and 2 mM magnesium acetate. Negative Control #1 siRNA was also

obtained from Invitrogen. Huh-7 cells in 6 cm culture plates were transfected with 200 pmol siRNA using 10 μl Lipofectamine 2000 according to manufacturer's instructions.

2.7.7 Construction of the esiRNA1-deleted *ψPPM1K*-expressing plasmid

To delete the esiRNA1 sequence from *ψPPM1K*, the overlap extension method of PCR-based mutagenesis was used. First, two complementary mutagenic primers, forward 5'-CCTCAGCCTCCTGAGTACACCCCTGGCTAATTTT-3' and reverse 5'-AAAATTAGCCAGGGGTGTACTCAGGAGGCTGAGG-3', were synthesized. Two PCRs using the mutagenic forward primer/outer *ψPPM1K* reverse primer pair and the mutagenic reverse primer/outer *ψPPM1K* forward primer pair were performed to amplify the right and left *ψPPM1K* fragments, respectively. The two fragments were then mixed and further amplified using the outer *ψPPM1K* primers to generate the esiRNA1-deleted *ψPPM1K* fragment. Finally, this fragment was inserted between the *Eco*RI and *Xba*I sites of the pCI-neo vector to generate the esiRNA1-deleted *ψPPM1K*-expressing plasmid.

2.7.8 Mitochondrial activities

For indirect assay of mitochondrial membrane potential and permeability transition pore activity [103], overnight-plated monolayer cells on 100 mm dishes were exposed to 0.5 g/ml or 1.0 g/ml Rh123 in the growth medium [104]. The kinetics of dye uptake were determined by harvesting the cells after 10 min, 30 min, 5 h and 24 h incubation in Rh123-containing media. To determine Rh123 retention activity of cancer cells [103-105], monolayer cells exposed to Rh123-containing medium for 30 min were rinsed 3x with Hanks' balanced salt solution (HBSS) to remove the dye, and replenished with fresh medium for a further 18 h

incubation before harvest. Cells harvested by trypsinization were washed 2x with cold PBS and collected by centrifugation at 300x *g* at 4℃ for 5 min. With or without further reaction with fluorescent monoclonal antibody against cell surface markers, e.g. CD133/Prominin (BD Pharmingen), the doubly- or triply-labeled washed cells were analyzed in a fluorocytometer.

2.7.9 miRNA-mediated knockdown of *PPM1K* and *ψPPM1K*

The stable negative control (siCon: 5'FAM-UUC UCC GAA CGU GUC ACG UTT), has-miR-650 (miR-650: 5'-AGG AGG CAG CGC UCU CAG GAC) and has-miR-3174 (miR-3174: 5'-UAG UGA GUU AGA GAU GCA GAG CC-3') miRNAs were purchased from GeneDireX, Inc. (Las Vegas, NV, USA) and transfected into cells by Lipofetamine™ RNAiMax (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. The efficacy of mRNA knockdown after miRNA transfection for 48 h was determined by RT-qPCR.

2.7.10 Statistical analysis

Student's t-test was used for analysis of the cell assays. Significance was accepted at *P*-value < 0.05.

Chapter 3 Results

In this chapter, we described the results of computational analyses and subsequent described the experimental results. The detailed results were described in following sections.

3.1 Overview of TPGs

To verify the hypothesis that human pseudogenes may generate esiRNA to regulate protein-coding genes, we developed a computational pipeline (**Figure 2.1**). A total of 16,524 genes including 15,003 pseudogenes and 1,521 processed transcripts were collected from BioMart and integrated in the Ensembl Genome Browser, filtering by gene type as "pseudogene" or "pseudogene-related gene". The percentage of pseudogenes on each chromosome is similar to that of protein-coding genes (**Figure 3.1**). Of these pseudogenes, 1,404 are detectable by the Affymetrix Human Genome U133A/U133Plus2 microarray, and are thus considered to be transcribed pseudogenes (TPGs). Consequently, we determined the gene expression profiles (GDS596) by $\log_2$ average intensity of TPGs in 79 normal human tissues. Hierarchical clustering showed that most TPGs were highly expressed in 19 tissues, especially in blood-related cells where expression levels were ≥ 4-fold higher than average (**Figure 3.2**). Another assessment of gene expression in six different tissues (GSE5364) showed that TPGs were often expressed differently in tumor material compared to paired non-tumor tissue (**Figure 3.3a**). In particular, most TPGs in HCC samples were under-expressed in comparison to normal tissues (GSE6222) (**Figure 3.3b**).

3.2 TPGs v.s. NGs

According to the sequence mapping results, total of 1313 TPG-NG pairs were obtained. The statistics showed that the classifications of highly similar (TPG-similarity≥0.8), similar

(0.5<TPG-similarity<0.8), and low similar (TPG-similarity≤0.5) are 1004, 179, 130 TPGs, respectively (**Figure 3.4**). To further understand the biological functions of these TPGs, enriched GO terms and KEGG pathways of TPG-paired NGs were annotated by using DAVID. The enriched KEGG pathways and GO terms with *P*-value small than 0.001 are presented in **Figure 3.5**. The results showed that the GO terms in 3 groups of TPG-paired NGs are alike involving in protein binding, nucleic acid binding and nucleotide binding in molecular function-term and cellular metabolic process, biosynthetic process and cellular component assembly in biological process-term, respectively (**Figure 3.5**). Total of 1000 TPG-NG pairs, there are 288 TPGs with the same probesets of paired NGs, 264 paired NGs with no probesets, and 448 TPGs with the unique probesets in Affymetrix U133A chip. There are 236 TPGs with the same probesets of paired NGs, 262 paired NGs without probesets, and 619 TPGs with the unique probesets in Affymetrix U133Plus2 chip. To detect the expression profiles of TPGs and paired NGs, the intensity of TPG referenced by paired NG with $\log_2$ ratio in a number of public gene expression profiles. Hierarchical clustering showed that about half of TPGs (200/448) were highly expressed than its paired NG. in 79 normal tissues (**Figure 3.6**). Another assessment of gene expression in 61 tumor tissues (GSE2109) showed that nearly half of TPGs (250/619) were highly expressed than its paired NG. (**Figure 3.7**).

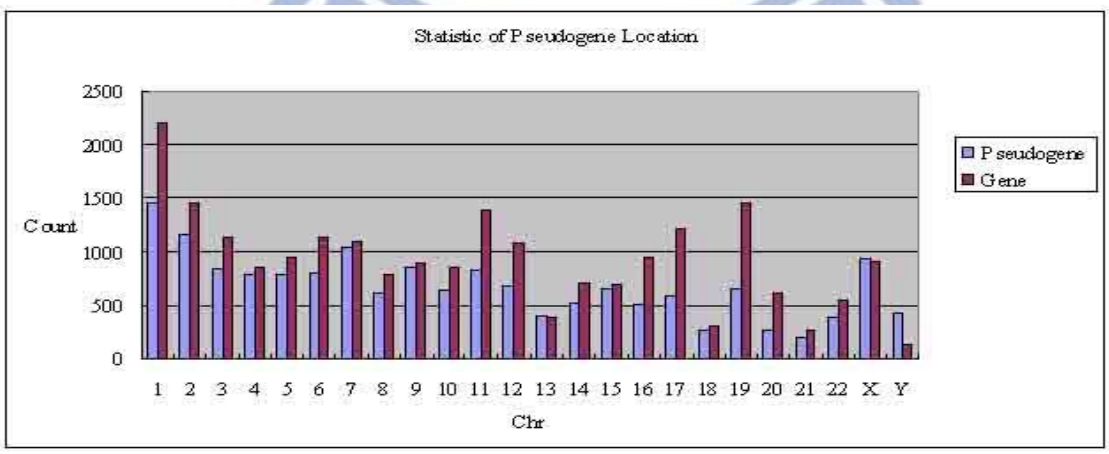


Figure 3.1 Chromosome location of pseudogenes and protein-coding genes.

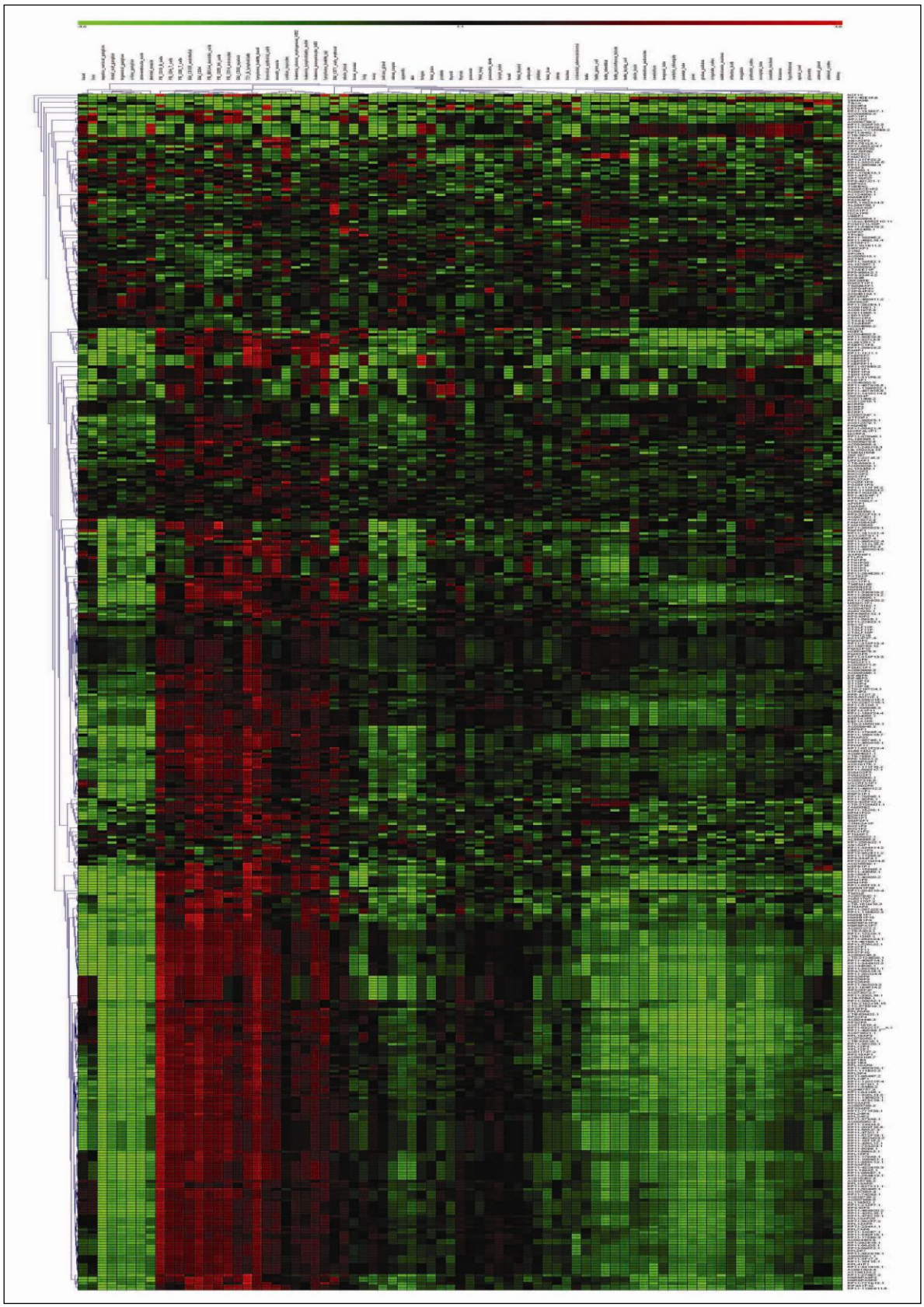


Figure 3.2 Gene expression profiles of TPGs in 79 human physiologically normal tissues.

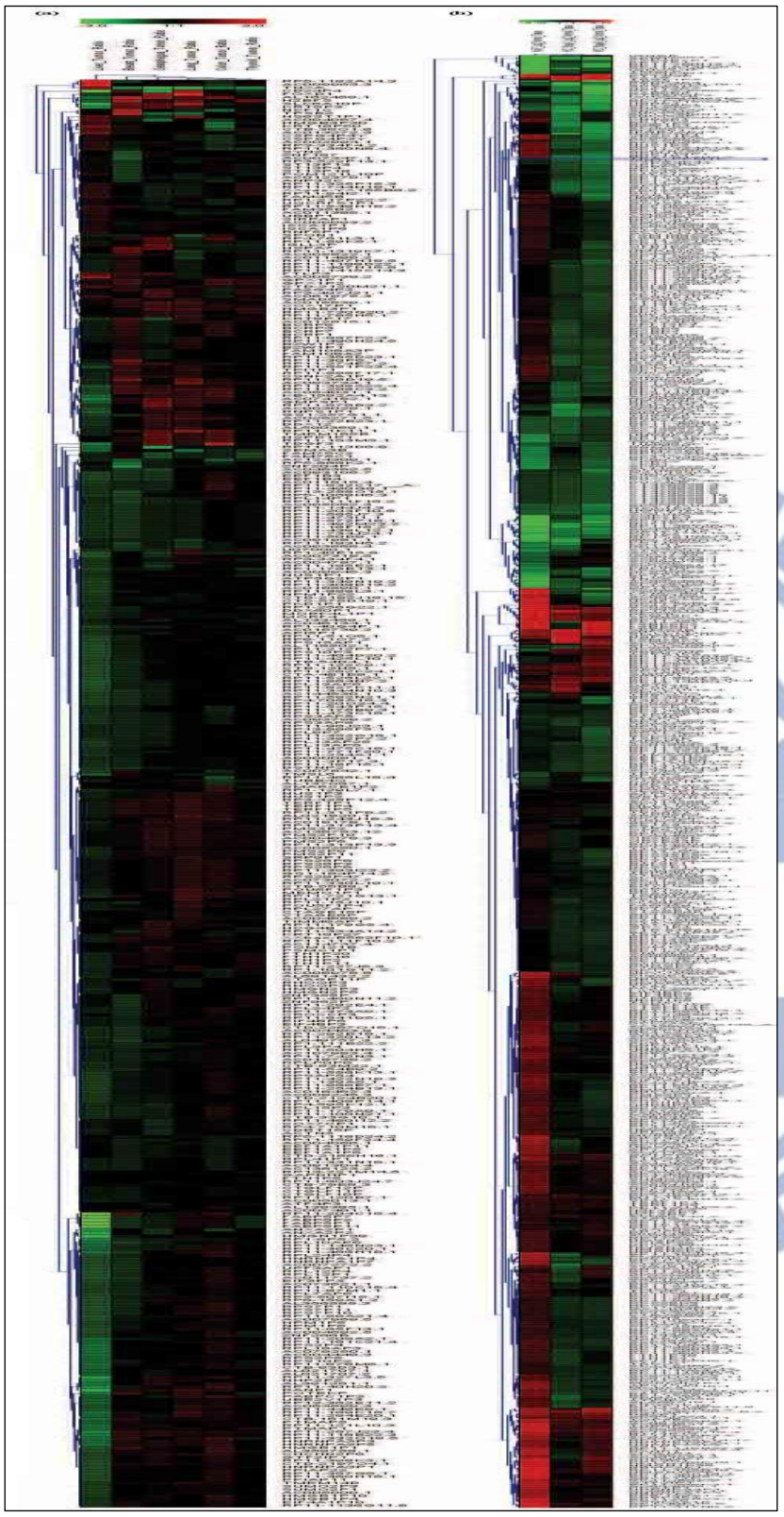


Figure 3.3 Gene expression profiles of TPGs. (a) Expression heat-map of TPGs in pair of cancers compared to non-malignant tissues. (b) TPGs expression profiles in the tumor tissues/cells referenced by normal liver tissues.

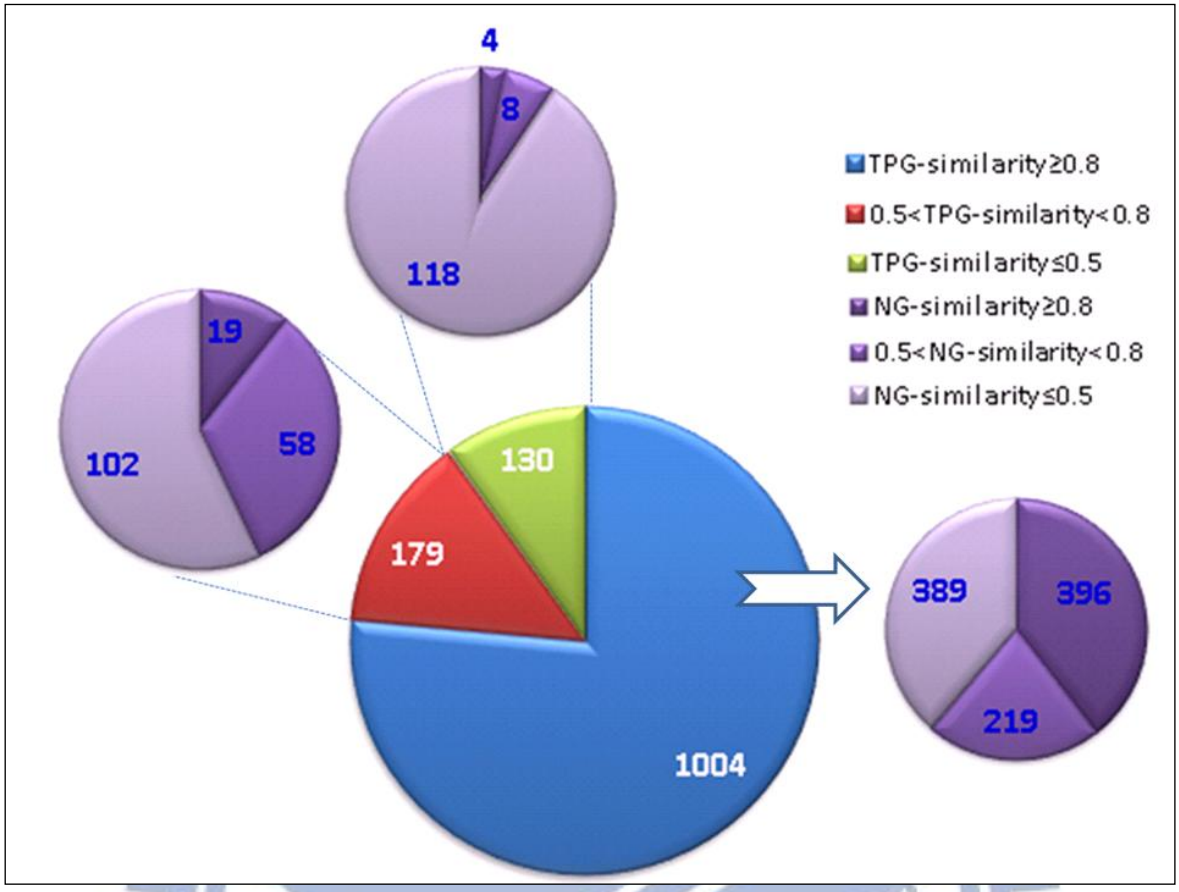


Figure 3.4 Statistics of classification of TPG-NG pairs.

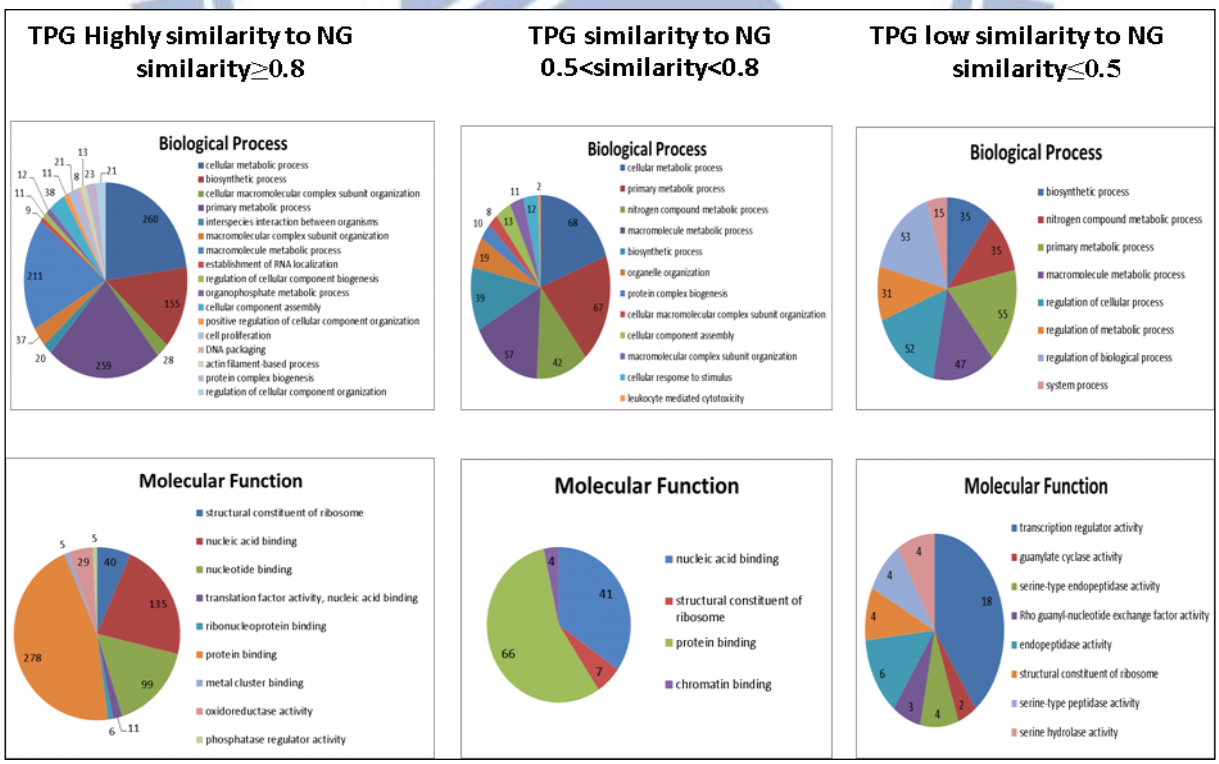


Figure 3.5 Enriched GO terms of NGs in classification of TPG-NG pairs.

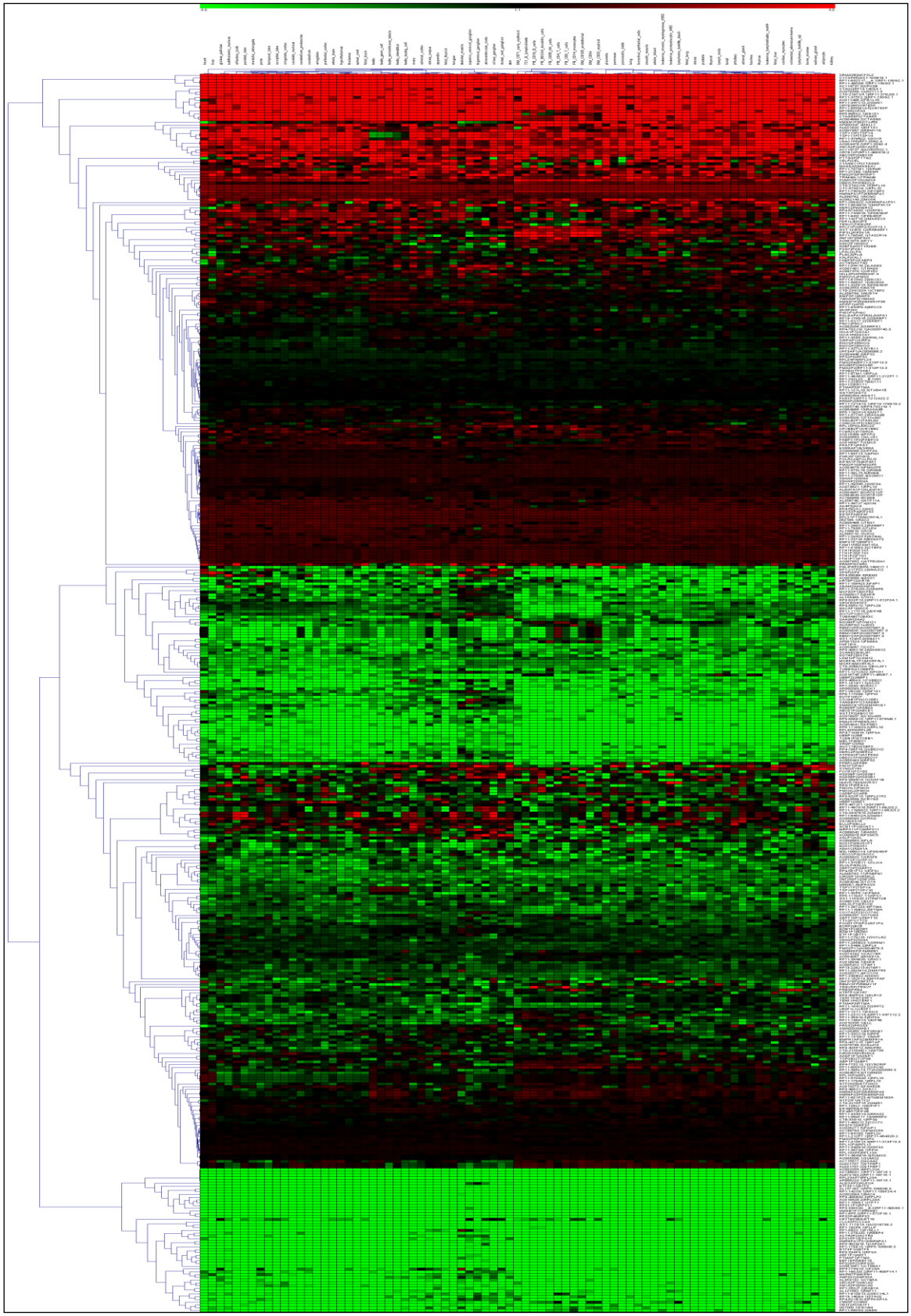


Figure 3.6 Gene expression profiles of TPGs referenced by its paired NG in 79 human physiologically normal tissues.

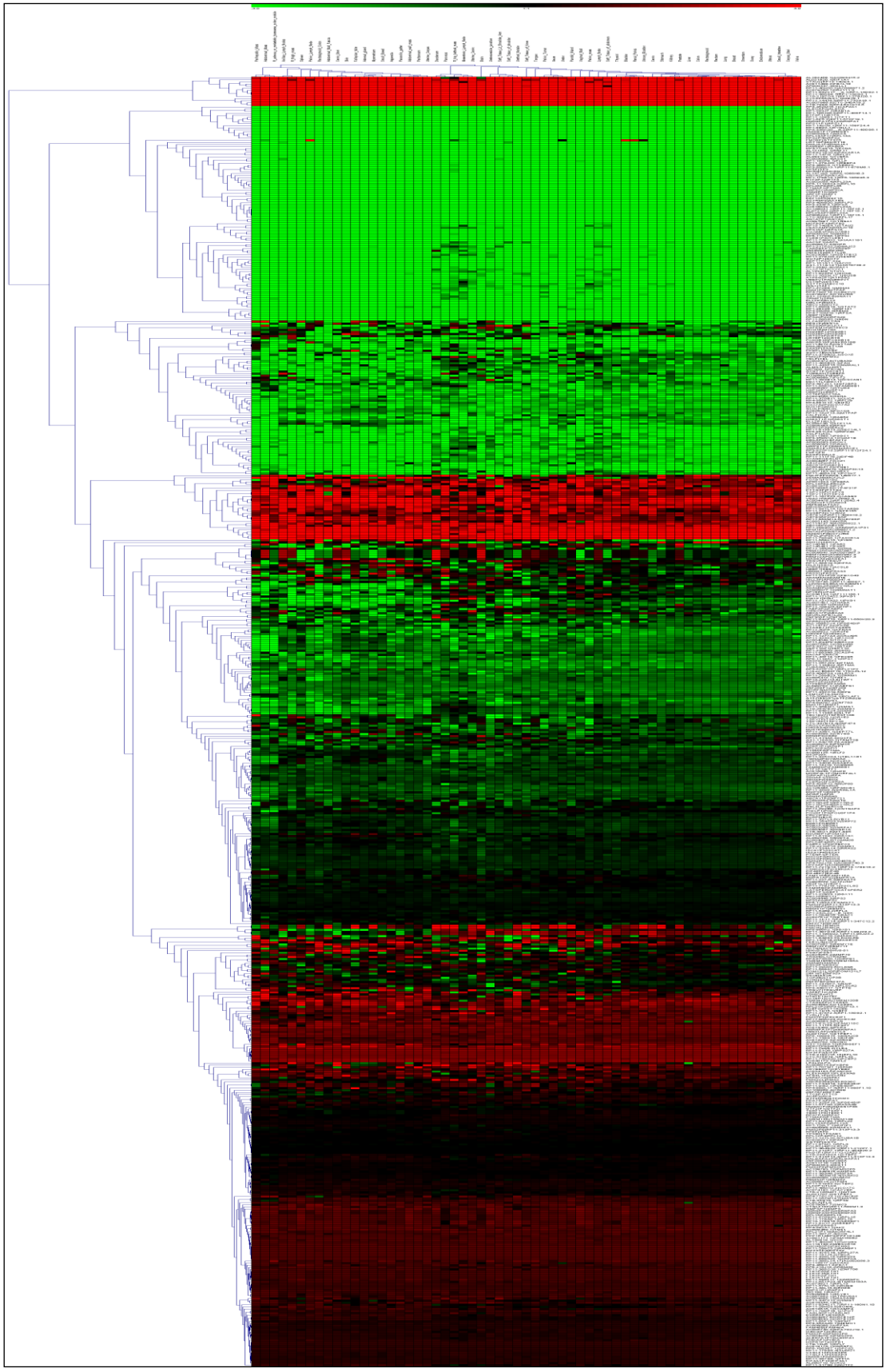


Figure 3.7 Gene expression profiles of TPGs referenced by its paired NG in 61 human tumor tissues.

3.3 TPGs-derived esiRNAs-target interactions

To identify the miRNA-derived from human TPGs, the miRNA sequences were obtained from miRBase release 18 and mapped to human TPGs. The results showed that total of 13 miRNAs may derive from human TPGs (**Table 3.1**). One of these miRNAs, has-miR-622, was derived from *KRT18P27* (keratin 18 pseudogene 27), located on nt 858-879, as similar as miRBase database. We also found that has-miR-622 down-regulation of HRAS and KRAS experimented by luciferase reporter assay and Western blot (**Table 3.1**). Total of 63 non-duplicate piRNAs, which supported by sRNA deep sequencing data, possibly were derived from TPGs (**Table 3.2**). In addition to miRNAs and piRNAs, TPGs may produce a large number of sRNAs which supported by in public deep sequencing data from various sRNA libraries (**Appendix I**).

Table 3.1 Human TPG-derived miRNA-target interactions.

Ensembl_ID	Gene_Name	Description	Chr	miRNA	Targets	Validation Method
ENSG00000224419	KRT18P27	keratin 18 pseudogene 27	13	hsa-miR-622	HRAS, KRAS	Luciferase reporter assay, Western blot
ENSG00000099725	PRKY	protein kinase, Y-linked, pseudogene	Y	hsa-miR-5585-3p	SLC20A2, PKD1, UST, RABL3, MAPK1, NPSR1, COLQ, etc.	Prediction
ENSG00000180771	SRSF8	serine/arginine-rich splicing factor 8	11	hsa-miR-5096	MGAT4A, ANK3, LCORL, YWHAG, TTPAL, CCDC132, etc.	Prediction
ENSG00000185485	SDHAP1	succinate dehydrogenase complex, subunit A, flavoprotein pseudogene 1	3	hsa-miR-1273g-3p	C3ORF49,ARD1B,SMURF1,NFAT5,CA13,NCOA6,PNPLA8, etc	Prediction
ENSG00000196912	ANKRD36B	ankyrin repeat domain 36B	2	hsa-miR-1273g-3p	see above	Prediction
ENSG00000205771	CATSPER2P1	cation channel, sperm associated 2 pseudogene 1 [Source:HGNC Symbol;Acc:31054]	15	hsa-miR-5096	see above	Prediction
ENSG00000224479	AC136289.1		3	hsa-miR-1273g-3p	see above	Prediction
ENSG00000234420	ZNF37BP	zinc finger protein 37B, pseudogene [Source:HGNC Symbol;Acc:13103]	10	hsa-miR-5096	see above	Prediction
ENSG00000242193	RP11-568K15.1		1	hsa-miR-1273g-3p	see above	Prediction
ENSG00000243468	INGX	inhibitor of growth family, X-linked, pseudogene	X	hsa-miR-5585-3p	see above	Prediction
ENSG00000251474	RPL32P3	ribosomal protein L32 pseudogene 3	3	hsa-miR-1273g-3p	see above	Prediction
ENSG00000255185	RP11-106J23.2		16	hsa-miR-5096		Prediction
ENSG00000257267	ZNF271	zinc finger protein 271	18	hsa-miR-5096		Prediction

Table 3.2 The deep sequence data supported piRNAs-derived from TPGs.

No	piRNA-Seq	TPG-ID	TPG-Symbol
1	GCGGATGATAATCTCAAGACACCT	TPG0000769,TPG0001000,	PKD1P1,RP11-347C12.2,
2	GCTCACGCCTGTAATCCCAGCACTTTG	TPG0000115,TPG0001051,TPG0001065,TPG0001066,TPG0001193,TPG0001199,TPG0001250,TPG0001291,TPG0001310,TPG0001313,TPG0001324,TPG0001330,TPG0001375,TPG0001376,TPG0001396,	AC009093.2,BMS1P1,RP9P,CATSPER2P1,FAM92A3,CEACAM22P,GSTTP2,RP11-568K15.1,INGX,RPLP0P2,RP11-515B12.1,RRN3P2,RP11-106J23.2,AP003068.18,RP3-405J10.2,
3	TGTAATCCCAGCACTTTGGGAGGCC	TPG0000155,TPG0000327,TPG0000115,TPG0000953,TPG0000970,TPG0000990,TPG0001051,TPG0001066,TPG0001157,TPG0001199,TPG0001238,TPG0001302,TPG0001305,TPG0001310,TPG0001324,TPG0001330,TPG0001376,	ZNF204P,AC011498.2,AC009093.2,AC002472.8,AP002387.1,SRSF8,BMS1P1,CATSPER2P1,AC005077.12,CEACAM22P,RP11-231P20.2,ZNF702P,ABCC13,INGX,RP11-515B12.1,RRN3P2,AP003068.18,
4	CTGGTCTCGAACTCCTGACCTCA	TPG0000133,TPG0000907,TPG0000327,TPG0000118,TPG0001030,TPG0001175,	AC012615.1,AC007401.1,AC011498.2,AL133458.1,UBE2Q2P1,AC004985.13,
5	TGGTCTCGAACTCCTGACCTCA	TPG0000133,TPG0000327,TPG0001030,	AC012615.1,AC011498.2,UBE2Q2P1,
6	TAAATGCGGTGGCATCGACAAAAGAAC	TPG0000574,TPG0000111,	EEF1A1P6,EEF1A1P5,
7	TGACAAAGAAAAGAAGGAACAA	TPG0000598,TPG0000335,TPG0000594,TPG0000170,TPG0000150,TPG0000755,TPG0000751,TPG0000366,TPG0000168,TPG0001036,	AC073072.7,RP11-393I23.2,RPS26P47,RPS26P6,RPS26P8,CTB-55B8.1,RP11-330L19.1,GS1-184P14.2,RPS26P3,AC004057.1,
8	GGCTGGTCTCGAACTCCTGACCTCA	TPG0000133,TPG0000327,TPG0000005,TPG0000907,TPG0000344,TPG0001030,	AC012615.1,AC011498.2,CROCCP3,AC007401.1,POM121L10P,UBE2Q2P1,

9	GCCTCCCAAAGTGCTGGGATTACAGGC	TPG0000138,TPG0000327,TPG0000802,TPG0000344,TPG0000955,TPG0001051,TPG0001090,TPG0001092,TPG0001116,TPG0001191,TPG0001276,TPG0001313,TPG0001326,TPG0001328,	RP5-886K2.1,AC011498.2,ZNF286B,POM121L10P,NEURL3,BMS1P1,SEPT7P2,POLR2J4,SMPD4P1,GS1-124K5.2,PLGLA,RPLP0P2,RP11-219E24.1,RP11-33B1.1,
10	CAAAGTGCTGGGATTACAGGCGTGAGC	TPG0000138,TPG0000327,TPG0000802,TPG0000955,TPG0000974,TPG0001051,TPG0001079,TPG0001116,TPG0001191,TPG0001276,TPG0001308,TPG0001313,TPG0001326,TPG0001328,	RP5-886K2.1,AC011498.2,ZNF286B,NEURL3,SCAND2,BMS1P1,AC016629.2,SMPD4P1,GS1-124K5.2,PLGLA,PTPRVP,RPLP0P2,RP11-219E24.1,RP11-33B1.1,
11	GCCTCCCAAAGTGCTGGGATTACA	TPG0000137,TPG0000138,TPG0000327,TPG0000925,TPG0000955,TPG0001051,TPG0001090,TPG0001092,TPG0001112,TPG0001116,TPG0001191,TPG0001276,TPG0001281,TPG0001328,	AL021707.1,RP5-886K2.1,AC011498.2,FER1L4,NEURL3,BMS1P1,SEPT7P2,POLR2J4,AC093642.5,SMPD4P1,GS1-124K5.2,PLGLA,RP11-147I3.1,RP11-33B1.1,
12	GATCGAGACCATCCTGGCTAAC	TPG0000327,TPG0000987,TPG0001310,	AC011498.2,AC034193.5,INGX,
13	CCAGGCTGGAGTGCAGTGGTGC	TPG0000327,TPG0000975,TPG0001003,TPG0001130,TPG0001351,	AC011498.2,FKBP9L,HTR7P1,AC136289.1,RPL32P3,
14	CACTTTGGGAGGCCGAGGCGGG	TPG0000327,TPG0000999,TPG0001103,TPG0001199,TPG0001305,TPG0001325,TPG0001354,	AC011498.2,PI4KAP2,PI4KAP1,CEACAM22P,ABCC13,RP3-449O17.1,GUSBP3,
15	TTCTTGCCAAGAGAATTAATGTG	TPG0000480,	RP11-492M23.2,

16	AAAGTGCTGGGATTACAGGCGTGAGC	TPG0000584,TPG0000138,TPG0000327,TPG0000802,TPG0000955,TPG0000974,TPG0001051,TPG0001079,TPG0001116,TPG0001191,TPG0001276,TPG0001308,TPG0001313,TPG0001326,TPG0001328,	RP11-380N8.4,RP5-886K2.1,AC011498.2,ZNF286B,NEURL3,SCAND2,BMS1P1,AC016629.2,SMPD4P1,GS1-124K5.2,PLGLA,PTPRVP,RPLP0P2,RP11-219E24.1,RP11-33B1.1,
17	AAAGTGCTGGGATTACAGGCGTGAG	TPG0000584,TPG0000138,TPG0000327,TPG0000802,TPG0000955,TPG0000974,TPG0001010,TPG0001051,TPG0001079,TPG0001092,TPG0001116,TPG0001191,TPG0001276,TPG0001308,TPG0001313,TPG0001326,TPG0001328,	RP11-380N8.4,RP5-886K2.1,AC011498.2,ZNF286B,NEURL3,SCAND2,CIDECP,BMS1P1,AC016629.2,POLR2J4,SMPD4P1,GS1-124K5.2,PLGLA,PTPRVP,RPLP0P2,RP11-219E24.1,RP11-33B1.1,
18	AGAACATGATGGCTGCCTGTGACCC	TPG0000219,	RP11-396A22.4,
19	CTGGTCTCGAACTCCTGACCTCAGGT	TPG0000118,TPG0000907,TPG0001175,	AL133458.1,AC007401.1,AC004985.13,
20	ATTCTCCTGCCTCAGCCTCCTGAGTA	TPG0000118,TPG0001071,TPG0001086,TPG0001313,TPG0001377,	AL133458.1,AP000525.8,KRT42P,RPLP0P2,RP11-725K16.4,
21	CTCGAACTCCTGACCTCAGGTGATC	TPG0000118,TPG0000344,TPG0000005,TPG0001151,TPG0001175,	AL133458.1,POM121L10P,CROCCP3,GTF2H2B,AC004985.13,
22	TGGTCTCGAACTCCTGACCTCAGGT	TPG0000005,TPG0000118,TPG0000907,TPG0000344,TPG0001175,	CROCCP3,AL133458.1,AC007401.1,POM121L10P,AC004985.13,
23	CTCTTGTTGCCCAGGCTGGAGTGCA	TPG0000138,TPG0000005,TPG0000118,TPG0000991,TPG0001000,TPG0001072,TPG0001073,TPG0001112,TPG0001134,TPG0001184,TPG0001194,TPG0001358,TPG0001394,	RP5-886K2.1,CROCCP3,AL133458.1,HLA-P,RP11-347C12.2,HLA-P,HLA-P,AC093642.5,HLA-P,HLA-P,HLA-P,RP11-61L23.2,RP11-175P13.3,

24	CCTCCCAAAGTGCTGGGATTACAGG	TPG0000137,TPG0000118,TPG0000925,TPG0001112,TPG0001328,	AL021707.1,AL133458.1,FER1L4,AC093642.5,RP11-33B1.1,
25	GGCCTCCCAAAGTGCTGGGATTACA	TPG0000138,TPG0000118,TPG0000802,TPG0000344,TPG0000955,TPG0001090,TPG0001112,TPG0001116,TPG0001191,TPG0001313,TPG0001326,TPG0001328,TPG0001343,	RP5-886K2.1,AL133458.1,ZNF286B,POM121L10P,NEURL3,SEPT7P2,AC093642.5,SMPD4P1,GS1-124K5.2,RPLP0P2,RP11-219E24.1,RP11-33B1.1,FTLP10,
26	TCGAACTCCTGACCTCAGGTGATC	TPG0000005,TPG0000118,TPG0000344,TPG0001151,TPG0001175,	CROCCP3,AL133458.1,POM121L10P,GTF2H2B,AC004985.13,
27	TTCTTGCCAAGAGAATTAATGTGC	TPG0000247,TPG0000280,TPG0000749,TPG0000763,TPG0000330,TPG0000328,TPG0000208,TPG0000745,TPG0000314,TPG0000331,	RPL21P2,RP3-522P13.1,RP11-734J24.1,RP11-50D9.1,RP11-473I1.1,RP11-565J7.3,AC005062.3,RP11-439L12.1,RP11-203F10.6,RP11-572P18.1,
28	TACTCAGGAGGCTGAGGCAGGAGAAT	TPG0000155,TPG0001046,TPG0001051,TPG0001068,TPG0001092,TPG0001168,TPG0001354,TPG0001376,TPG0001395,	ZNF204P,ZNF876P,BMS1P1,EGFEM1P,POLR2J4,NACAP1,GUSBP3,AP003068.18,RP11-486A14.1,
29	AGAACAAGGAGCATGTGATTGAGGCC	TPG0000045,TPG0000582,TPG0001214,	RP11-209A2.1,CTC-575D19.1,RP11-62H7.2,
30	TGGTAGATGGAAAACCGGTGAATCTG	TPG0000806,	RP11-284E20.1,
31	TGGCCAACATGGTGAAACCCTGTCTCT	TPG0000124,	AC011385.1,
32	CAGCCTGGCCAACATGGTGAAACCC	TPG0000124,TPG0000785,TPG0000938,TPG0000994,TPG0001051,TPG0001091,TPG0001151,TPG0001199,TPG0001250,TPG0001396,	AC011385.1,AL691432.2,ZDHHC8P1,DNM1P46,BMS1P1,RP11-726G1.1,GTF2H2B,CEACAM22P,GSTTP2,RP3-405J10.2,
33	CCTGTAATCCCAGCACTTTGGGAGGC	TPG0000155,TPG0001238,TPG0001302,	ZNF204P,RP11-231P20.2,ZNF702P,

34	TGTAATCCCAGCACTTTGGGAGGC	TPG0000928,TPG0000953,TPG0001066,TPG0001157,TPG0001193,TPG0001249,TPG0001291,TPG0001313,TPG0001324,TPG0001330,TPG0001375,TPG0001396,	CRYBB2P1,AC002472.8,CATSPER2P1,AC005077.12,FAM92A3,RPS20P22,RP11-568K15.1,RPLP0P2,RP11-515B12.1,RRN3P2,RP11-106J23.2,RP3-405J10.2,
35	ACTGCACTCCAGCCTGGGCAACA	TPG0000539,TPG0000953,TPG0000999,TPG0001103,TPG0001396,	ANXA2P2,AC002472.8,PI4KAP2,PI4KAP1,RP3-405J10.2,
36	GGAGGCTGAGGCAGGAGAATGGC	TPG0000539,TPG0000987,TPG0001030,TPG0001266,TPG0001310,TPG0001354,TPG0001396,	ANXA2P2,AC034193.5,UBE2Q2P1,AC140481.7,INGX,GUSBP3,RP3-405J10.2,
37	TGAAGCTGCAGAACCAACGAGGTGGCC	TPG0000617,TPG0000436,TPG0000682,	FTH1P2,FTH1P16,FTH1P11,
38	AAATGCGGTGGCATCGACAAAAGAAC	TPG0000461,TPG0000111,TPG0000574,	EEF1A1P11,EEF1A1P5,EEF1A1P6,
39	TGGCCAACATGGTGAAACCCTGTCTCTAC	TPG0000785,TPG0001051,TPG0001250,TPG0001306,	AL691432.2,BMS1P1,GSTTP2,RP11-58H15.1,
40	ACCAGCCTGGCCAACATGGTGAAACCC	TPG0000785,TPG0000938,TPG0000994,TPG0001151,TPG0001199,TPG0001396,	AL691432.2,ZDHHC8P1,DNM1P46,GTF2H2B,CEACAM22P,RP3-405J10.2,
41	TGCCTCAGCCTCCCAAGTAGCTGGGA	TPG0000134,TPG0000005,TPG0001010,	AL162458.1,CROCCP3,CIDECP,
42	CTCGAACTCCTGACCTCAGGTGATCC	TPG0000005,TPG0000118,	CROCCP3,AL133458.1,
43	TTGCCCAGGCTGGAGTGCAATGG	TPG0000138,TPG0000005,TPG0000960,TPG0000961,TPG0001000,TPG0001039,TPG0001112,TPG0001358,TPG0001394,	RP5-886K2.1,CROCCP3,PMCHL1,PMCHL2,RP11-347C12.2,ZNF833P,AC093642.5,RP11-61L23.2,RP11-175P13.3,
44	CCATGTTGGTCAGGCTGGTCTCGAACT	TPG0000138,	RP5-886K2.1,
45	TGCCTCAGCCTCCCAAGTAGCTGGGAC	TPG0000138,TPG0000486,TPG0000233,TPG0000327,TPG0001034,TPG0001090,	RP5-886K2.1,DDX50P1,AC007347.1,AC011498.2,ADAM5P,SEPT7P2,

46	TCCCGGGTTCAAGCGATTCTCCTG	TPG0000138,TPG0001108,	RP5-886K2.1,RP11-465B22.3,
47	CACTCTGTCACCCAGGCTGGAGTGCA	TPG0000138,TPG0001124,TPG0001294,	RP5-886K2.1,AFG3L1P,CTB-161A2.3,
48	CTCCCAAAGTGCTGAGATTACAGG	TPG0000138,TPG0001003,TPG0001223,	RP5-886K2.1,HTR7P1,CCT6P3,
49	CTCGGCCTCCCAAAGTGCTGGGATT	TPG0000486,TPG0000138,TPG0000118,TPG0000802,TPG0000344,TPG0000955,TPG0001090,TPG0001116,TPG0001191,TPG0001313,TPG0001326,TPG0001328,TPG0001343,	DDX50P1,RP5-886K2.1,AL133458.1,ZNF286B,POM121L10P,NEURL3,SEPT7P2,SMPD4P1,GS1-124K5.2,RPLP0P2,RP11-219E24.1,RP11-33B1.1,FTLP10,
50	GGCCTCCCAAAGTGCTGGGATT	TPG0000486,TPG0000138,TPG0001090,	DDX50P1,RP5-886K2.1,SEPT7P2,
51	TGGGATTACAGGCATGAGCCACC	TPG0000907,TPG0000344,TPG0000986,TPG0001017,TPG0001070,TPG0001090,TPG0001157,TPG0001396,	AC007401.1,POM121L10P,HERC2P3,RP11-687M24.3,HERC2P9,SEPT7P2,AC005077.12,RP3-405J10.2,
52	CAGACAGAAGGATGTAAAGGATGGAA	TPG0000742,	RP11-234A1.1,
53	ATGAGAGCCAAGTGGAGGAAGA	TPG0000890,TPG0000426,TPG0000887,	AC138123.2,RPL41P1,AC091304.4,
54	TGTGACTTCCTGGAGACTCACTTCCTAG	TPG0000415,	FTLP3,
55	TACTTTGGGAAAGTTGGTATGA	TPG0000259,	RPL27AP,
56	GCTCACTGCAACCTCCGCCTCCC	TPG0000134,TPG0001079,TPG0001090,TPG0001151,TPG0001343,	AL162458.1,AC016629.2,SEPT7P2,GTF2H2B,FTLP10,
57	ATTCTCCTGCCTCAGCCTCC	TPG0000134,TPG0000964,TPG0001071,TPG0001130,TPG0001356,	AL162458.1,RP11-423H2.1,AP000525.8,AC136289.1,TPTE2P1,
58	GCCTAAGAAGAACCGGATTGCC	TPG0000185,	RP11-212P7.1,
59	GGCCTTCCCACACTTTGTCCTCACT	TPG0000344,	POM121L10P,
60	CCCAGGCTAGAGTGCAGTGGTGCA	TPG0000344,	POM121L10P,

61	CTCTGTCACCCAGGCTGGAGTGCA	TPG0000344,TPG0000138,TPG0001124,TPG0001265,TPG0001294,	POM121L10P,RP5-886K2.1,AFG3L1P,ZNF542,CTB-161A2.3,
62	GCCTCCCAAAGTGCTGGGATTACAGG	TPG0000137,TPG0000118,TPG0000925,TPG0001112,TPG0001328,	AL021707.1,AL133458.1,FER1L4,AC093642.5,RP11-33B1.1,
63	ATCCACCCGCCTCAGCCTCCCAAAGT	TPG0000137,TPG0001051,	AL021707.1,BMS1P1,

3.4 pseudoMap (http://pseudomap.mbc.nctu.edu.tw/)

To enable the systematic compilation and updating of these results and additional information, we have developed a database, pseudoMap, capturing various types of information, including sequence data, TPG and cognate annotation, deep sequencing data, RNA-folding structure, gene expression profiles, miRNA annotation and target prediction. As a web-based system, pseudoMap can thoroughly identify TPGs, including TPGs act as a miRNA regulators and TPGs-derived esiRNA-target interactions in humans (**Figure 3.8**). There are two ways to access pseudoMap: by browsing the database content, or by searching for a particular TPG. **Figure 3.9A** displays the interface of output results of the browse gateway. The interface contains general information of TPGs, the relationships of TPG and its cognate gene with miRNA-mediated repression termed as "miRNA Regulator", TPG-derived esiRNA-target interactions named as "esiRNA", and "Expression" showed the gene expression profiles. **Figure 3.9B** provides a detailed view of miRNA regulator, which displays more fine-grained information. Above results indicated the relationships between TPGs and cognate genes by a miRNA decoy mechanism such as that observed by Poliseno *et al.* [48]. And, the "Expression" presents the gene expression profiles of not only distinct TPG and corresponding parental gene, but also TPG referenced by cognate in various experimental conditions (**Figure 3.9C**). Moreover, the view of esiRNA indicates the TPG-derived esiRNAs and graphical display of deep sequencing data (**Figure 3.9D**). The red line represents the TPG as well as the blue line refers to esiRNAs. We also estimate the esiRNA-target interactions (**Figure 3.9E**) and the RNA folding structure of TPG-derived esiRNA (**Figure 3.9F**). All of the results and sequences can be downloaded for further experimental tests. In pseudoMap, we also incorporate the external sources, such as UCSC genome browser

[80] for a genomic view, GeneCards [34] for gene annotation and miRBase [94] for miRNA annotation (**Figure 3.9G**). Additionally, pseudoMap also consists of a tutorial and knowledge of pseudogenes.

In the search gateway, the TPG ID, Ensembl ID, TPG symbol, parental gene symbol and gene description are allowed for further analysis. **Figure 3.10** displays the interface of output results with the search a particular research term. The interface contains the general information of TPG, miRNA regulators, gene expression profiles and TPG-derived esiRNAs.

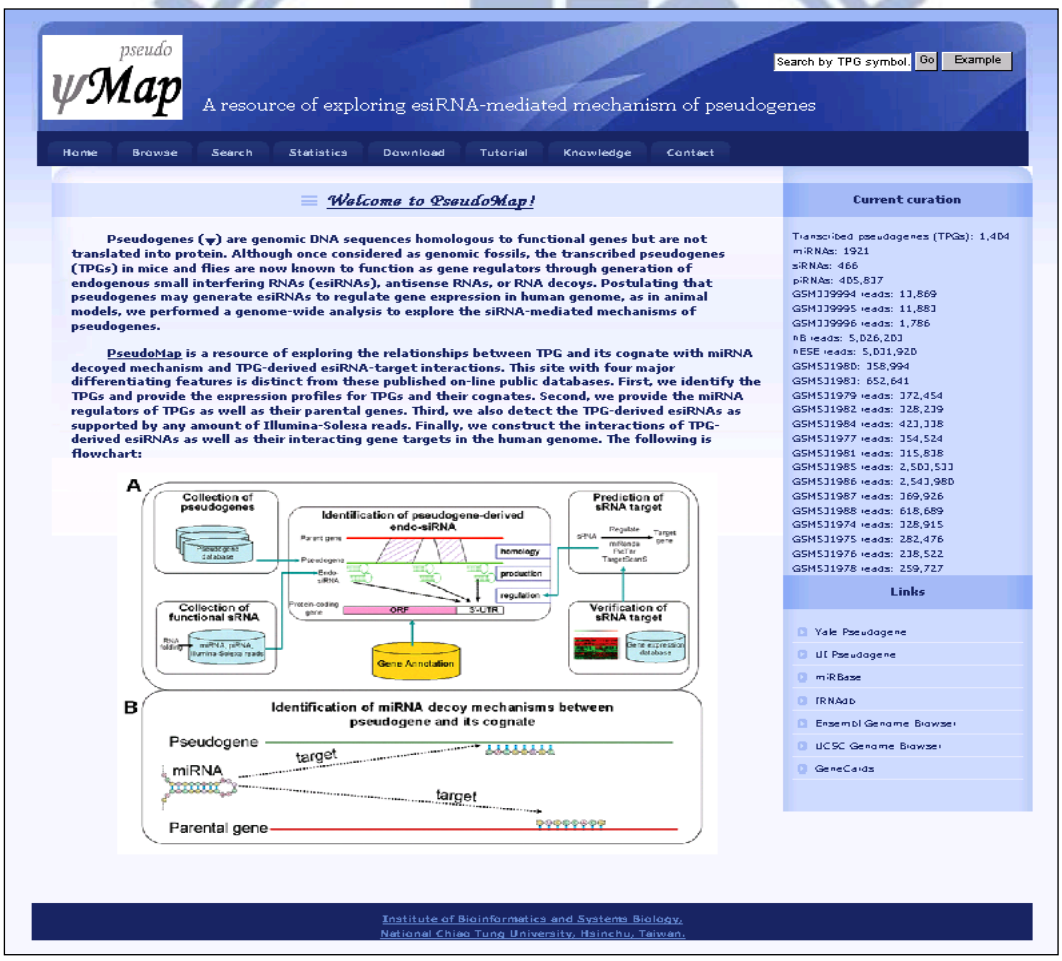


Figure 3.8 pseudoMap: a resource of exploring esiRNA-mediated mechanisms in human transcribed pseudogenes.

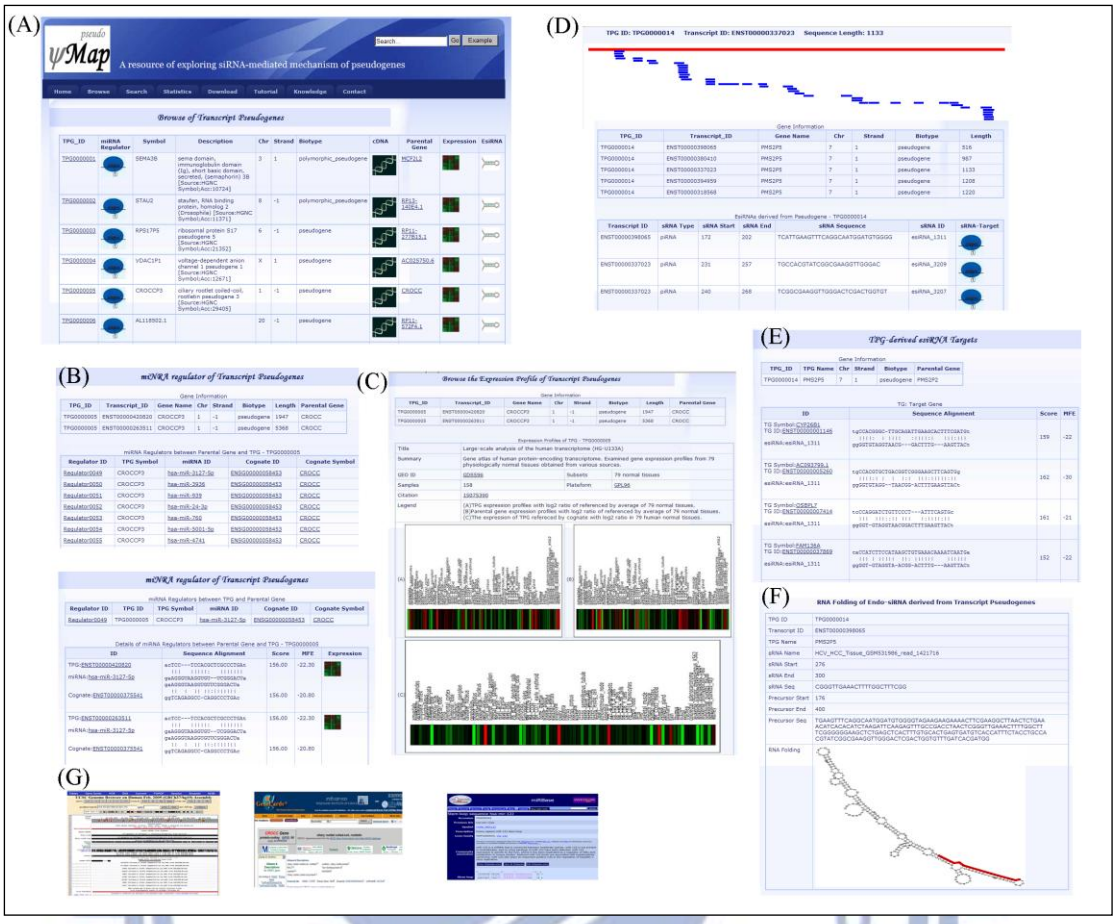


Figure 3.9 Web interface of pseudoMap.

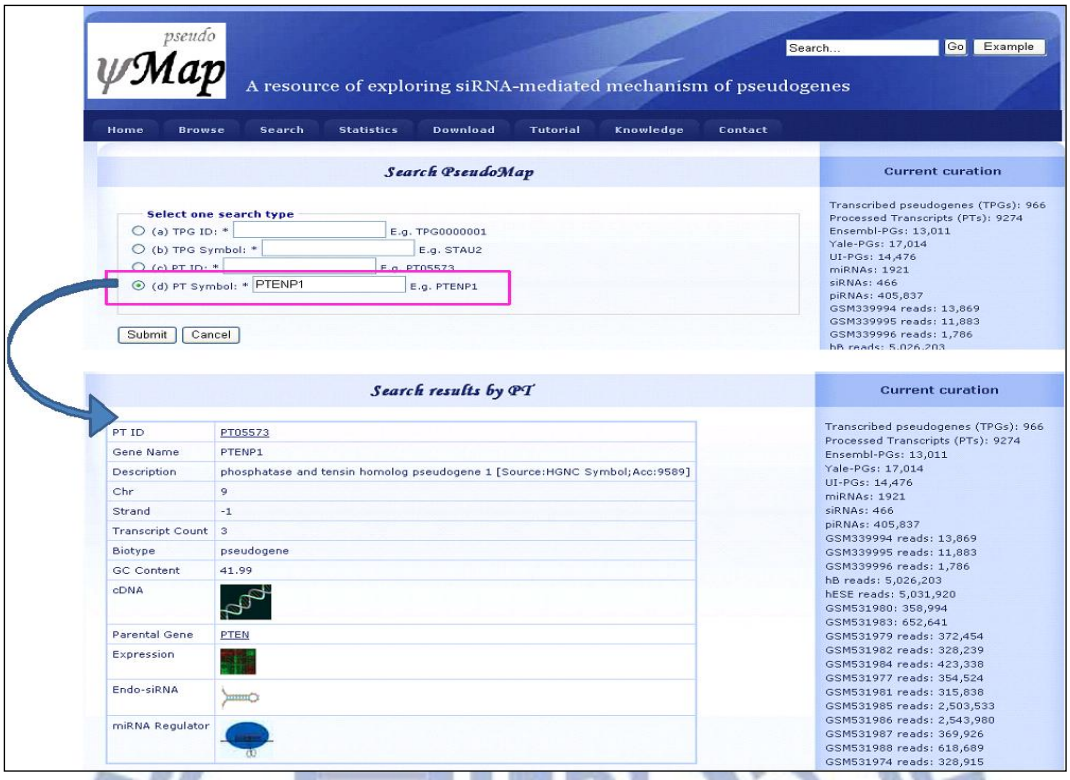


Figure 3.10 Search interface of pseudoMap.

3.4.1 Comparison with other previous databases related to pseudogenes

A few databases have been constructed to explore pseudogenes. In particular, PseudoGene database [64] identifies pseudogenes using various computational methods in genomes; HOPPSIGEN [11] represents the homologous processed pseudogenes shared between the mouse and human genomes that contains location information and potential function; as a web-based system, PseudoGeneQuest [101] identifies novel human pseudogenes based on a user provided protein sequence; in addition, the University of Iowa's UI Pseudogenes website contains human pseudogenes and the candidates for gene conversion [106]. However, these databases focus on automatic detection of pseudogenes by using a variety of homology-based approaches. Our database, pseudoMap, aims to providing comprehensive resource for genome-wide identifying the functions and regulators of human transcribed

pseudogenes. In briefly, there are three major differentiating features from currently public databases of pseudogenes. First, pseudoMap elucidates the relationships of TPG and its cognate gene with miRNA decoys mechanism. Second, to explore the interaction of TPG and its parental gene, pseudoMap provides the gene expression profiles of TPG and its cognate gene in various experimental conditions. Third, pseudoMap curates the TPG-derived esiRNAs, which supported by deep sequencing data, as well as their interacting gene targets in the human genome. **Table 3.3** lists the detailed comparisons of pseudoMap with other previous databases related to pseudogenes.

3.4.2 Applications

PseudoMap provides two major applications, one is the non-coding RNA products of TPGs, as like animal models, may generate esiRNAs to regulate protein-coding genes in humans. In this process, pseudoMap supplies next-generation sequencing data from sRNA libraries to support the candidates of TPG-derived esiRNAs and gene expression profiles to verify the target interactions, respectively. Another application is that both the gene and pseudogene contain miRNA target sites, if the pseudogene competes for the freely available repressor molecules that would be free the gene to reduce the miRNA-mediated repression. Another words, the pseudogene may act as a "miRNA decoy" to release the repression of its cognate gene. pseudoMap provides another insight into the pathway of miRNA-target interactions with TPG-mediated mechanism.

Table 3.3 Comparisons of pseudoMap with currently public databases of pseudogenes.

Supported features	pseudoMap (our database)	PseudoGene database	UI Pseudogene	Hoppsigen
Web interface	http://pseudomap.mbc.nctu.edu.tw/	http://www.pseudogene.org/	https://genome.uiowa.edu/pseudogenes/	http://pbil.univ-lyon1.fr/databases/hoppsigen.html
Description	pseudoMap provides a comprehensive resource for genome-wide identifying the functions and regulators of human pseudogenes.	This site contains a comprehensive database of identified pseudogenes, utilities used to find pseudogenes, various publication data sets and a pseudogene knowledgebase.	This site serves as a repository for all pseudogenes in the human genome and provides a ranked list of human pseudogenes that have been identified as candidates for gene conversion.	Hoppsigen is a nucleic database of homologous processed pseudogenes.
Species supported	Human	Eukaryote and prokaryote	Human	Human
Sequence download	Yes	Yes	Yes	Yes
Pseudogene information	Yes	Yes	Yes	Yes
Parental gene information	Yes	Yes	-	-
Knowledge of pseudogenes	Yes	Yes	-	Yes
miRNA-pseudogene interactions	Yes	-	-	-
miRNA-parental gene interactions	Yes	-	-	-
Gene expression profiles	Yes (both of pseudogene and its parental gene)	-	-	-
Pseudogene-derived siRNAs	Yes	-	-	-
Deep sequencing data for profiling TPG-derived siRNAs	Yes	-	-	-

3.5 Bio-experimental results

In accordant with predicted distinct esiRNAs, RNA structure folding, multiple target sites, gene expression profiles as well as the functions of both the cognate and target gene, the *ψPPM1K*, a partial retrotranscript from *PPM1K* (protein phosphatase, Mg^{2+}/Mn^{2+} dependent, 1K), was first collected for detailed model studies. The detailed experimental results were described following sections.

3.5.1 *ψPPM1K* and its cognate gene

PPM1K (NM_152542.3), located on chromosome 4q22.1, produces a 3,725 nt mRNA encoding a mitochondrial matrix serine/threonine protein phosphatase shown to regulate the membrane permeability transition pore (MPTP) essential for cell survival and organ development [107]. Pseudogene *PPM1K (ψPPM1K)* transcription is supported by mRNA and EST evidences and also located on chromosome 4q22.1 (**Figure 3.11**). *ψPPM1K*, 474 bp in length, is partially retrotranscribed from *PPM1K* (**Figure 3.12a**) and bases 155-456 of the *ψPPM1K* transcript show > 79% similarity to the antisense-strand of *PPM1K* (**Figure 3.12b**). Multiple alignments of the 5'- and 3'-ends of *ψPPM1K* DNA sequences in 44 vertebrate species revealed a high degree of conservation among rhesus, mouse, dog and elephant. The *ψPPM1K* sequence also contains SINE and Alu repeat elements (**Figure 3.11**).

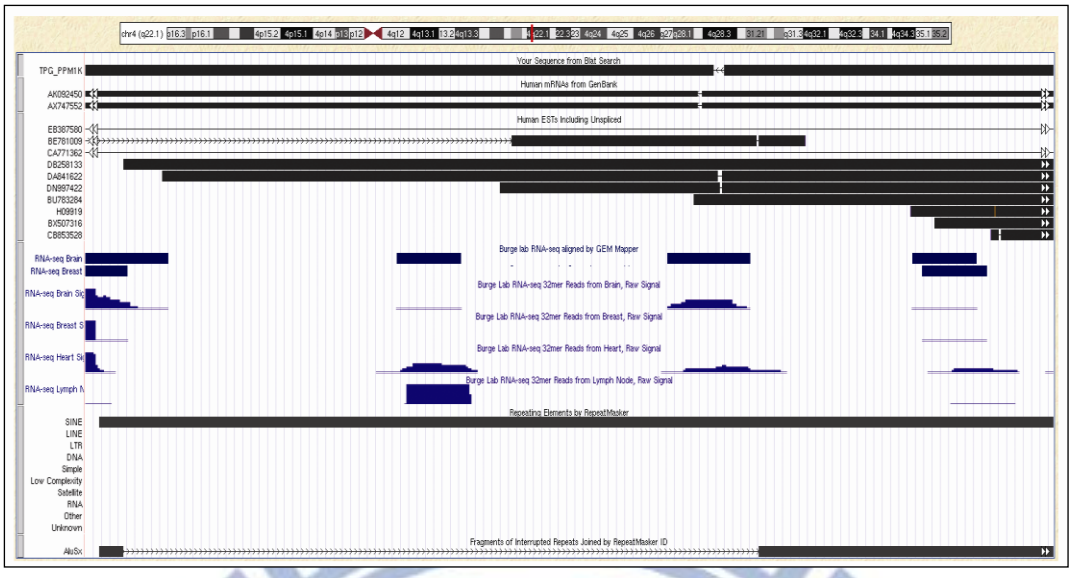


Figure 3.11 The genomic view of *ψPPM1K*.

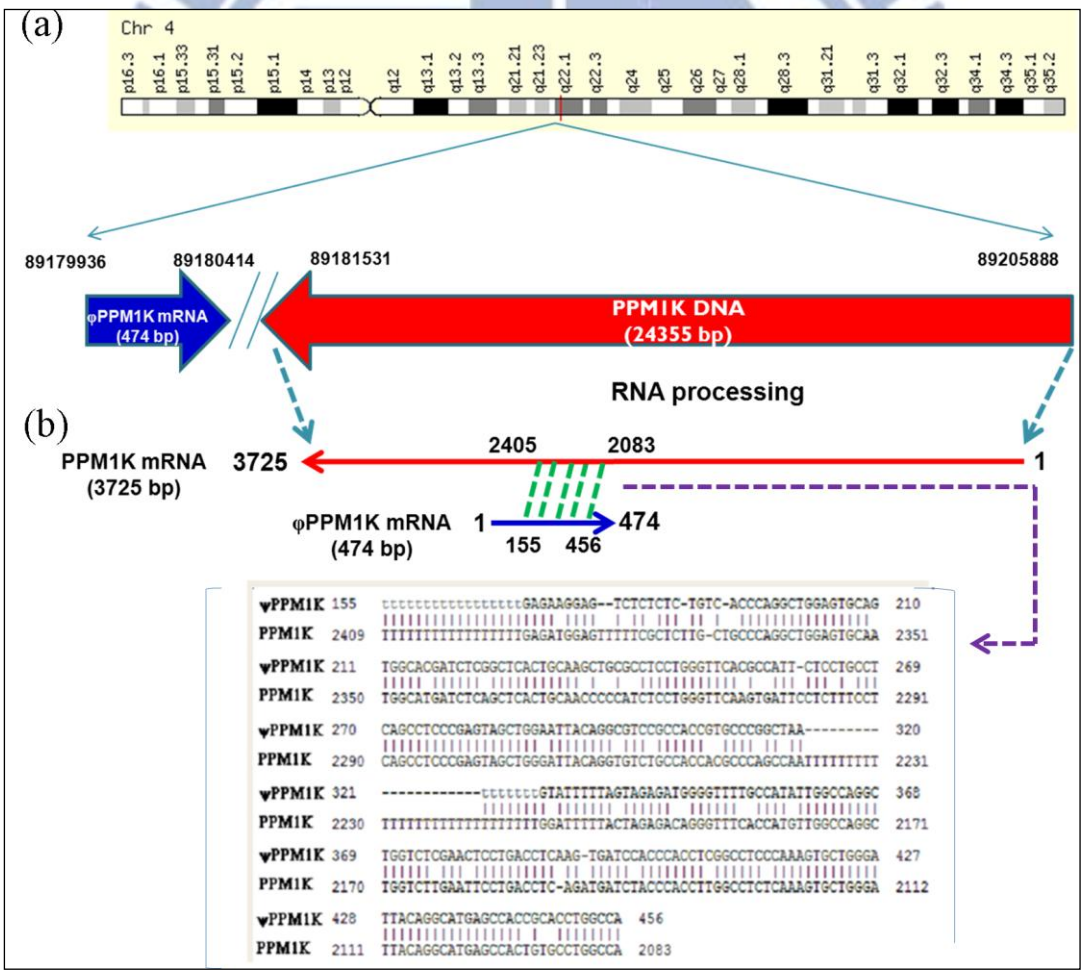


Figure 3.12 Schematic representation of *ψPPM1K* and its parental gene, PPM1K. (a) *ψPPM1K* is located on chromosome 4 proximal to *PPM1K*. (b) Alignment of *ψPPM1K* and its cognate gene *PPM1K*.

3.5.2 Determination of *ψPPM1K*-derived esiRNAs and their candidate target genes

To predict *ψPPM1K*-derived esiRNAs, we aligned the sequences of fsRNAs with length 18-40 nt, obtained from fRNAdb, to the *ψPPM1K* sequence and found 76 positive examples that also matched publicly available deep sequencing data from various sRNA libraries. Positives can be grouped into three blocks, indicating precursor esiRNA1 as being located in the 24-144 nt region (depicted green), precursor esiRNA2 in the 170-273 nt region (depicted red) and precursor esiRNA3 in the 328-455 nt region (depicted pink), respectively (**Figure 3.13a**). The sequences of precursor esiRNAs are shown in **Figure 3.13b** and their respective mature esiRNAs are given in color. We were mostly interested in the *ψPPM1K*-unique esiRNAs, esiRNA1 and esiRNA2, and therefore did not study precursor esiRNA3 further, as it is also present in the cognate gene *PPM1K* (**Figure 3.13c**). **Figure 3.13d** shows that the precursor of esiRNA1 (esiRNA1 depicted in green) folds into a hairpin structure with a MFE value of -40.4 by Mfold [77]. Subsequently, we performed a genome-wide search for esiRNA1-target interactions and obtained 57 candidate target genes with scores ≥ 200 (**Table 3.4**). These targets could be classified into functional categories involving nucleotide binding, and transferase and serine/threonine protein kinase enzyme activities (**Figure 3.14a-c**). Analyses of transcription factor binding sites (TFBS) in these target genes revealed that they shared motifs for RSRFC4, GFI1, NF1, GATA1 and ELK1 (**Figure 3.14d**). As illustrated in Figure **Figure 3.13e**, the predicted *ψPPM1K*-specific esiRNAs are expected to regulate cognate gene *PPM1K* and target gene *NEK8* (Never in mitosis A-related kinase 8) through association with multiple target sites. We also observed that *ψPPM1K*-specific esiRNAs not only

complemented the 3'-UTR but also the coding region of *NEK8*. Additionally, gene expression profiles (GSE6222) also revealed that *ψPPM1K* is expressed at lower levels than *NEK8* and *PPM1K* in HCC tissues and cells in comparison to normal tissues (**Figure 3.13f**).

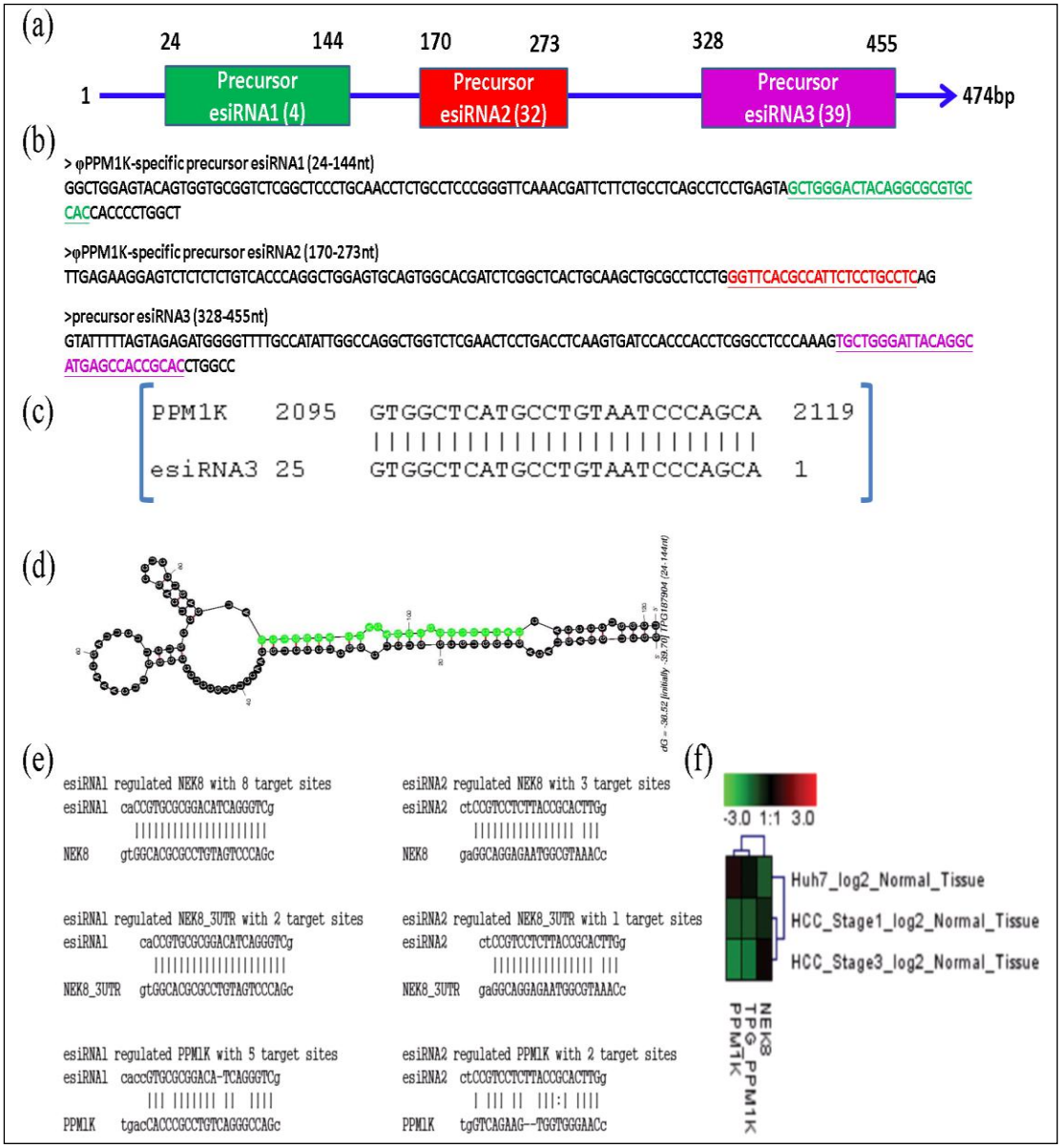


Figure 3.13 Candidates *ψPPM1K*-derived esiRNAs and their targets. (a) Location and read counts of transcribed *ψPPM1K* RNA from sRNA deep sequencing data. (b) Sequences of precursor esiRNAs, with esiRNAs shown in color. (c) The alignment of esiRNA3 mapping to *PPM1K* gene. (d) Hairpin structure prediciton of precursor esiRNA1. (e) Matches of esiRNA1 and esiRNA2 sequences with target gene *NEK8* and parental gene *PPM1K*. (f) Expression profiles of *ψPPM1K*, *PPM1K* and *NEK8* in HCC tissues/cells.

Table 3.4 The top 20 candidates, that score more than 200, of predicted esiRNA1 targeted genes by modification of TargetScan, RNAhybrid and miRanda.

Ensembl_ID	Symbol	Description
ENSG00000124608	AARS2	Probable alanyl-tRNA synthetase, mitochondrial Precursor (EC 6.1.1.7)
ENSG00000161533	ACOX1	Peroxisomal acyl-coenzyme A oxidase 1 (AOX)(EC 1.3.3.6)
ENSG00000078295	ADCY2	Adenylate cyclase type 2 (EC 4.6.1.1)
ENSG00000152213	ARL11	ADP-ribosylation factor-like protein 11 (ADP-ribosylation factor-like tumor suppressor protein 1)
ENSG00000204217	BMPR2	Bone morphogenetic protein receptor type-2 Precursor (EC 2.7.11.30)
ENSG00000008300	CELSR3	Solute carrier family 26 member 6 (Pendrin-like protein 1)
ENSG00000006634	DBF4	Protein DBF4 homolog A (Activator of S phase kinase)
ENSG00000160602	NEK8	Serine/threonine-protein kinase Nek8 (EC 2.7.11.1)
ENSG00000137992	DBT	Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial Precursor (EC 2.3.1.168)
ENSG00000154144	TBRG1	Transforming growth factor beta regulator 1 (Nuclear interactor of ARF and Mdm2)
ENSG00000166938	DIS3L	DIS3-like exonuclease 1 (EC 3.1.13.-)
ENSG00000116199	FAM20B	Protein FAM20B Precursor
ENSG00000143458	GABPB2	GA-binding protein subunit beta-2 (GABP subunit beta-2)(GABPB-2)
ENSG00000162654	GBP4	Guanylate-binding protein 4 (Guanine nucleotide-binding protein 4)
ENSG00000159921	GNE	Bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase (UDP-GlcNAc-2-epimerase/ManAc kinase)
ENSG00000177885	GRB2	Growth factor receptor-bound protein 2 (Adapter protein GRB2)(SH2/SH3 adapter GRB2)(Protein Ash)
ENSG00000177602	GSG2	Serine/threonine-protein kinase haspin (EC 2.7.11.1)
ENSG00000214801	GUSBL1	Beta-glucuronidase-like protein 1 (SMA3-like protein)
ENSG00000204592	HLA-E	major histocompatibility complex, class I, E precursor
ENSG00000135100	HNF1A	Hepatocyte nuclear factor 1-alpha (HNF-1A)

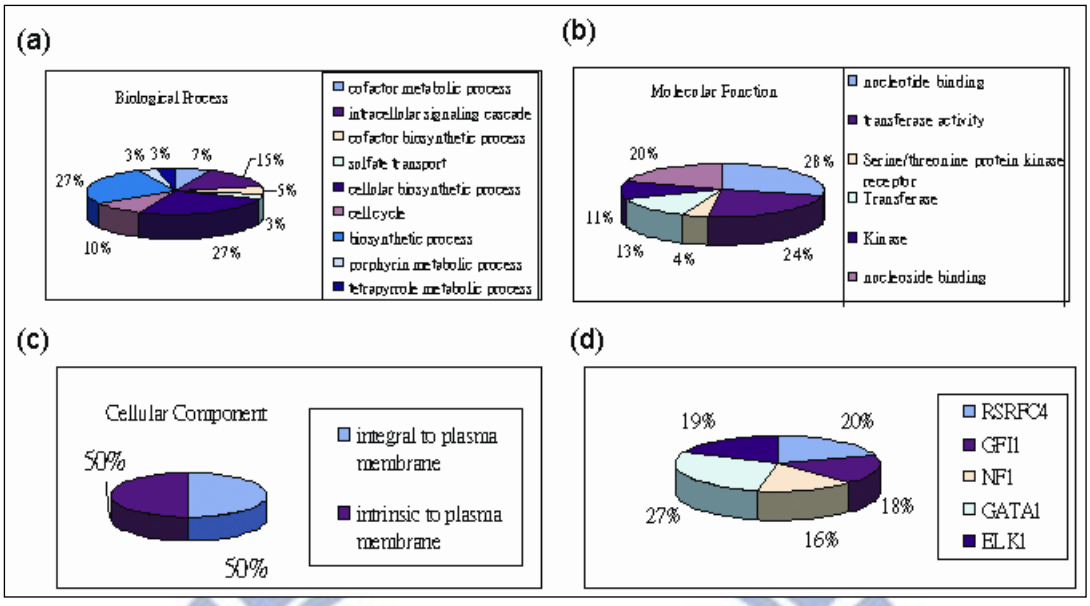


Figure 3.14 Characterization of esiRNA1-targeted genes. The gene-ontology was categorized according to biological process (a), molecular function (b) and cellular component (c) to determine the common cellular functions affected by esiRNA1. (d) Common transcriptional factor binding sites (TFBS) of target genes.

3.5.3 Expression of *PPM1K*, precursors of *ψPPM1K*-derived esiRNAs and *NEK8* in HCC patient samples

To compare the expression of *PPM1K*, *ψPPM1K*-derived esiRNAs and target gene, *NEK8*, in clinical samples, we examined 41 HCC surgical specimens together with matched non-tumor tissues by RT-qPCR. **Figure 3.15a** shows that the HCC tumor tissues expressed significantly lower levels of both *PPM1K* (0.76-fold; *P*=0.001) and precursors of *ψPPM1K*-derived esiRNA1 (24-144 nt; 0.59-fold; *P*=0.007) and esiRNA2 (170-273 nt; 0.61-fold; *P*=0.007), than paired non-tumor tissues, whereas expression of the target gene *NEK8* was elevated (1.29-fold) in the tumor samples. These results are consistent with those described above (**Figure 3.13f**).

3.5.1 Decreased cell growth and clonogenic activity in *ψPPM1K*-transduced Huh-7 cells

RT-qPCR analyses (**Figure 3.15b**) showed that *PPMIK* mRNA levels are higher in Huh-7 than in HepG2 cells and that the expression of *PPM1K* mRNA is higher than that of *PPMIK* in both cell lines (n=7). To examine the effect of overexpression of *ψPPM1K*, Huh-7 and HepG2 cells were transfected with *ψPPM1K*-harboring recombinant plasmid vector and four stable transfectant clones, TPG1, TPG2, TPG7 (for Huh-7) and TPG (for HepG2) were isolated; a control clone transfected with empty vector was also produced (**Figure 3.16a**). Because of the effects of *NEK8*/*NEK9* on cell cycle regulation [108, 109] and their relatively high expression in human breast cancer [109], we examined cell growth behavior of *ψPPM1K*-transfected versus control Huh-7 cells. A cell proliferation assay showed that *ψPPM1K*-overexpressing cell lines have a slower proliferation rate than the vector control cells (TPG7 cells: P=0.036 for Day 2, P=0.018 for Day 4; TPG1 cells: P=0.045 for Day 4) (**Figure 3.16b**). Rates of colony expansion after plating also indicated slower proliferation of the three transfected Huh-7 TPG cell clones in comparison to the control clone (**Figure 3.16c**). Furthermore, assessment of colony formation in soft agar suspension cultures under normoxic and hypoxic conditions (**Figure 3.16d**) showed that the three stably transfected *ψPPM1K*-expressing cell clones all consistently exhibited less clonogenic activity than the vector control (TPG1 cells: P=1.54E-06, TPG2 cells: P=0.0001, TPG7 cells: P=0.013 for normoxic conditions; TPG1cells: P=2.22E-05, TPG2 cells: P=5.71E-05, TPG7 cells: P=0.0011 for hypoxic conditions). These effects on cell proliferation suggest that *ψPPM1K* might act as a tumor suppressor in Huh-7 cells.

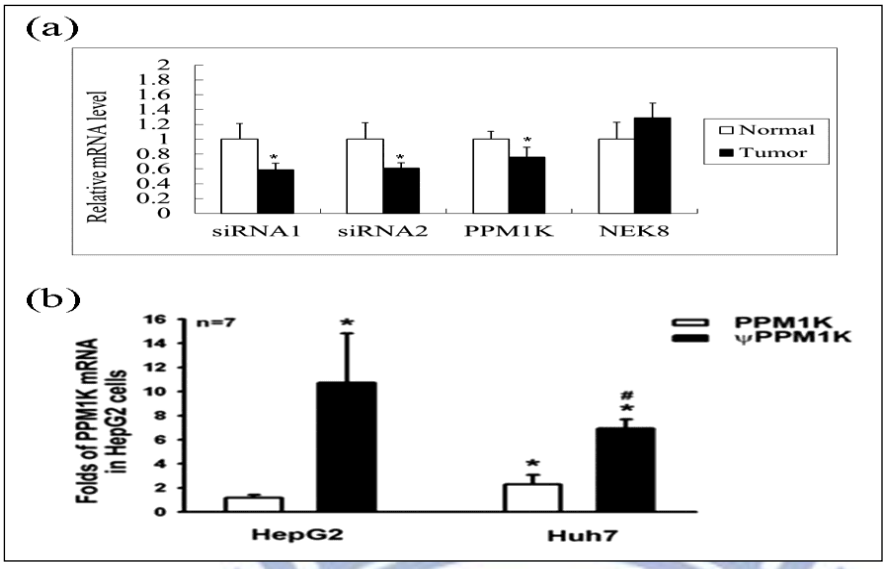


Figure 3.15 Expression patterns of HCC tissues and cell lines. (a) RT-qPCR of two esiRNA precursors, *PPM1K* and *NEK8*, in paired HCC tissues (b) RNA levels of *PPM1K* and *ψPPM1K* in HepG2 and Huh-7 cells.

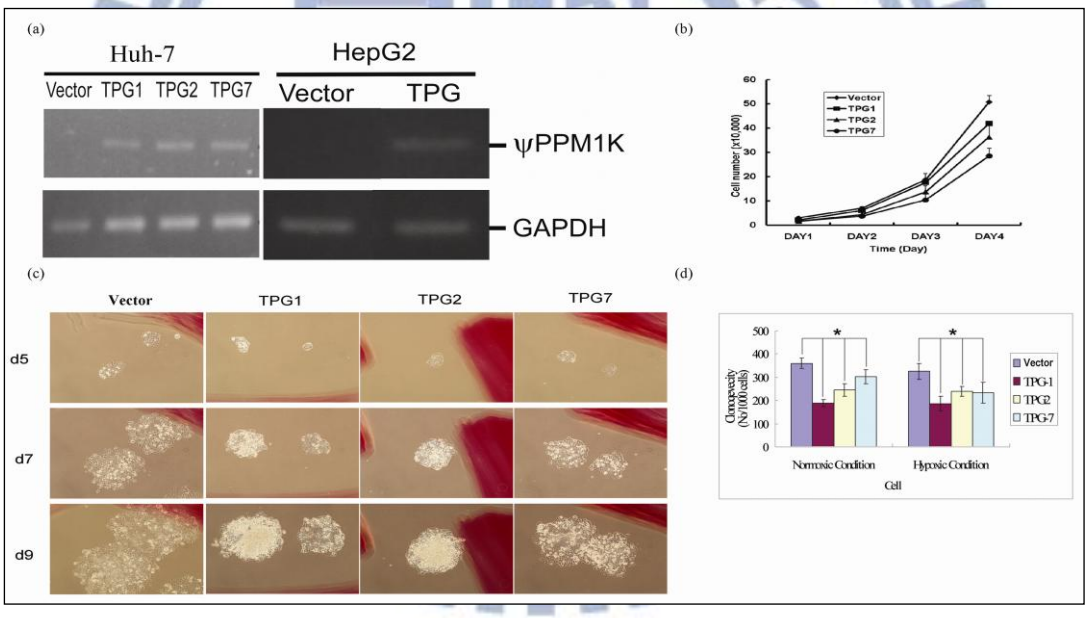


Figure 3.16 Effect of overexpressed *ψPPM1K* on cell growth and clonogenic activity in transfected Huh-7 cell clones. (a) HCC line Huh-7 and HepG2 cells were transfected with *ψPPM1K*-expressing recombinant plasmid to isolate stably transfected cell clones. (b) All three *ψPPM1K*-expressing cell lines have a slower proliferation rate than the vector control cell line. (c) Serial photographs of the same colonies at day 5, day 7 and day 9 showing the two-dimensional growth of mock2, TPG1, TPG2 and TPG7 transfected Huh-7 clones on plastic culture dishes. (d) Clonogenic activity of mock2, TPG1, TPG2 and TPG7.

3.5.2 Expression of *NEK8* in *ψPPM1K*-transfected HCC cells

The *in silico* results indicated that *ψPPM1K*-derived esiRNAs may regulate protein-coding genes, such as *NEK8*, *TBRG1* and *BMPR2* (**Table 3.4**). Respectively, *NEK8* encodes a member of the serine/threonine protein kinase family related to NIMA of *Aspergillus nidulans* and is overexpressed in human breast tumor [109]; *TBRG1* acts as a growth inhibitor that activates *TP53* to cause G1 arrest and collaborates with *CDKN2A* to restrict proliferation, but does not require either protein to inhibit DNA synthesis [110, 111]; while *BMPR2* encodes a member of the BMP receptor family of transmembrane serine/threonine kinases [112]. To examine the effect of overexpression of *ψPPM1K* in Huh-7 and HepG2 stable transfectant clones, the mRNA levels of *NEK8*, *TBRG1*, and *BMPR2* were determined by RT-qPCR. The results showed that *NEK8* expression was reduced in all *ψPPM1K*-expressing HCC cell lines relative to the vector control cells (P=0.033 in TPG7 Huh-7 cells; P=0.047 in TPG HepG2 cells) (**Figure 3.17a-b**). Thus, *ψPPM1K* down-regulated *NEK8* in both Huh-7 and HepG2 cell lines.

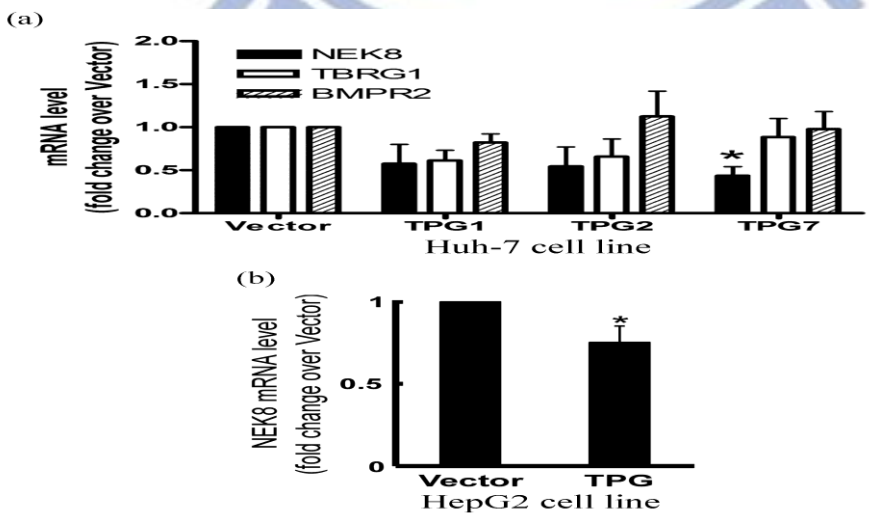


Figure 3.17 Expression of target genes in HCC cell clones.

3.5.3 Expression of *ψPPM1K*-derived esiRNAs

Before verifying down-regulation of *NEK8* by *ψPPM1K*-derived esiRNAs, we first demonstrated esiRNA expression in the Huh-7 cell line and TPG-expressing cell clones derived from it. The Northern blots showed that both of esiRNA1 and esiRNAs2 were expressed in Huh-7 cells and that the expression of esiRNA1 mRNA is higher than that of esiRNA2 (**Figure 3.18a**). Next, a TaqMan MicroRNA Assay (Applied Biosystems) was used to show that the esiRNA1 level in TPG7 cells was significantly higher than that in the vector control cells (2-fold; $P < 0.05$) (**Figure 3.18b**). These results suggest that esiRNA1 is expressed in Huh-7 cells, but is produced at higher levels when *ψPPM1K* is overexpressed from a recombinant plasmid.

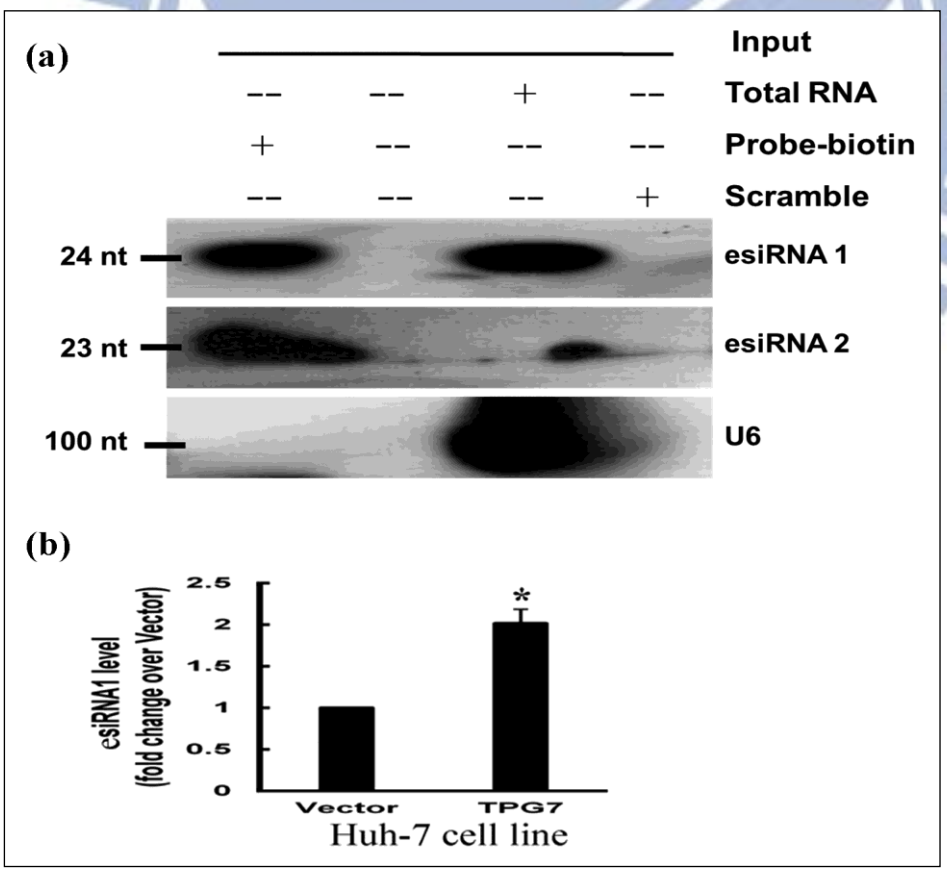


Figure 3.18 Expression of *ψPPM1K*-dervied esiRNAs.

3.5.4 FISH localization of *ψPPM1K*-derived esiRNAs

FISH results showed that *ψPPM1K*-derived esiRNA1 was located in the cytoplasm of interphase cells (**Figure 3.19, upper panels**) and that esiRNA2 was mainly present the in nuclear area of both interphase and prometaphase cells (**Figure 3.19, lower panels**), suggesting that esiRNA1, like miRNAs, regulates its targets in the cytoplasm while esiRNA2 might not.

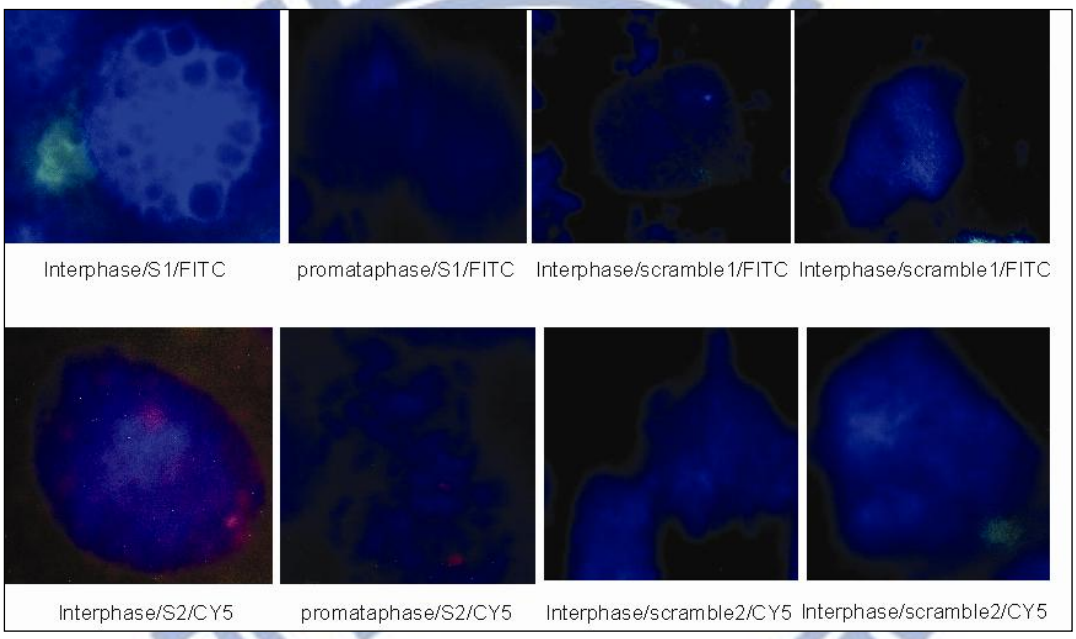


Figure 3.19 Localization of pseudogene-derived esiRNAs by FISH analysis.

3.5.5 Expression of *NEK8* in Huh-7 cells transfected with synthetic siRNA1

To verify the effect of *ψPPM1K*-dervied esiRNA1 on *NEK8* expression, we transfected Huh-7 cells with synthetic siRNA1 identical in sequence to esiRNA1 (**Figure 3.20a**) and determined *NEK8* mRNA levels at 48 h. *NEK8* mRNA levels were significantly lower in Huh-7 cells transfected with the synthetic siRNA1 than those

transfected with negative control siRNAs (*P*=0.001), implying that synthetic siRNA homologous to the esiRNA1 can directly down-regulate *NEK8* gene expression in Huh-7 cell lines (Figure **Figure 3.20b**).

3.5.6 *NEK8* expression and cell growth activation in esiRNA1-deleted *ψPPM1K*-overexpressing cells

To establish that *NEK8* is the major target of *ψPPM1K* through esiRNA1, we constructed an esiRNA1-deletion mutant *ψPPM1K*-expression plasmid to transfect HepG2 cells (named mTPG cells) and determined *NEK8* expression levels. The results showed that *NEK8* was down-regulated in TPG HepG2 cells but not in mTPG cells (**Figure 3.20c**). Additionally, we also co-expressed *NEK8* in *ψPPM1K*-transfected cells and carried out the cell proliferation assay used previously. This demonstrated that *NEK8* can counteract the inhibitory effects of *ψPPM1K* on HCC cell proliferation (**Figure 3.20d**).

3.5.7 *ψPPM1K*-derived esiRNA1 down-regulation of *PPM1K*

ψPPM1K-derived esiRNA1 is also predicted to target *PPM1K* (**Figure 3.13d**) and therefore we examined the expression of *PPM1K* in mTPG HepG2 cells. *PPM1K* mRNA levels were significantly lower in TPG HepG2 cells than in control cells (*P*=0.001), but were unaffected in mTPG HepG2 cells, implying that *ψPPM1K*-derived esiRNA1 can directly down-regulate *PPM1K* expression (**Figure 3.21a**).

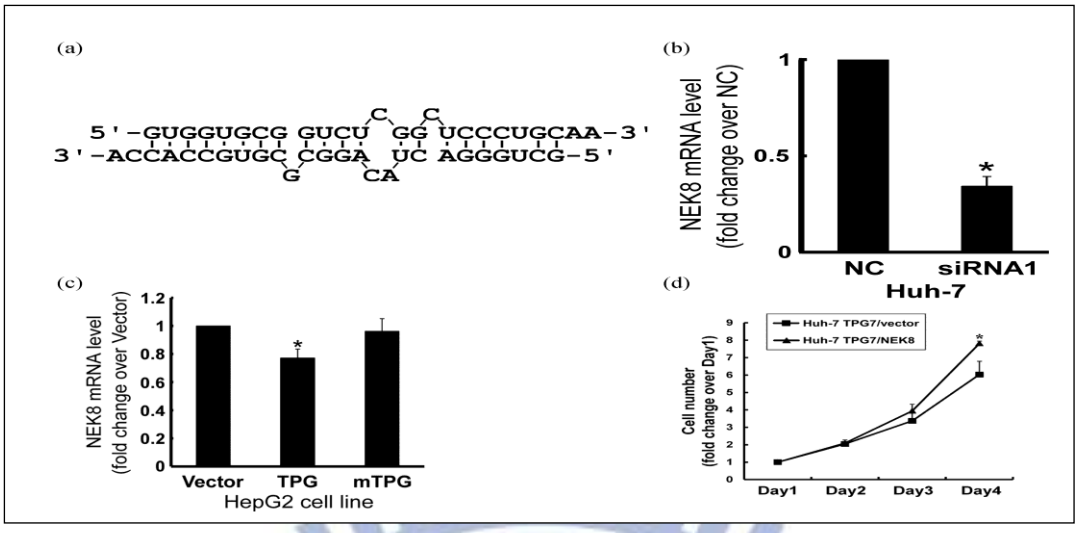


Figure 3.20 Expression of *NEK8* in Huh-7 cells transfected with synthetic siRNA1. (a) The sequence of synthetic siRNA1. (b) Huh-7 cells were transfected with synthetic siRNA1 (c) Expression of *NEK8* in an esiRNA1-deletion mutant cell line. (d) Growth of Huh-7 TPG7 cells transfected with either *NEK8*-overexpressing plasmid or empty vector analyzed by cell proliferation assay.

3.5.8 *ψPPM1K* alters mitochondrial functions in *ψPPM1K*-transduced Huh-7 cells

It has been shown that, following cellular uptake of rhodamine 123 (Rh123) and specific accumulation of this fluorescent dye in mitochondria, human cancer cells tend to retain significantly more Rh123 than normal cells [104, 105]. It is thought that this difference in the uptake and retention of Rh123 may reflect MPTP activity [103]. As the phosphatase encoded by *PPM1K*/*PP2Cm* is located in the mitochondrial matrix and regulates MPTP activity, the effect of *ψPPM1K* overexpression is of interest. We found that rates of Rh123 uptake were higher for TPG1 and TPG2, while remaining approximately the same as the control for TPG7 (**Figure 3.21b** and **Figure 3.22**). The mitochondrial Rh123 retention assay showed significantly reduced Rh123

retention in TPG7 cells than in control vector cells (**Figure 3.21c**), suggesting faster Rh123 release or relatively normalized MPTP stability in TPG7 cells. However, no such change in Rh123 retention was observed in TPG1 or TPG2 cells in comparison to control cells.

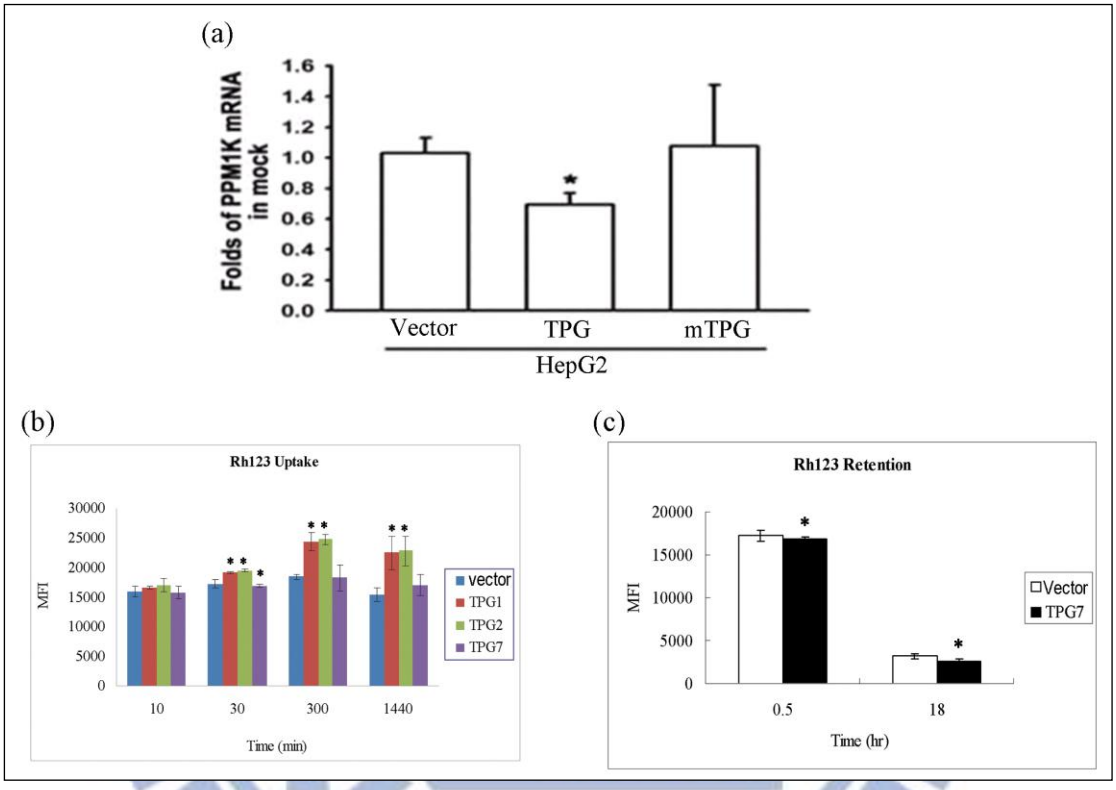


Figure 3.21 *ψPPM1K* alters *PPM1K* expression and mitochondrial function.

3.5.9 miRNA regulation of *ψPPM1K* and *PPM1K*

Considering the recently reported novel findings that pseudogenes *PTENP1* and *KRAS1P* can de-repress their cognate genes by a miRNA decoy mechanism [48], we took a similar approach to investigate miRNA-target interactions (MTI) in *ψPPM1K* and *PPM1K*. The *in silico* results revealed many miRNAs that potentially interact with multiple possible target sites in both *ψPPM1K* and *PPM1K* (**Table 3.5**).

Furthermore, seed matches for *PPM1K*-targeting miRNAs were perfectly conserved in *ψPPM1K* (**Figure 3.23a**). To verify this prediction, we assessed the expression levels of *PPM1K* and *ψPPM1K* in miRNA-transfected TPG7 Huh-7 cells and vector control cells. The results showed that has-miR-3174 (miR-3174) down-regulated *PPM1K* in both cell lines, while *ψPPM1K* was also down-regulated in vector control cells compared to a negative control sRNA (siCon) (**Figure 3.23b**).

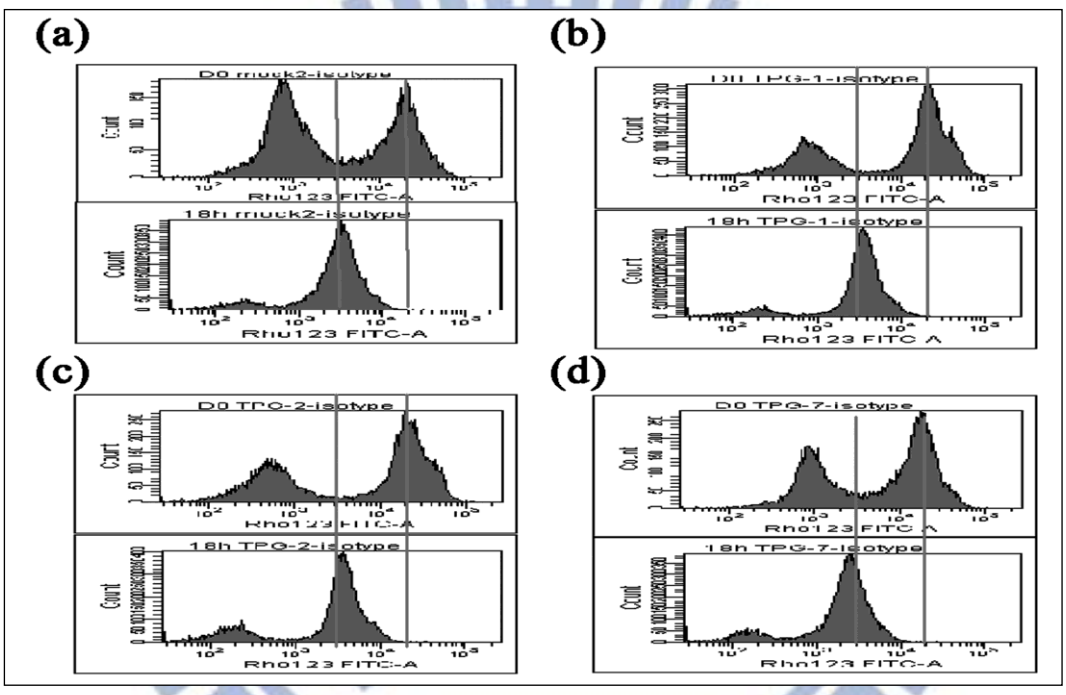


Figure 3.22 FACS analysis of mitochondrial Rh123 uptake and release from transfected mock2 (a), TPG1 (b), TPG2 (c), TPG7 (d) transfected HCC Huh-7 clones.

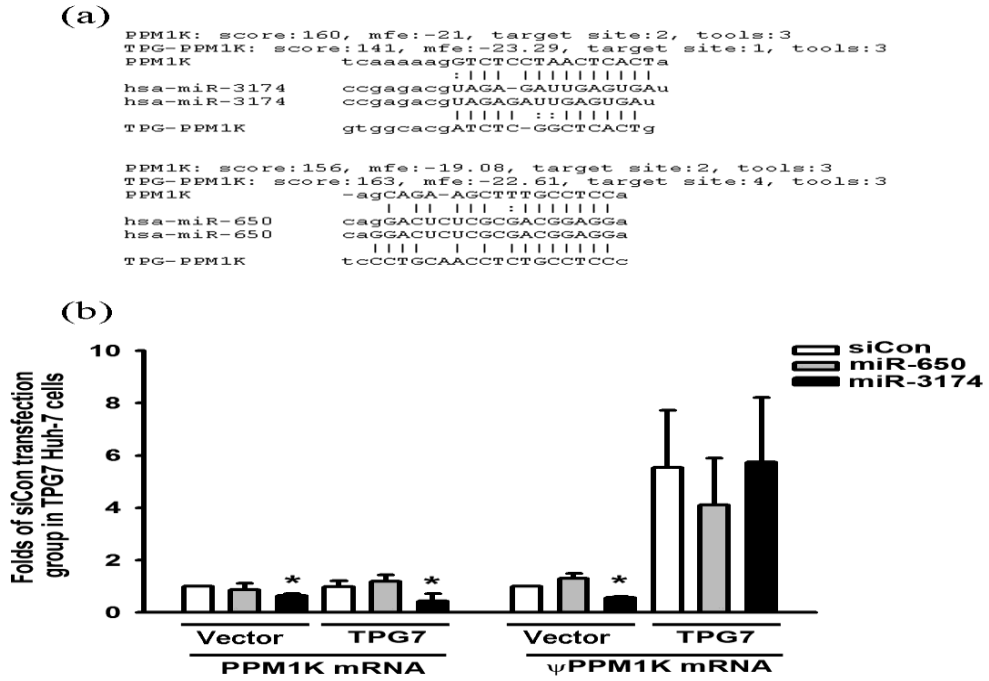


Figure 3.23 miRNA regulation of *PPM1K* and *ψPPM1K*. (a) Alignments of miRNAs target with *PPM1K* and *ψPPM1K*. (b) miR-3174 down-regulation of PPM1K and *ψPPM1K*.

Table 3.5 The miRNA target sites in *ψPPM1K* and its parental genes.

miRNA	target sites of TPG-PPM1K	target sites of PPM1K_CDS	target sites of PPM1K_3UTR
hsa-miR-4685-5p	3	10	1
hsa-miR-638	3	5	1
hsa-miR-5192	3	8	1
hsa-miR-1273d	2	4	1
hsa-miR-5189	4	14	1
hsa-miR-4722-5p	5	8	1
hsa-miR-4656	4	8	1
hsa-miR-4656	4	8	1
hsa-miR-4656	4	8	1
hsa-miR-892b	1	8	2
hsa-miR-1972	1	12	1
hsa-miR-4459	6	11	1
hsa-miR-3192	5	12	1
hsa-miR-658	4	10	1

hsa-miR-1202	6	8	2
hsa-miR-663b	2	6	1
hsa-miR-4741	1	7	1
hsa-miR-3137	2	10	2
hsa-miR-3138	2	12	2
hsa-miR-3187-5p	5	10	1
hsa-miR-4707-5p	1	7	1
hsa-miR-92a-3p	2	5	2
hsa-miR-5195-3p	3	6	1
hsa-miR-5090	3	9	1
hsa-miR-4454	2	3	1
hsa-miR-3619-3p	1	6	1
hsa-miR-3064-5p	2	6	1
hsa-miR-1301	4	10	2
hsa-let-7b-5p	4	6	1
hsa-miR-1183	3	17	1
hsa-let-7c	3	7	1
hsa-miR-1915-3p	2	8	1
hsa-miR-4449	2	9	2
hsa-miR-4682	1	6	2
hsa-miR-1912	1	5	2
hsa-miR-25-3p	2	4	1
hsa-miR-4257	3	3	1
hsa-let-7d-5p	3	6	1
hsa-miR-323a-5p	1	6	1
hsa-miR-885-3p	3	7	1
hsa-miR-193a-3p	1	4	1
hsa-miR-4758-5p	1	8	1
hsa-miR-106b-3p	1	7	1
hsa-miR-5572	3	8	1
hsa-miR-92b-3p	2	3	2
hsa-miR-4538	3	5	1
hsa-miR-3934	2	6	1
hsa-miR-1273g-5p	4	6	1
hsa-miR-4687-3p	2	12	3
hsa-miR-711	1	4	1
hsa-miR-4479	1	9	1

hsa-miR-3132	3	7	1
hsa-miR-4689	2	9	1
hsa-miR-510	3	6	1
hsa-miR-5684	2	3	1
hsa-miR-5091	1	2	1
hsa-miR-298	3	11	2
hsa-miR-1271-3p	5	11	1
hsa-miR-3170	1	3	1
hsa-miR-3944-5p	1	3	1
hsa-miR-4299	2	7	1
hsa-miR-4714-3p	1	3	1
hsa-miR-1229	1	8	1
hsa-miR-4436b-3p	4	6	1
hsa-miR-941	1	6	1
hsa-miR-4793-3p	3	7	1
hsa-let-7e-5p	4	9	2
hsa-let-7a-5p	3	5	1
hsa-miR-4751	1	6	1
hsa-miR-2467-3p	5	5	1
hsa-miR-3174	1	5	1
hsa-miR-5190	1	7	2
hsa-miR-138-5p	1	12	1
hsa-miR-365a-5p	1	6	2
hsa-miR-3156-5p	1	4	1
hsa-miR-194-3p	1	10	2
hsa-miR-143-5p	2	11	2
hsa-miR-1538	3	9	1
hsa-miR-876-3p	1	3	1
hsa-miR-149-5p	1	10	1
hsa-miR-3135b	1	12	1
hsa-miR-4433-3p	4	4	2
hsa-miR-1293	4	7	1
hsa-miR-4745-3p	2	8	1
hsa-miR-135a-3p	1	4	1
hsa-miR-2355-3p	1	8	1
hsa-miR-4266	3	2	1
hsa-miR-4708-3p	1	5	1

hsa-miR-5093	1	6	1
hsa-miR-4675	1	13	2
hsa-miR-4496	1	7	2
hsa-miR-5008-5p	1	6	1
hsa-miR-766-3p	2	5	1
hsa-miR-422a	1	6	1
hsa-miR-508-5p	2	9	1
hsa-miR-1265	1	6	1
hsa-miR-644b-3p	1	4	1
hsa-miR-423-3p	1	7	1
hsa-miR-3605-5p	1	6	1
hsa-miR-378c	3	10	3
hsa-miR-770-5p	2	8	1
hsa-miR-1224-5p	1	6	1
hsa-miR-4474-3p	1	9	1
hsa-miR-4322	1	6	1
hsa-miR-4489	1	5	1
hsa-miR-4518	5	9	1
hsa-miR-3154	2	10	1
hsa-miR-4446-3p	2	9	2
hsa-miR-3139	1	7	2
hsa-miR-4440	1	5	1
hsa-miR-602	2	5	1
hsa-miR-1182	1	9	1
hsa-miR-98	2	4	1
hsa-let-7f-5p	3	3	1
hsa-miR-219-1-3p	2	4	1
hsa-miR-296-3p	2	4	1
hsa-miR-1266	1	10	2
hsa-miR-27a-5p	1	6	1
hsa-miR-4733-5p	1	5	1
hsa-miR-3185	1	6	1
hsa-miR-1295b-5p	1	4	1
hsa-miR-936	1	5	1
hsa-miR-378h	2	6	1
hsa-miR-1225-5p	1	7	1
hsa-miR-105-3p	1	3	1

hsa-miR-4253	1	3	2
hsa-miR-708-5p	1	10	3
hsa-miR-557	1	6	3
hsa-miR-4736	1	7	1
hsa-miR-3929	4	8	1
hsa-miR-5707	1	3	1
hsa-miR-584-3p	2	8	1
hsa-miR-4638-5p	1	12	1
hsa-miR-525-5p	1	7	1
hsa-miR-4683	4	5	1
hsa-miR-148a-3p	1	4	1
hsa-miR-3198	1	7	2
hsa-miR-4794	3	5	1
hsa-miR-4418	1	4	1
hsa-miR-195-3p	1	4	1
hsa-miR-181a-2-3p	2	7	1
hsa-miR-588	2	3	1
hsa-miR-3129-5p	1	5	1
hsa-miR-4684-3p	1	6	2
hsa-miR-34a-5p	1	6	1
hsa-miR-4698	1	4	2
hsa-miR-514a-5p	4	8	2
hsa-miR-550a-5p	4	11	2
hsa-miR-5581-3p	1	4	1
hsa-miR-20b-3p	2	8	1
hsa-miR-4660	1	7	1
hsa-miR-216a	1	5	1
hsa-miR-3167	1	1	1
hsa-miR-649	2	1	1
hsa-miR-520a-5p	1	8	1
hsa-miR-4784	2	8	1
hsa-miR-452-5p	1	4	1
hsa-miR-4327	2	2	1
hsa-miR-3681-3p	1	4	1
hsa-miR-29a-3p	1	3	1
hsa-miR-769-5p	2	7	2
hsa-miR-518c-3p	1	2	1

hsa-miR-432-5p	1	6	2
hsa-miR-4661-5p	2	4	1
hsa-miR-1267	1	2	1
hsa-miR-1258	1	5	2
hsa-miR-921	1	10	2
hsa-miR-4539	3	6	2
hsa-miR-500a-5p	1	13	1
hsa-miR-4712-5p	1	7	1
hsa-miR-3940-5p	1	5	1
hsa-miR-4458	3	2	1
hsa-miR-4453	1	3	1
hsa-miR-1304-3p	2	6	1
hsa-miR-214-5p	3	10	1
hsa-miR-92a-2-5p	2	7	1
hsa-miR-490-3p	3	4	1
hsa-miR-4672	2	3	1
hsa-miR-3194-3p	1	8	1
hsa-miR-431-3p	1	5	1
hsa-miR-3657	1	4	1
hsa-miR-199a-5p	1	3	1
hsa-miR-548b-3p	1	2	1
hsa-miR-516a-5p	1	4	1
hsa-miR-2467-5p	4	7	2
hsa-miR-3940-3p	2	7	1
hsa-miR-5004-5p	1	7	1
hsa-miR-5580-3p	1	3	1
hsa-miR-4517	1	3	2
hsa-miR-199a-3p	1	6	3
hsa-miR-199a-3p	1	6	3
hsa-miR-199a-3p	1	6	3
hsa-miR-199b-3p	1	6	3
hsa-miR-1251	2	5	1
hsa-miR-5197-3p	1	5	1
hsa-miR-5197-3p	1	5	1
hsa-miR-5197-3p	1	5	1
hsa-miR-26b-3p	1	5	3
hsa-miR-1294	1	7	1

hsa-miR-17-3p	1	11	1
hsa-miR-550a-3-5p	1	7	2
hsa-let-7g-5p	4	5	2
hsa-miR-28-5p	1	4	2
hsa-miR-664-5p	3	6	1
hsa-miR-132-5p	1	8	1
hsa-miR-3680-5p	1	5	1
hsa-miR-134	1	8	1
hsa-miR-3916	1	10	1
hsa-miR-2114-5p	1	7	2
hsa-miR-516b-5p	1	5	1
hsa-miR-1284	1	2	1
hsa-miR-499a-5p	1	2	1
hsa-miR-340-3p	1	7	1
hsa-miR-3157-5p	2	4	1
hsa-miR-219-2-3p	1	4	1
hsa-miR-544b	1	12	1
hsa-miR-3150b-3p	2	8	1
hsa-miR-3614-5p	1	5	1
hsa-miR-4267	2	6	1

Chapter 4 Discussion

The present study incorporated both bioinformatics and experimental approaches to screen more than 20,000 human pseudogenes, and we found *in silico* 448 TPGs that might regulate protein-coding genes through derivation of esiRNAs. In particular, we focus on *ψPPM1K*, a partial retrotranscript from *PPM1K* (protein phosphatase, Mg^{2+}/Mn^{2+} dependent, 1K). This TPG contains distinct sequences with inverted repeats capable of folding into a hairpin structure for processing into two esiRNAs that may target many cellular genes. The experimental results contained five major discoveries. First, we found that *ψPPM1K*-derived esiRNAs were expressed at higher levels in human liver tissue than in paired HCC tumor samples (**Figure 3.15**), and this is reflected in publicly available gene expression profile data (**Figure 3.14**f). Second, growth inhibitory effects were observed in *ψPPM1K*-transfected cells (**Figure 3.16**), implicating *ψPPM1K*-derived esiRNAs in the control of cell growth. This also demonstrates that esiRNAs can be derived from pseudogene transcripts in human somatic cells, thereby extending similar findings in mouse germ cells (oocytes) [51, 52]. Third, at least two esiRNAs that are expressed in liver tissue and in transfected HCC cells are derived from distinct *ψPPM1K* sequences not present in the cognate *PPM1K* gene (**Figure 3.13** and **Figure 3.15**). Fourth, we provided direct evidence that *ψPPM1K*-derived esiRNA1 down-regulates target gene *NEK8* as well as the parental *PPM1K* gene, and inhibits cell growth through down-regulation of *NEK8* (**Figure 3.20**, **Figure 3.21** and **Figure 3.22**). Moreover, the proliferation rate of *ψPPM1K*-transfected cells in which *NEK8* was co-expressed was higher than control cells, supporting the hypothesis that *NEK8* can counteract the inhibitory effects of *ψPPM1K* on HCC cell proliferation (**Figure 3.20d**). This is consistent with reports

that *NEK* family members regulate cell cycle progression [108] and that *NEK8*/*NEK9* are overexpressed in human breast cancer [108, 109]. Thus, in a broad sense, *ψPPM1K* could be considered a tumor suppressor gene. Fifth, a miRNA transfection assay showed that miR-3174 significantly down-regulated *PPM1K* in vector control cells and TPG7 cells; it also down-regulated *ψPPM1K* in vector control cells compared to a stable negative control (siCon) (**Figure 3.23b**). The miR-3174 down-regulation of *PPM1K* and *ψPPM1K*, however, is distinct from the recently elucidated decoy mechanism by which the tumor suppressor pseudogene *PTENP1* helps to regulate its cognate tumor suppressor gene, *PTEN* [48].

This study indicates two probable pathways for the generation of *ψPPM1K*-derived esiRNAs: one from dsRNAs formed by the antisense transcript of *ψPPM1K* and its cognate gene sequence in the 420-449 nt region, which gives rise to esiRNA3 (**Figure 3.13c**), and the other from a hairpin structure resulting from inverted repeats in *ψPPM1K*, giving esiRNA1 (**Figure 3.13b** and **Figure 4.1**).

It is noteworthy that *PPM1K*, encoded mitochondrial matrix serine/threonine protein phosphatase containing an N-terminal mitochondrial localization signal and a central PP2C catalytic domain, regulates the mitochondrial MPTP and is essential for cellular survival and development [107]. Another report presented that human cancer cells have abnormally high uptake and retention of mitochondria-specific dye, Rh123 [104, 105], which might represent the MPTP activity [103]. It is uncertain whether or not *PPM1K* is associated with carcinogenesis and/or oncogenesis or how it may interact with its tumor suppressor pseudogene *ψPPM1K*. In this regard, our finding indicated that *ψPPM1K* down-reuglated *PPM1K* through by producing esiRNA1 (**Figure 3.21a**) as well as *ψPPM1K* may decrease uptake and retention of Rh123, suggesting altered MPTP activity (**Figure 3.21b** and **Figure 3.22**).

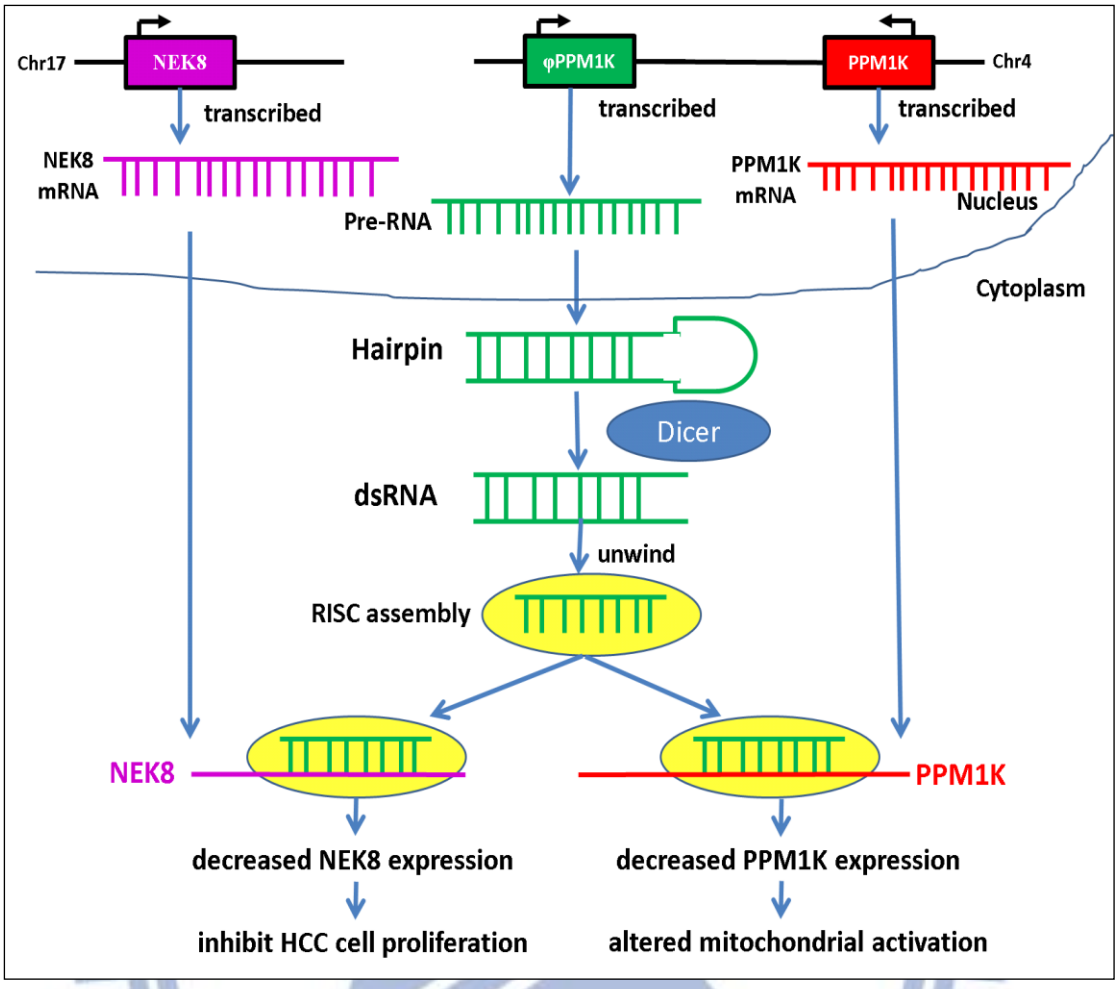


Figure 4.1 Possible genetic regulatory mechanisms involved in *ψPPM1K*.

Chapter 5 Conclusion

In conclusion, we have shown that pseudogene-derived esiRNAs can, independently of the cognate gene, regulate cell growth-related target genes in HCC. Specifically, our bioinformatics pipeline and subsequent biological experiments demonstrated that *ψPPM1K*-derived esiRNA1 modulates HCC cell proliferation via direct down-regulation of *NEK8* and may also regulate mitochondrial activation through decreased *PPM1K* cognate gene function (**Figure 4.1**). This suggests a tumor suppressor role for the esiRNAs derived from *ψPPM1K*. To explore whether other pseudogene-derived esiRNAs identified in somatic human cells [113] also possess regulatory functions, it will be of interest to perform similar experimental tests on the 447 transcribed pseudogenes and their esiRNA derivatives predicted from our *in silico* analyses.

Chapter 6 Future works

In pseudoMap, we focus on analyzing the miRNAs, post-transcriptional regulations, of TPGs. To complete the regulations of TPGs, we will analyse the transcriptional regulations of TPGs in the future. Identifying transcriptional start sties (TSS) of TPGs is the first step for deciphering transcriptional regulatory mechanism of TPGs. To archive this future aim, several high-throughput sequencing datasets will be used to analyse the TSSs. The cap analysis gene expression (CAGE) tags can be massively generated using a biotinylated cap-trapper with specific linkers to ensure that the sequences after 5' cap of cDNAs are reserved [95]. Based on this attribute, CAGE tags are extensively and adopted to identify the TSSs of genes with 5' cap transcripts [80]. Similar to CAGE tags, TSS Seq tags initially denominated by the database of transcriptional start site (DBTSS) are also the 5'-end sequences of human cDNAs base on use of the TSS Seq method [96]. More than 300 million TSS Seq tags were generated by integrating the oligo-capping method and Solexa sequencing technology, offering an abundant resource to detect TPG TSSs. Histone methylation significantly influences gene expression. H3K4me3, which represents histone H3 as trimethylated at its lysine 4 residue, is enriched around TSS and positively correlated with gene expression, regardless of whether or not the genes are transcribed productively. As a massive parallel signature sequencing technique, ChIP-Seq performs well in chromatin modifications and provides high-resolution profiling of histone methylations in human genome [114]. In addition, the transcription factors and binding site of TPGs need further studying.

In accordance with the studies of ψNOS transcript acts as a natural antisense regulator of neuronal NOS protein synthesis in snails **[44, 45],** we want to know how many TPGs act as a natural antisense regulator of its parental gene. In the fact, a number of

TPGs have indicated that the sequences of TPG with antisense complementary to its parental gene. These interesting cases need further studies.

The recently studies showed that lncRNA functions as various mechanisms, such as act as signal, decoy, guide and scaffold [43]. The similar mechanisms may involve in TPGs, as the TPGs seem to be another type of lncRNA. In this study, we have verified that TPGs act as esiRNA to regulate protein-coding genes as well as miRNAs co-regulate TPG and its parental gene. However, other mechanisms need to be study in the future. After all, no matter how computational methods are well-designed, experimental confirmation is the only way to make putative information become knowledge in the field of biological science.

Reference

1. Torrents, D., et al., *A genome-wide survey of human pseudogenes.* Genome Res, 2003. **13**(12): p. 2559-67.
2. Mighell, A.J., et al., *Vertebrate pseudogenes.* FEBS Lett, 2000. **468**(2-3): p. 109-14.
3. Balasubramanian, S., et al., *Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes.* Genome Biol, 2009. **10**(1): p. R2.
4. Harrison, P. and Z. Yu, *Frame disruptions in human mRNA transcripts, and their relationship with splicing and protein structures.* BMC Genomics, 2007. **8**: p. 371.
5. Harrison, P.M., et al., *Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability.* Nucleic Acids Res, 2005. **33**(8): p. 2374-83.
6. Vinckenbosch, N., I. Dupanloup, and H. Kaessmann, *Evolutionary fate of retroposed gene copies in the human genome.* Proc Natl Acad Sci U S A, 2006. **103**(9): p. 3220-5.
7. Zheng, D., et al., *Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution.* Genome Res, 2007. **17**(6): p. 839-51.
8. Zheng, D., et al., *Integrated pseudogene annotation for human chromosome 22: evidence for transcription.* J Mol Biol, 2005. **349**(1): p. 27-45.
9. Khachane, A.N. and P.M. Harrison, *Assessing the genomic evidence for conserved transcribed pseudogenes under selection.* BMC Genomics, 2009. **10**: p. 435.
10. Imanishi, T., et al., *Integrative annotation of 21,037 human genes validated by full-length cDNA clones.* PLoS Biol, 2004. **2**(6): p. e162.
11. Morozova, N., et al., *Kinetic signatures of microRNA modes of action.* RNA, 2012. **18**(9): p. 1635-55.
12. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.
13. Birney, E., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. **447**(7146): p. 799-816.
14. Kapranov, P., et al., *RNA maps reveal new RNA classes and a possible function for pervasive transcription.* Science, 2007. **316**(5830): p. 1484-8.
15. Kapranov, P., A.T. Willingham, and T.R. Gingeras, *Genome-wide transcription*

and the implications for genomic organization. Nat Rev Genet, 2007. **8**(6): p. 413-23.

16. Alexander, R.P., et al., *Annotating non-coding regions of the genome.* Nat Rev Genet, 2010. **11**(8): p. 559-71.
17. Cooper, D.N., et al., *Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics.* Hum Mutat, 2010. **31**(6): p. 631-55.
18. Wery, M., M. Kwapisz, and A. Morillon, *Noncoding RNAs in gene regulation.* Wiley Interdiscip Rev Syst Biol Med, 2011. **3**(6): p. 728-38.
19. Bernard, D., et al., *A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression.* EMBO J, 2010. **29**(18): p. 3082-93.
20. Pandey, R.R., et al., *Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation.* Mol Cell, 2008. **32**(2): p. 232-46.
21. Terranova, R., et al., *Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos.* Dev Cell, 2008. **15**(5): p. 668-79.
22. Nagano, T., et al., *The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin.* Science, 2008. **322**(5908): p. 1717-20.
23. Sleutels, F., R. Zwart, and D.P. Barlow, *The non-coding Air RNA is required for silencing autosomal imprinted genes.* Nature, 2002. **415**(6873): p. 810-3.
24. Kanduri, C., *Kcnq1ot1: a chromatin regulatory RNA.* Semin Cell Dev Biol, 2011. **22**(4): p. 343-50.
25. Khalil, A.M., et al., *Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression.* Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11667-72.
26. Mondal, T., et al., *Characterization of the RNA content of chromatin.* Genome Res, 2010. **20**(7): p. 899-907.
27. Hung, T., et al., *Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters.* Nat Genet, 2011. **43**(7): p. 621-9.
28. Diederichs, S. and D.A. Haber, *Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression.* Cell, 2007. **131**(6): p. 1097-108.
29. Lee, K.J., et al., *Do human transposable element small RNAs serve primarily as genome defenders or genome regulators?* Mob Genet Elements, 2012. **2**(1): p. 19-25.

30. Ishizu, H., A. Nagao, and H. Siomi, *Gatekeepers for Piwi-piRNA complexes to enter the nucleus.* Curr Opin Genet Dev, 2011. **21**(4): p. 484-90.
31. Chen, L.L. and G.G. Carmichael, *Decoding the function of nuclear long non-coding RNAs.* Curr Opin Cell Biol, 2010. **22**(3): p. 357-64.
32. Clark, M.B. and J.S. Mattick, *Long noncoding RNAs in cell biology.* Semin Cell Dev Biol, 2011. **22**(4): p. 366-76.
33. Seto, A.G., R.E. Kingston, and N.C. Lau, *The coming of age for Piwi proteins.* Mol Cell, 2007. **26**(5): p. 603-9.
34. Siomi, M.C., et al., *PIWI-interacting small RNAs: the vanguard of genome defence.* Nat Rev Mol Cell Biol, 2011. **12**(4): p. 246-58.
35. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-97.
36. Jovanovic, M. and M.O. Hengartner, *miRNAs and apoptosis: RNAs to die for.* Oncogene, 2006. **25**(46): p. 6176-87.
37. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.
38. Lund, E., et al., *Nuclear export of microRNA precursors.* Science, 2004. **303**(5654): p. 95-8.
39. Yi, R., et al., *Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs.* Genes Dev, 2003. **17**(24): p. 3011-6.
40. Kim, D.H. and J.J. Rossi, *Strategies for silencing human disease using RNA interference.* Nat Rev Genet, 2007. **8**(3): p. 173-84.
41. Shriner, D., *Moving toward System Genetics through Multiple Trait Analysis in Genome-Wide Association Studies.* Front Genet, 2012. **3**: p. 1.
42. Sana, J., et al., *Novel classes of non-coding RNAs and cancer.* J Transl Med, 2012. **10**: p. 103.
43. Wang, K.C. and H.Y. Chang, *Molecular mechanisms of long noncoding RNAs.* Mol Cell, 2011. **43**(6): p. 904-14.
44. Korneev, S.A., J.H. Park, and M. O'Shea, *Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene.* J Neurosci, 1999. **19**(18): p. 7711-20.
45. Korneev, S. and M. O'Shea, *Evolution of nitric oxide synthase regulatory genes by DNA inversion.* Mol Biol Evol, 2002. **19**(8): p. 1228-33.
46. Hirotsune, S., et al., *An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene.* Nature, 2003. **423**(6935): p. 91-6.
47. Gray, T.A., et al., *The putatively functional Mkrn1-p1 pseudogene is neither expressed nor imprinted, nor does it regulate its source gene in trans.* Proc Natl Acad Sci U S A, 2006. **103**(32): p. 12039-44.

48. Poliseno, L., et al., *A coding-independent function of gene and pseudogene mRNAs regulates tumour biology.* Nature, 2010. **465**(7301): p. 1033-8.
49. Ghildiyal, M., et al., *Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells.* Science, 2008. **320**(5879): p. 1077-81.
50. Brennecke, J., et al., *Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila.* Cell, 2007. **128**(6): p. 1089-103.
51. Tam, O.H., et al., *Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes.* Nature, 2008. **453**(7194): p. 534-8.
52. Watanabe, T., et al., *Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes.* Nature, 2008. **453**(7194): p. 539-43.
53. Czech, B., et al., *An endogenous small interfering RNA pathway in Drosophila.* Nature, 2008. **453**(7196): p. 798-802.
54. Kawamura, Y., et al., *Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells.* Nature, 2008. **453**(7196): p. 793-7.
55. Okamura, K., et al., *The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs.* Nature, 2008. **453**(7196): p. 803-6.
56. Karro, J.E., et al., *Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation.* Nucleic Acids Res, 2006.
57. Khelifi, A., L. Duret, and D. Mouchiroud, *HOPPSIGEN: a database of human and mouse processed pseudogenes.* Nucleic Acids Res, 2005. **33**(Database issue): p. D59-66.
58. Bischof, J.M., et al., *Genome-wide identification of pseudogenes capable of disease-causing gene conversion.* Hum Mutat, 2006. **27**(6): p. 545-52.
59. Griffiths-Jones, S., et al., *miRBase: microRNA sequences, targets and gene nomenclature.* Nucleic Acids Res, 2006. **34**(Database issue): p. D140-4.
60. Birney, E., et al., *Ensembl 2006.* Nucleic Acids Res, 2006. **34**(Database issue): p. D556-61.
61. Hinrichs, A.S., et al., *The UCSC Genome Browser Database: update 2006.* Nucleic Acids Res, 2006. **34**(Database issue): p. D590-8.
62. Mituyama, T., et al., *The Functional RNA Database 3.0: databases to support mining and annotation of functional RNAs.* Nucleic Acids Res, 2009. **37**(Database issue): p. D89-92.
63. Carninci, P., et al., *The transcriptional landscape of the mammalian genome.* Science, 2005. **309**(5740): p. 1559-63.
64. Vitali, M., et al., *Analysis of the genes coding for subunit 10 and 15 of cytochrome c oxidase in Alzheimer's disease.* J Neural Transm, 2009. **116**(12): p. 1635-41.
65. He, S., et al., *NONCODE v2.0: decoding the non-coding.* Nucleic Acids Res,

2008. **36**(Database issue): p. D170-2.
66. Griffiths-Jones, S., et al., *Rfam: annotating non-coding RNAs in complete genomes.* Nucleic Acids Res, 2005. **33**(Database issue): p. D121-4.
67. Pang, K.C., et al., *RNAdb 2.0--an expanded database of mammalian non-coding RNAs.* Nucleic Acids Res, 2007. **35**(Database issue): p. D178-82.
68. Lestrade, L. and M.J. Weber, *snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs.* Nucleic Acids Res, 2006. **34**(Database issue): p. D158-62.
69. Barrett, T., et al., *NCBI GEO: mining tens of millions of expression profiles--database and tools update.* Nucleic Acids Res, 2006.
70. Brazma, A., et al., *Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.* Nat Genet, 2001. **29**(4): p. 365-71.
71. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
72. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Res, 1994. **22**(22): p. 4673-80.
73. Kruger, J. and M. Rehmsmeier, *RNAhybrid: microRNA target prediction easy, fast and flexible.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W451-4.
74. Lewis, B.P., et al., *Prediction of mammalian microRNA targets.* Cell, 2003. **115**(7): p. 787-98.
75. John, B., et al., *Human MicroRNA targets.* PLoS Biol, 2004. **2**(11): p. e363.
76. Mathews, D.H., et al., *Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.* J Mol Biol, 1999. **288**(5): p. 911-40.
77. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15.
78. Costa, F.F., *Non-coding RNAs: new players in eukaryotic biology.* Gene, 2005. **357**(2): p. 83-94.
79. Griffiths-Jones, S., et al., *miRBase: tools for microRNA genomics.* Nucleic Acids Res, 2008. **36**(Database issue): p. D154-8.
80. Carninci, P., et al., *Genome-wide analysis of mammalian promoter architecture and evolution.* Nat Genet, 2006. **38**(6): p. 626-35.
81. Morin, R.D., et al., *Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells.* Genome Res, 2008. **18**(4): p. 610-21.

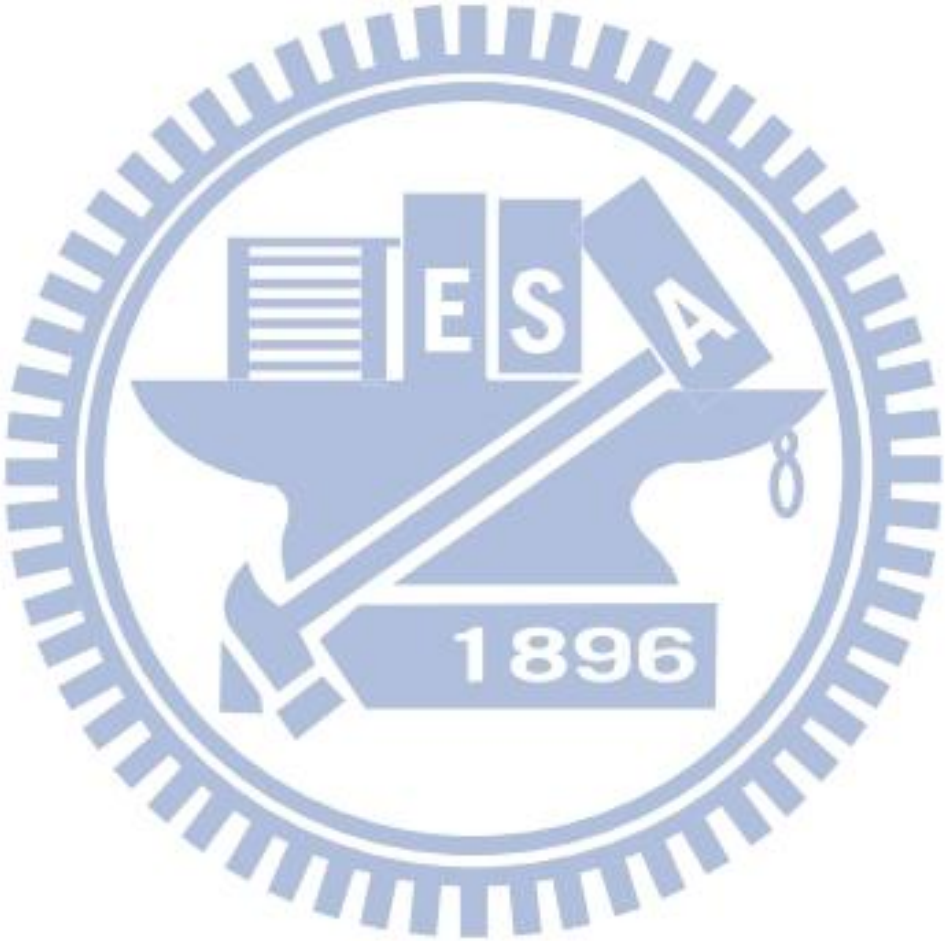
82. Seila, A.C., et al., *Divergent transcription from active promoters.* Science, 2008. **322**(5909): p. 1849-51.
83. Yeo, G.W., et al., *Alternative splicing events identified in human embryonic stem cells and neural progenitors.* PLoS Comput Biol, 2007. **3**(10): p. 1951-67.
84. Hsu, S.D., et al., *miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes.* Nucleic Acids Res, 2008. **36**(Database issue): p. D165-9.
85. Friedman, R.C., et al., *Most mammalian mRNAs are conserved targets of microRNAs.* Genome Res, 2009. **19**(1): p. 92-105.
86. Grimson, A., et al., *MicroRNA targeting specificity in mammals: determinants beyond seed pairing.* Mol Cell, 2007. **27**(1): p. 91-105.
87. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.* Cell, 2005. **120**(1): p. 15-20.
88. Barrett, T. and R. Edgar, *Gene expression omnibus: microarray data storage, submission, retrieval, and analysis.* Methods Enzymol, 2006. **411**: p. 352-69.
89. Su, A.I., et al., *A gene atlas of the mouse and human protein-encoding transcriptomes.* Proc Natl Acad Sci U S A, 2004. **101**(16): p. 6062-7.
90. Yu, K., et al., *A precisely regulated gene expression cassette potently modulates metastasis and survival in multiple solid cancers.* PLoS Genet, 2008. **4**(7): p. e1000129.
91. Hsu, S.D., et al., *miRTarBase: a database curates experimentally validated microRNA-target interactions.* Nucleic Acids Res, 2011. **39**(Database issue): p. D163-9.
92. Okuda, S., et al., *KEGG Atlas mapping for global analysis of metabolic pathways.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W423-6.
93. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
94. Kozomara, A. and S. Griffiths-Jones, *miRBase: integrating microRNA annotation and deep-sequencing data.* Nucleic Acids Res, 2011. **39**(Database issue): p. D152-7.
95. Shiraki, T., et al., *Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15776-81.
96. Yamashita, R., et al., *DBTSS provides a tissue specific dynamic view of Transcription Start Sites.* Nucleic Acids Res, 2010. **38**(Database issue): p. D98-104.

97. Jover-Gil, S., H. Candela, and M.R. Ponce, *Plant microRNAs and development.* Int J Dev Biol, 2005. **49**(5-6): p. 733-44.
98. Chan, C.H., et al., *Subcellular and functional proteomic analysis of the cellular responses induced by Helicobacter pylori.* Mol Cell Proteomics, 2006. **5**(4): p. 702-13.
99. Yuo, C.Y., et al., *5-(N-ethyl-N-isopropyl)-amiloride enhances SMN2 exon 7 inclusion and protein expression in spinal muscular atrophy cells.* Ann Neurol, 2008. **63**(1): p. 26-34.
100. Liu, T.C., et al., *Epigenetic alteration of the SOCS1 gene in chronic myeloid leukaemia.* Br J Haematol, 2003. **123**(4): p. 654-61.
101. Zhou, X., et al., *MicroRNA profiling using microParaflo microfluidic array technology.* Methods Mol Biol, 2012. **822**: p. 153-82.
102. Fleming, E.J., et al., *What's new is old: resolving the identity of Leptothrix ochracea using single cell genomics, pyrosequencing and FISH.* PLoS One, 2011. **6**(3): p. e17769.
103. Rottenberg, H. and S. Wu, *Quantitative assay by flow cytometry of the mitochondrial membrane potential in intact cells.* Biochim Biophys Acta, 1998. **1404**(3): p. 393-404.
104. Nadakavukaren, K.K., J.J. Nadakavukaren, and L.B. Chen, *Increased rhodamine 123 uptake by carcinoma cells.* Cancer Res, 1985. **45**(12 Pt 1): p. 6093-9.
105. Darzynkiewicz, Z., et al., *Interaction of rhodamine 123 with living cells studied by flow cytometry.* Cancer Res, 1982. **42**(3): p. 799-806.
106. de Yebenes, V.G. and A.R. Ramiro, *MicroRNA activity in B lymphocytes.* Methods Mol Biol, 2010. **667**: p. 177-92.
107. Lu, G., et al., *A novel mitochondrial matrix serine/threonine protein phosphatase regulates the mitochondria permeability transition pore and is essential for cellular survival and development.* Genes Dev, 2007. **21**(7): p. 784-96.
108. O'Regan, L., J. Blot, and A.M. Fry, *Mitotic regulation by NIMA-related kinases.* Cell Div, 2007. **2**: p. 25.
109. Bowers, A.J. and J.F. Boylan, *Nek8, a NIMA family kinase member, is overexpressed in primary human breast tumors.* Gene, 2004. **328**: p. 135-42.
110. Tompkins, V., et al., *Identification of novel ARF binding proteins by two-hybrid screening.* Cell Cycle, 2006. **5**(6): p. 641-6.
111. Tompkins, V.S., et al., *A novel nuclear interactor of ARF and MDM2 (NIAM) that maintains chromosomal stability.* J Biol Chem, 2007. **282**(2): p. 1322-33.
112. Nohno, T., et al., *Identification of a human type II receptor for bone*

morphogenetic protein-4 that forms differential heteromeric complexes with bone morphogenetic protein type I receptors. J Biol Chem, 1995. **270**(38): p. 22522-6.

113. Okamura, K., W.J. Chung, and E.C. Lai, *The long and short of inverted repeat genes in animals: microRNAs, mirtrons and hairpin RNAs.* Cell Cycle, 2008. **7**(18): p. 2840-5.

114. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.



Appendix I - TPG-derived esiRNAs

Table A1 The list of TPGs production of endo-siRNAs which supported by in public deep sequencing data from various sRNA libraries.

No.	TPG	Description	Chr	Gene Start	Gene End	Strand	Biotype	Transcript Count
1	TPG-PPM1K		4	89,398,481	89,398,959	1	pseudogene	1
2	SEMA3B	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B [Source:HGNC Symbol;Acc:10724]	3	50,304,990	50,314,977	1	polymorphic_pseudogene	16
3	CROCCP3	ciliary rootlet coiled-coil, rootletin pseudogene 3 [Source:HGNC Symbol;Acc:29405]	1	16,793,931	16,819,196	-1	pseudogene	2
4	AL118502.1		20	820,090	820,620	-1	pseudogene	1
5	RP11-22B23.1		12	9,429,831	9,466,684	1	pseudogene	3
6	SMAD5	SMAD family member 5 [Source:HGNC Symbol;Acc:6771]	5	135,468,534	135,524,435	1	polymorphic_pseudogene	13
7	PMS2P5	postmeiotic segregation increased 2 pseudogene 5 [Source:HGNC Symbol;Acc:9130]	7	74,306,894	74,366,314	1	pseudogene	5
8	HMGB1P1	high mobility group box 1 pseudogene 1 [Source:HGNC Symbol;Acc:4993]	20	56,063,448	56,064,083	-1	pseudogene	1
9	RPL13AP25	ribosomal protein L13a pseudogene 25 [Source:HGNC	13	55,014,839	55,015,450	-1	pseudogene	1

		Symbol;Acc:36981]						
10	RP11-122C9.1		1	97,144,430	97,145,176	-1	pseudogene	1
11	RPL14P1	ribosomal protein L14 pseudogene 1 [Source:HGNC Symbol;Acc:31384]	12	63,359,095	63,359,764	1	pseudogene	2
12	AC004453.8		7	44,507,441	44,507,939	1	pseudogene	2
13	SPON1	spondin 1, extracellular matrix protein [Source:HGNC Symbol;Acc:11252]	11	13,983,914	14,289,646	1	polymorphic_pseudogene	3
14	SYN2	synapsin II [Source:HGNC Symbol;Acc:11495]	3	12,045,862	12,232,907	1	polymorphic_pseudogene	7
15	NCF1C	neutrophil cytosolic factor 1C pseudogene [Source:HGNC Symbol;Acc:32523]	7	74,572,445	74,587,848	-1	pseudogene	4
16	ZNF355P	zinc finger protein 355, pseudogene [Source:HGNC Symbol;Acc:17218]	21	14,467,626	14,486,111	-1	pseudogene	1
17	BCRP6	breakpoint cluster region pseudogene 6 [Source:HGNC Symbol;Acc:39074]	22	21,645,049	21,648,875	-1	pseudogene	1
18	AC005280.1		14	73,957,644	73,960,105	1	pseudogene	1
19	CSPG4P4Y	chondroitin sulfate proteoglycan 4 pseudogene 4, Y-linked [Source:HGNC Symbol;Acc:38535]	Y	27,624,416	27,632,852	1	pseudogene	2
20	CSPG4P3Y	chondroitin sulfate proteoglycan 4 pseudogene 3, Y-linked [Source:HGNC Symbol;Acc:38111]	Y	26,332,656	26,338,017	-1	pseudogene	1
21	AC007318.5		2	65,432,137	65,433,396	1	pseudogene	2
22	TERF1P1	telomeric repeat binding factor (NIMA-interacting) 1 pseudogene 1 [Source:HGNC Symbol;Acc:16733]	21	15,148,407	15,149,587	1	pseudogene	1

23	PMS2P2	postmeiotic segregation increased 2 pseudogene 2 [Source:HGNC Symbol;Acc:9127]	7	74,880,905	74,988,262	-1	pseudogene	5
24	RP11-313P13.4		7	72,507,941	72,515,008	1	pseudogene	1
25	RP11-209A2.1		6	27,179,023	27,179,663	1	pseudogene	1
26	AC026271.5		17	18,553,508	18,554,855	1	pseudogene	2
27	CTAGE15P	CTAGE family, member 15, pseudogene [Source:HGNC Symbol;Acc:37295]	7	143,268,819	143,271,478	1	pseudogene	2
28	CTAGE6P	CTAGE family, member 6, pseudogene [Source:HGNC Symbol;Acc:28644]	7	143,452,341	143,454,751	-1	pseudogene	1
29	AL359758.1		1	145,139,025	145,139,569	-1	pseudogene	2
30	AC074182.1		2	184,472,170	184,472,847	-1	pseudogene	1
31	FAM75C2	family with sequence similarity 75, member C2 [Source:HGNC Symbol;Acc:24508]	9	90,744,220	90,749,900	-1	pseudogene	2
32	RPS3AP5	ribosomal protein S3A pseudogene 5 [Source:HGNC Symbol;Acc:23744]	10	86,320,199	86,321,019	1	pseudogene	1
33	RP11-308D13.2		4	140,619,298	140,619,708	-1	pseudogene	2
34	CTD-2192J16.15		19	12,754,089	12,754,733	-1	pseudogene	1
35	RP11-349N19.2		X	122,648,324	122,648,723	1	pseudogene	1
36	CTGLF12P	centaurin, gamma-like family, member 12 pseudogene	10	49,217,898	49,239,743	-1	pseudogene	2

		[Source:HGNC Symbol;Acc:23661]						
37	RP13-221M14.5	heterogeneous nuclear ribonucleoprotein C-like (LOC440563), mRNA [Source:RefSeq DNA;Acc:NM_001136561]	1	13,182,960	13,184,326	-1	pseudogene	2
38	AC124309.1		15	23,888,695	23,890,979	-1	pseudogene	1
39	RP11-1319K7.1		5	69,390,202	69,424,261	-1	pseudogene	2
40	SRP9P1	signal recognition particle 9 pseudogene 1 [Source:HGNC Symbol;Acc:23402]	10	93,565,803	93,567,253	-1	pseudogene	2
41	RPL24P4	ribosomal protein L24 pseudogene 4 [Source:HGNC Symbol;Acc:21371]	6	42,924,083	42,924,503	-1	pseudogene	1
42	RP11-556K13.1		1	102,251,892	102,254,059	-1	pseudogene	2
43	RPS7P1	ribosomal protein S7 pseudogene 1 [Source:HGNC Symbol;Acc:17081]	17	26,794,814	26,795,460	1	pseudogene	2
44	FOXD1	forkhead box D1 [Source:HGNC Symbol;Acc:3802]	5	72,742,083	72,744,352	-1	pseudogene	2
45	CD24P4	CD24 molecule pseudogene 4 [Source:HGNC Symbol;Acc:1649]	Y	21,154,139	21,154,595	-1	pseudogene	1
46	CTD-2287O16.1		5	115,387,607	115,387,986	-1	pseudogene	2
47	RPL12P4	ribosomal protein L12 pseudogene 4 [Source:HGNC Symbol;Acc:16587]	20	53,691,155	53,691,685	1	pseudogene	2

48	RP11-887P2.3		12	94,034,598	94,035,362	-1	pseudogene	2
49	RP5-1182A14.3		1	16,972,069	16,976,914	1	pseudogene	3
50	AC114737.4		7	76,607,934	76,673,092	1	pseudogene	3
51	AC007842.1		19	40,448,563	40,449,634	-1	pseudogene	1
52	HS6ST1P1	heparan sulfate 6-O-sulfotransferase 1 pseudogene 1 [Source:HGNC Symbol;Acc:31835]	1	21,754,795	21,756,030	1	pseudogene	2
53	PMS2P10	postmeiotic segregation increased 2 pseudogene 10 [Source:HGNC Symbol;Acc:9124]	7	74,916,804	74,928,928	-1	pseudogene	2
54	AC016839.1		18	23,749,724	23,751,313	-1	pseudogene	1
55	AC138783.12		7	74,702,428	74,714,647	-1	pseudogene	2
56	FAM35B2	family with sequence similarity 35, member B2 [Source:HGNC Symbol;Acc:34038]	10	47,379,727	47,420,712	1	pseudogene	2
57	AC004448.3		17	19,349,246	19,350,169	-1	pseudogene	2
58	RPS2P4	ribosomal protein S2 pseudogene 4 [Source:HGNC Symbol;Acc:19814]	14	105,302,331	105,303,221	-1	pseudogene	1
59	EEF1A1P5	eukaryotic translation elongation factor 1 alpha 1 pseudogene 5 [Source:HGNC Symbol;Acc:3200]	9	135,894,816	135,896,553	1	pseudogene	2
60	RP11-497H16.5		5	69,812,214	69,851,093	-1	pseudogene	1
61	AC002994.1		17	59,667,894	59,668,541	-1	pseudogene	1
62	AC009093.2		16	29,259,550	29,261,449	1	pseudogene	1

63	ZNF187	zinc finger protein 187 [Source:HGNC Symbol;Acc:12978]	6	28,234,788	28,245,974	1	pseudogene	5
64	AL133458.1		6	167,357,853	167,360,151	-1	pseudogene	1
65	TPSB2	tryptase beta 2 (gene/pseudogene) [Source:HGNC Symbol;Acc:14120]	16	1,277,272	1,280,214	-1	polymorphic_pseudogene	3
66	EIF4BP6	eukaryotic translation initiation factor 4B pseudogene 6 [Source:HGNC Symbol;Acc:37939]	7	104,308,196	104,310,023	1	pseudogene	1
67	RP11-460N11.2		9	69,830,288	69,848,759	1	pseudogene	1
68	GPX1P1	glutathione peroxidase pseudogene 1 [Source:HGNC Symbol;Acc:4560]	X	13,396,854	13,397,459	-1	pseudogene	1
69	CTB-161M19.3		5	118,309,491	118,309,817	1	pseudogene	1
70	AC011385.1		5	137,881,629	137,882,992	-1	pseudogene	1
71	RP11-267J23.4		12	12,264,097	12,264,423	1	pseudogene	2
72	BZW1P2	basic leucine zipper and W2 domains 1 pseudogene 2 [Source:HGNC Symbol;Acc:33954]	3	116,364,592	116,365,574	-1	pseudogene	2
73	PPIAP22	peptidylprolyl isomerase A (cyclophilin A) pseudogene 22 [Source:HGNC Symbol;Acc:17236]	21	20,230,097	20,230,594	1	pseudogene	2
74	AC004878.3		7	74,945,134	74,957,275	-1	pseudogene	2
75	FAM108A3P	family with sequence similarity 108, member A3, pseudogene [Source:HGNC Symbol;Acc:33538]	1	146,076,838	146,080,009	1	pseudogene	1
76	AC012615.1		19	1,925,185	1,926,009	1	pseudogene	1
77	AL162458.1		20	44,644,111	44,644,231	1	pseudogene	1

78	AL021707.1		22	39,128,914	39,129,588	1	pseudogene	1
79	RP5-886K2.1		1	24,032,291	24,033,424	-1	pseudogene	1
80	RP11-243J18.3		1	155,581,011	155,720,105	1	pseudogene	3
81	RP11-744H18.1		1	149,036,435	149,067,496	1	pseudogene	1
82	FAM108A2	family with sequence similarity 108, member A2 [Source:HGNC Symbol;Acc:28394]	1	147,618,673	147,621,845	1	pseudogene	1
83	CTGLF11P	centaurin, gamma-like family, member 11 pseudogene [Source:HGNC Symbol;Acc:23660]	10	47,708,307	47,730,235	1	pseudogene	2
84	RPL12P1	ribosomal protein L12 pseudogene 1 [Source:HGNC Symbol;Acc:13976]	6	33,367,836	33,368,333	-1	pseudogene	1
85	AC011737.2		2	204,055,503	204,056,000	1	pseudogene	2
86	AC019178.1		2	190,788,062	190,788,942	1	pseudogene	2
87	POTEKP	POTE ankyrin domain family, member K, pseudogene [Source:HGNC Symbol;Acc:30182]	2	132,349,325	132,384,839	1	pseudogene	2
88	RPS26P8	ribosomal protein S26 pseudogene 8 [Source:HGNC Symbol;Acc:31329]	17	43,685,909	43,686,349	1	pseudogene	2
89	AC083899.3		2	87,352,513	87,423,770	1	pseudogene	1
90	RP11-15J10.1		9	70,181,520	70,216,621	-1	pseudogene	1
91	ZNF204P	zinc finger protein 204, pseudogene [Source:HGNC Symbol;Acc:12995]	6	27,325,604	27,339,304	-1	pseudogene	2
92	COX17P1	COX17 cytochrome c oxidase assembly homolog (S.	13	47,064,997	47,065,397	1	pseudogene	2

		cerevisiae) pseudogene 1 [Source:HGNC Symbol;Acc:24341]						
93	POM121B	POM121 membrane glycoprotein B (pseudogene) [Source:HGNC Symbol;Acc:34004]	7	72,707,497	72,716,496	1	pseudogene	2
94	RP11-473C18.1		15	43,407,897	43,408,691	-1	pseudogene	2
95	IGLL3P	immunoglobulin lambda-like polypeptide 3, pseudogene [Source:HGNC Symbol;Acc:5872]	22	25,714,223	25,716,047	1	pseudogene	1
96	HNRNPA1P4	heterogeneous nuclear ribonucleoprotein A1 pseudogene 4 [Source:HGNC Symbol;Acc:32234]	8	83,203,859	83,204,614	-1	pseudogene	1
97	RP11-592N21.1		15	71,633,466	71,634,086	-1	pseudogene	2
98	RPL15P3	ribosomal protein L15 pseudogene 3 [Source:HGNC Symbol;Acc:21538]	6	12,514,342	12,515,006	1	pseudogene	2
99	RPS26P3	ribosomal protein S26 pseudogene 3 [Source:HGNC Symbol;Acc:23411]	9	9,090,874	9,091,323	1	pseudogene	2
100	RPS26P6	ribosomal protein S26 pseudogene 6 [Source:HGNC Symbol;Acc:31090]	8	101,907,974	101,908,387	1	pseudogene	2
101	RP11-641D5.1		3	169,201,459	169,201,868	-1	pseudogene	1
102	AC093106.7		7	131,346,711	131,347,383	-1	pseudogene	1
103	RP11-212P7.1		7	128,210,295	128,210,742	-1	pseudogene	1
104	RPS7P11	ribosomal protein S7 pseudogene 11 [Source:HGNC Symbol;Acc:35841]	17	44,798,948	44,799,533	-1	pseudogene	1

105	ANXA2P1	annexin A2 pseudogene 1 [Source:HGNC Symbol;Acc:538]	4	154,228,944	154,229,869	-1	pseudogene	1
106	HSPD1P1	heat shock 60kDa protein 1 (chaperonin) pseudogene 1 [Source:HGNC Symbol;Acc:35133]	5	21,882,694	21,884,421	1	pseudogene	3
107	RPL18AP3	ribosomal protein L18a pseudogene 3 [Source:HGNC Symbol;Acc:31387]	12	104,659,056	104,659,669	1	pseudogene	2
108	NPM1P5	nucleophosmin 1 (nucleolar phosphoprotein B23, numatrin) pseudogene 5 [Source:HGNC Symbol;Acc:7925]	12	9,847,980	9,848,823	1	pseudogene	1
109	RPLP0P6	ribosomal protein, large, P0 pseudogene 6 [Source:HGNC Symbol;Acc:36404]	2	38,708,920	38,710,012	1	pseudogene	2
110	RP4-595K12.1		1	76,047,784	76,048,240	1	pseudogene	1
111	RPSAP54	ribosomal protein SA pseudogene 54 [Source:HGNC Symbol;Acc:36373]	13	21,535,449	21,536,334	1	pseudogene	1
112	HMGB3P30	high mobility group box 3 pseudogene 30 [Source:HGNC Symbol;Acc:39333]	X	111,933,143	111,933,748	1	pseudogene	2
113	RP11-344N10.3		10	74,765,591	74,766,353	-1	pseudogene	1
114	HMGB1P10	high mobility group box 1 pseudogene 10 [Source:HGNC Symbol;Acc:4994]	22	26,956,492	26,957,054	1	pseudogene	1
115	AC004692.5		7	43,313,758	43,314,181	1	pseudogene	1
116	NHP2P2	NHP2 ribonucleoprotein homolog (yeast) pseudogene 2 [Source:HGNC Symbol;Acc:22797]	7	27,087,788	27,088,247	1	pseudogene	1
117	AC005062.3		7	20,042,346	20,042,909	1	pseudogene	2

118	EEF1B3	eukaryotic translation elongation factor 1 beta 3 [Source:HGNC Symbol;Acc:3209]	5	67,455,046	67,455,721	-1	pseudogene	1
119	RP11-327L3.5		9	35,971,341	35,972,315	-1	pseudogene	1
120	NPM1P6	nucleophosmin 1 (nucleolar phosphoprotein B23, numatrin) pseudogene 6 [Source:HGNC Symbol;Acc:7926]	8	62,114,909	62,115,779	1	pseudogene	1
121	RPS10P2	ribosomal protein S10 pseudogene 2 [Source:HGNC Symbol;Acc:16594]	20	14,738,209	14,738,702	1	pseudogene	1
122	PTMAP5	prothymosin, alpha pseudogene 5 [Source:HGNC Symbol;Acc:9628]	13	82,264,203	82,264,535	1	pseudogene	1
123	RP11-396A22.4		13	41,958,909	41,959,319	-1	pseudogene	1
124	CTAGE11P	CTAGE family, member 11, pseudogene [Source:HGNC Symbol;Acc:37293]	13	75,812,080	75,814,432	-1	pseudogene	1
125	AC093799.2		7	98,015,113	98,015,896	-1	pseudogene	1
126	RP11-121L10.3		11	90,015,728	90,017,340	1	pseudogene	1
127	CTB-13H5.1		5	149,473,860	149,474,619	1	pseudogene	2
128	RPS15AP1	ribosomal protein S15a pseudogene 1 [Source:HGNC Symbol;Acc:16598]	20	21,146,846	21,147,236	-1	pseudogene	1
129	SLC6A10P	solute carrier family 6 (neurotransmitter transporter, creatine), member 10, pseudogene [Source:HGNC Symbol;Acc:11043]	16	32,888,797	32,896,822	-1	pseudogene	2
130	HNRNPA3P3	heterogeneous nuclear ribonucleoprotein A3 pseudogene 3	X	139,114,503	139,115,665	1	pseudogene	2

		[Source:HGNC Symbol;Acc:39772]						
131	AC007347.1		16	54,146,970	54,148,378	1	pseudogene	1
132	RP11-144J4.3		12	122,849,327	122,850,176	1	pseudogene	1
133	HMGB1P4	high mobility group box 1 pseudogene 4 [Source:HGNC Symbol;Acc:4996]	2	171,458,172	171,458,738	-1	pseudogene	2
134	DDX12	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 12 [Source:HGNC Symbol;Acc:2737]	12	9,570,309	9,600,825	-1	pseudogene	3
135	TMEM14D	transmembrane protein 14D [Source:HGNC Symbol;Acc:15660]	10	70,304,246	70,304,590	-1	pseudogene	1
136	RP3-507I15.1		X	140,232,663	140,233,358	1	pseudogene	1
137	RP11-367G6.1		6	24,976,646	24,977,204	1	pseudogene	1
138	AC073621.1		17	17,286,691	17,287,326	-1	pseudogene	2
139	RPL21P2	ribosomal protein L21 pseudogene 2 [Source:HGNC Symbol;Acc:16539]	20	4,040,280	4,040,760	-1	pseudogene	1
140	TERF1P4	telomeric repeat binding factor (NIMA-interacting) 1 pseudogene 4 [Source:HGNC Symbol;Acc:38500]	X	83,003,833	83,004,947	-1	pseudogene	1
141	GPX1P2	glutathione peroxidase pseudogene 2 [Source:HGNC Symbol;Acc:4561]	21	28,515,663	28,516,268	-1	pseudogene	1
142	RPL27AP	ribosomal protein L27a pseudogene [Source:HGNC Symbol;Acc:16250]	20	42,281,163	42,281,931	1	pseudogene	1
143	BCRP3	breakpoint cluster region pseudogene 3 [Source:HGNC Symbol;Acc:1016]	22	25,028,882	25,046,799	1	pseudogene	2

144	HNRNPA1P7	heterogeneous nuclear ribonucleoprotein A1 pseudogene 7 [Source:HGNC Symbol;Acc:31015]	18	29,991,471	29,993,200	-1	pseudogene	3
145	XXbac-B562F10.11		22	20,804,409	20,806,111	1	pseudogene	1
146	BCRP7	breakpoint cluster region pseudogene 7 [Source:HGNC Symbol;Acc:39075]	22	18,843,134	18,846,153	-1	pseudogene	1
147	ZNF962P	zinc finger protein 962, pseudogene [Source:HGNC Symbol;Acc:39250]	13	19,041,312	19,059,588	-1	pseudogene	1
148	AL356585.3		GL000212.1	35,322	36,221	-1	pseudogene	1
149	RP11-490H24.5		12	104,424,557	104,425,321	1	pseudogene	2
150	RP3-334F4.2		6	54,467,285	54,467,948	-1	pseudogene	1
151	RP3-522P13.1		6	25,261,467	25,261,948	1	pseudogene	1
152	RP1-40E16.8		6	3,177,278	3,179,998	-1	pseudogene	1
153	RP11-30P6.1		6	85,995,794	86,000,567	1	pseudogene	1
154	RP11-460H18.1		10	32,477,251	32,477,745	-1	pseudogene	1
155	RP1-179E13.1		6	127,004,181	127,004,986	1	pseudogene	1
156	ISCA1P1	iron-sulfur cluster assembly 1 homolog (S. cerevisiae) pseudogene 1 [Source:HGNC Symbol;Acc:33263]	5	62,072,704	62,073,090	-1	pseudogene	2
157	RP11-204C16.4		X	41,535,013	41,535,747	-1	pseudogene	1

158	RPS10P3	ribosomal protein S10 pseudogene 3 [Source:HGNC Symbol;Acc:23684]	9	90,631,215	90,631,708	1	pseudogene	1
159	RP6-159A1.2		X	65,176,131	65,177,038	-1	pseudogene	1
160	AC016739.2		2	177,065,636	177,065,980	-1	pseudogene	1
161	RP11-889L3.1		5	177,482,605	177,483,195	1	pseudogene	1
162	MORF4L1P1	mortality factor 4 like 1 pseudogene 1 [Source:HGNC Symbol;Acc:20400]	1	220,426,912	220,427,878	1	pseudogene	1
163	RP11-475C16.1		6	153,603,419	153,603,894	-1	pseudogene	1
164	RP1-256G22.1		6	5,609,460	5,610,428	-1	pseudogene	1
165	RP1-129L7.1		6	66,803,324	66,804,802	-1	pseudogene	1
166	RP11-203F10.6		1	204,315,904	204,316,386	-1	pseudogene	1
167	RP11-181C21.4		1	78,276,507	78,277,621	-1	pseudogene	2
168	RP1-217P22.2		6	38,730,681	38,732,470	1	pseudogene	1
169	RP3-425P12.4		6	25,140,212	25,141,649	1	pseudogene	2
170	RP1-292B18.1		6	151,420,665	151,421,069	1	pseudogene	1
171	KRT18P50	keratin 18 pseudogene 50 [Source:HGNC Symbol;Acc:33420]	6	96,438,983	96,440,225	1	pseudogene	1
172	RP1-6P5.2		6	128,960,333	128,960,798	-1	pseudogene	1
173	AC011498.2		19	4,477,048	4,479,576	1	pseudogene	1
174	RP11-565J7.3		1	212,224,829	212,225,371	-1	pseudogene	2
175	RP3-334F4.1		6	54,489,403	54,490,286	1	pseudogene	1
176	RP11-473I1.1		16	9,250,219	9,250,761	-1	pseudogene	2

177	RP11-572P18.1		10	122,114,177	122,114,718	-1	pseudogene	2
178	FAM45B	family with sequence similarity 45, member B [Source:HGNC Symbol;Acc:30886]	X	129,629,133	129,630,206	1	pseudogene	2
179	RP11-393I23.2		1	58,521,805	58,522,152	-1	pseudogene	1
180	POM121L10P	POM121 membrane glycoprotein-like 10, pseudogene [Source:HGNC Symbol;Acc:35448]	22	25,041,133	25,055,114	-1	pseudogene	2
181	KRT18P27	keratin 18 pseudogene 27 [Source:HGNC Symbol;Acc:33396]	13	90,882,638	90,883,936	1	pseudogene	1
182	EIF4BP3	eukaryotic translation initiation factor 4B pseudogene 3 [Source:HGNC Symbol;Acc:37936]	9	98,908,307	98,910,141	1	pseudogene	1
183	RP11-51O6.1		16	61,089,349	61,089,818	-1	pseudogene	2
184	KB-1592A4.13		22	21,495,913	21,496,660	-1	pseudogene	1
185	RP11-651P23.4		3	149,699,965	149,701,095	-1	pseudogene	2
186	AC011816.4		3	37,058,014	37,058,505	-1	pseudogene	1
187	AL391261.1		14	66,479,548	66,480,507	-1	pseudogene	1
188	SUMO2P3	SMT3 suppressor of mif two 3 homolog 2 (S. cerevisiae) pseudogene 3 [Source:HGNC Symbol;Acc:39013]	7	55,799,837	55,800,124	1	pseudogene	1
189	AD000091.1		19	15,722,467	15,722,934	1	pseudogene	1
190	GS1-184P14.2		X	24,447,690	24,448,037	-1	pseudogene	1
191	RP1-111B22.2		6	108,325,561	108,326,772	-1	pseudogene	1
192	BCRP1	breakpoint cluster region pseudogene 1 [Source:HGNC Symbol;Acc:39073]	22	24,655,897	24,659,732	-1	pseudogene	1

193	RP11-56A21.4		10	52,024,738	52,026,575	1	pseudogene	1
194	RP11-40H20.2		1	27,533,421	27,534,287	1	pseudogene	1
195	RP5-1118D24.2		1	12,280,849	12,281,166	-1	pseudogene	1
196	HSPA7	heat shock 70kDa protein 7 (HSP70B) [Source:HGNC Symbol;Acc:5240]	1	161,576,081	161,578,007	1	pseudogene	1
197	HMGB3P1	high mobility group box 3 pseudogene 1 [Source:HGNC Symbol;Acc:16240]	20	33,421,378	33,422,265	-1	pseudogene	1
198	TDGF3	teratocarcinoma-derived growth factor 3, pseudogene [Source:HGNC Symbol;Acc:11703]	X	109,764,358	109,764,924	1	pseudogene	1
199	AC007272.2		2	201,927,812	201,928,774	1	pseudogene	1
200	ATP5G2P1	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit C2 (subunit 9) pseudogene 1 [Source:HGNC Symbol;Acc:844]	1	209,441,143	209,441,568	-1	pseudogene	1
201	SIRPAP1	signal-regulatory protein alpha pseudogene 1 [Source:HGNC Symbol;Acc:9663]	22	30,938,523	30,940,292	1	pseudogene	2
202	RP13-258O15.1		X	91,931,518	91,931,953	-1	pseudogene	1
203	MEMO1P1	mediator of cell motility 1 pseudogene 1 [Source:HGNC Symbol;Acc:23274]	21	37,502,669	37,504,208	1	pseudogene	2
204	RP4-706A16.3		1	77,594,768	77,595,384	1	pseudogene	2
205	UQCRFS1P1	ubiquinol-cytochrome c reductase, Rieske iron-sulfur	22	40,271,293	40,272,119	1	pseudogene	2

		polypeptide 1 pseudogene 1 [Source:HGNC Symbol;Acc:12588]						
206	SMARCE1P2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1 pseudogene 2 [Source:HGNC Symbol;Acc:39732]	6	85,139,302	85,140,520	-1	pseudogene	1
207	RPL10AP6	ribosomal protein L10a pseudogene 6 [Source:HGNC Symbol;Acc:36504]	3	61,728,129	61,728,780	-1	pseudogene	2
208	TERF1P5	telomeric repeat binding factor (NIMA-interacting) 1 pseudogene 5 [Source:HGNC Symbol;Acc:39686]	13	19,254,634	19,255,846	1	pseudogene	2
209	TPI1P1	triosephosphate isomerase 1 pseudogene 1 [Source:HGNC Symbol;Acc:35449]	1	77,165,474	77,166,223	1	pseudogene	1
210	UPF3AP1	UPF3 regulator of nonsense transcripts homolog A (yeast) pseudogene 1 [Source:HGNC Symbol;Acc:30568]	17	16,648,945	16,650,075	-1	pseudogene	2
211	TMEM185B	transmembrane protein 185B (pseudogene) [Source:HGNC Symbol;Acc:18896]	2	120,979,500	120,980,552	-1	pseudogene	1
212	RPS7P10	ribosomal protein S7 pseudogene 10 [Source:HGNC Symbol;Acc:39903]	13	22,202,552	22,203,154	-1	pseudogene	1
213	FTH1P20	ferritin, heavy polypeptide 1 pseudogene 20 [Source:HGNC Symbol;Acc:37639]	2	181,737,594	181,738,141	-1	pseudogene	1
214	FTLP3	ferritin, light polypeptide pseudogene 3 [Source:HGNC Symbol;Acc:4000]	20	4,004,552	4,005,091	1	pseudogene	2
215	HNRNPA3P1	heterogeneous nuclear ribonucleoprotein A3 pseudogene 1	10	44,282,860	44,285,865	-1	pseudogene	2

		[Source:HGNC Symbol;Acc:13729]						
216	KB-1183D5.9		22	21,622,450	21,623,192	1	pseudogene	1
217	RP5-1068H6.3		20	4,610,073	4,610,865	-1	pseudogene	1
218	RPL41P1	ribosomal protein L41 pseudogene 1 [Source:HGNC Symbol;Acc:10356]	20	21,735,866	21,736,171	1	pseudogene	2
219	AC107983.4		17	18,476,045	18,476,415	1	pseudogene	2
220	RP11-543P15.1		12	3,320,775	3,321,096	-1	pseudogene	2
221	RP11-742N3.1		11	82,400,571	82,400,944	1	pseudogene	2
222	RP11-325P15.3	phosphodiesterase 4D interacting protein pseudogene (LOC728989), non-coding RNA [Source:RefSeq DNA;Acc:NR_024442]	1	146,490,895	146,596,107	-1	pseudogene	2
223	HCG4B	HLA complex group 4B (non-protein coding) [Source:HGNC Symbol;Acc:22919]	6	29,892,500	29,894,992	-1	pseudogene	2
224	RP11-282E4.1		9	66,664,325	66,665,950	1	pseudogene	1
225	FTH1P16	ferritin, heavy polypeptide 1 pseudogene 16 [Source:HGNC Symbol;Acc:3986]	11	77,445,520	77,446,071	-1	pseudogene	1
226	RP11-282K24.1		12	19,241,974	19,242,825	1	pseudogene	2
227	RP11-864N7.2		11	74,456,761	74,457,159	-1	pseudogene	1
228	ISCA1P6	iron-sulfur cluster assembly 1 homolog (S. cerevisiae) pseudogene 6 [Source:HGNC Symbol;Acc:38027]	2	129,276,362	129,276,751	-1	pseudogene	1
229	RP11-488L18.4		1	247,353,153	247,372,795	-1	pseudogene	2

230	XXyac-YX155 B6.2		1	147,552,868	147,566,198	1	pseudogene	2
231	NPM1P22	nucleophosmin 1 (nucleolar phosphoprotein B23, numatrin) pseudogene 22 [Source:HGNC Symbol;Acc:39679]	13	68,406,008	68,406,959	-1	pseudogene	1
232	CTA-481E9.1		16	18,026,576	18,027,133	1	pseudogene	1
233	ATF4P3	activating transcription factor 4 pseudogene 3 [Source:HGNC Symbol;Acc:788]	17	74,221,832	74,222,887	-1	pseudogene	1
234	GAPDHP1	glyceraldehyde-3-phosphate dehydrogenase pseudogene 1 [Source:HGNC Symbol;Acc:4159]	X	39,646,386	39,647,390	-1	pseudogene	1
235	PTMAP1	prothymosin, alpha pseudogene 1 (gene sequence 26) [Source:HGNC Symbol;Acc:9624]	6	30,601,409	30,601,675	1	pseudogene	1
236	EEF1A1P11	eukaryotic translation elongation factor 1 alpha 1 pseudogene 11 [Source:HGNC Symbol;Acc:3206]	1	96,912,486	96,913,874	1	pseudogene	1
237	RP11-298C3.2		X	25,663,839	25,664,444	-1	pseudogene	1
238	RP6-11O7.2		22	40,502,329	40,503,070	-1	pseudogene	1
239	RP11-313P13.3	postmeiotic segregation increased 2-like 2 pseudogene (PMS2L2), non-coding RNA [Source:RefSeq DNA;Acc:NR_003614]	7	72,476,064	72,555,749	1	pseudogene	5
240	AC009945.3		7	10,490,938	10,492,153	1	pseudogene	1
241	CTB-63M22.1		5	165,809,310	165,809,604	1	pseudogene	1
242	EIF4A1P10	eukaryotic translation initiation factor 4A1 pseudogene 10 [Source:HGNC Symbol;Acc:37930]	X	91,368,244	91,369,460	-1	pseudogene	1

243	ST13P13	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein) pseudogene 13 [Source:HGNC Symbol;Acc:38788]	10	104,975,466	104,976,557	1	pseudogene	1
244	UBE2V1P3	ubiquitin-conjugating enzyme E2 variant 1 pseudogene 3 [Source:HGNC Symbol;Acc:23983]	Y	3,734,347	3,734,763	-1	pseudogene	1
245	RP11-492M23.2		10	29,187,925	29,188,422	-1	pseudogene	2
246	RP11-110C15.4		3	185,135,283	185,136,566	-1	pseudogene	1
247	RP13-362E11.2		X	89,544,539	89,544,902	-1	pseudogene	1
248	DDX50P1	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 50 pseudogene 1 [Source:HGNC Symbol;Acc:18974]	2	32,426,669	32,429,068	1	pseudogene	1
249	RP11-94I2.1		1	146,011,091	146,025,308	1	pseudogene	1
250	RP4-710M16.1		1	56,673,239	56,674,128	-1	pseudogene	1
251	RP11-111F16.2		X	110,862,036	110,862,322	1	pseudogene	1
252	AC004797.1		17	47,500,970	47,502,298	-1	pseudogene	2
253	UBBP1	ubiquitin B pseudogene 1 [Source:HGNC Symbol;Acc:12464]	2	137,087,011	137,087,483	-1	pseudogene	1
254	TMSL6	thymosin-like 6 (pseudogene) [Source:HGNC Symbol;Acc:11888]	20	49,457,152	49,457,286	-1	pseudogene	1
255	RP11-331N16.1		1	243,138,307	243,138,386	1	pseudogene	1
256	RP11-632C17_		6	118,320,138	118,320,611	1	pseudogene	1

	_A.1							
257	FTH1P5	ferritin, heavy polypeptide 1 pseudogene 5 [Source:HGNC Symbol;Acc:3996]	6	50,880,425	50,880,969	-1	pseudogene	1
258	RP11-87M1.1		9	7,477,045	7,478,320	-1	pseudogene	1
259	FAM75C1	family with sequence similarity 75, member C1 [Source:HGNC Symbol;Acc:27846]	9	90,529,408	90,538,572	1	pseudogene	2
260	AC008738.2		19	33,790,853	33,793,430	-1	pseudogene	1
261	HMGN2P3	high mobility group nucleosomal binding domain 2 pseudogene 3 [Source:HGNC Symbol;Acc:33566]	16	26,043,860	26,044,132	-1	pseudogene	1
262	AC048382.5	golgin A6 family-like 5 (pseudogene) (GOLGA6L5), non-coding RNA [Source:RefSeq DNA;Acc:NR_003246]	15	85,047,738	85,060,078	-1	pseudogene	2
263	AC009245.3		7	137,406,731	137,407,592	1	pseudogene	1
264	RP11-320L13.2		X	70,182,814	70,183,140	1	pseudogene	1
265	PABPC1P3	poly(A) binding protein, cytoplasmic 1 pseudogene 3 [Source:HGNC Symbol;Acc:8560]	X	73,802,923	73,803,381	-1	pseudogene	1
266	U07000.1		22	23,605,323	23,606,541	1	pseudogene	1
267	RHOQP3	ras homolog gene family, member Q pseudogene 3 [Source:HGNC Symbol;Acc:37837]	2	130,970,443	130,971,063	-1	pseudogene	1
268	RP11-195E2.1		5	70,370,052	70,389,020	-1	pseudogene	1
269	CTGLF10P	centaurin, gamma-like family, member 10 pseudogene [Source:HGNC Symbol;Acc:23659]	10	46,174,140	46,195,980	1	pseudogene	1
270	RP11-96J23.1		15	83,040,957	83,041,655	-1	pseudogene	2

271	AC079250.1		2	47,917,855	47,918,385	-1	pseudogene	2
272	CTB-54D4.1		7	108,150,677	108,151,420	1	pseudogene	1
273	RP1-40G4P.1		22	27,282,872	27,283,511	-1	pseudogene	1
274	TIMM8AP1	translocase of inner mitochondrial membrane 8 homolog A (yeast) pseudogene 1 [Source:HGNC Symbol;Acc:17802]	2	162,933,867	162,934,161	-1	pseudogene	1
275	RP11-118B22.3		12	9,392,582	9,392,927	-1	pseudogene	1
276	AC013272.3		2	96,679,478	96,687,356	1	pseudogene	2
277	CTB-83D3.1	protein phosphatase 1, regulatory (inhibitor) subunit 2 pseudogene 3 (PPP1R2P3), non-coding RNA [Source:RefSeq DNA;Acc:NR_002168]	5	156,277,549	156,279,539	1	pseudogene	2
278	ANXA2P2	annexin A2 pseudogene 2 [Source:HGNC Symbol;Acc:539]	9	33,624,223	33,625,532	1	pseudogene	2
279	ST13P4	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein) pseudogene 4 [Source:HGNC Symbol;Acc:18487]	13	50,746,225	50,747,317	1	pseudogene	1
280	RP11-146N23.1		9	19,200,333	19,201,044	-1	pseudogene	1
281	RP11-311P8.2		X	74,546,607	74,547,834	-1	pseudogene	1
282	RP11-54B9.2		1	166,716,594	166,717,899	1	pseudogene	1
283	EEF1B4	eukaryotic translation elongation factor 1 beta 4 (pseudogene) [Source:HGNC Symbol;Acc:3210]	X	24,806,509	24,807,186	-1	pseudogene	1
284	GTF2IP1	general transcription factor IIi, pseudogene 1 [Source:HGNC Symbol;Acc:4660]	7	74,602,783	74,653,438	-1	pseudogene	2

285	RPL3P4	ribosomal protein L3 pseudogene 4 [Source:HGNC Symbol;Acc:19805]	14	99,439,216	99,439,638	-1	pseudogene	1
286	RHOQP2	ras homolog gene family, member Q pseudogene 2 [Source:HGNC Symbol;Acc:37836]	2	132,218,533	132,219,150	1	pseudogene	2
287	GS1-257G1.1		X	14,262,325	14,263,544	1	pseudogene	2
288	RP11-396K3.1		7	72,569,033	72,619,657	1	pseudogene	2
289	RP4-781L3.1		1	25,451,544	25,451,945	-1	pseudogene	2
290	EIF3FP3	eukaryotic translation initiation factor 3, subunit F pseudogene 3 [Source:HGNC Symbol;Acc:37625]	2	58,478,575	58,479,703	1	pseudogene	2
291	PMS2P9	postmeiotic segregation increased 2 pseudogene 9 [Source:HGNC Symbol;Acc:9135]	7	76,669,261	76,673,093	1	pseudogene	1
292	EEF1A1P6	eukaryotic translation elongation factor 1 alpha 1 pseudogene 6 [Source:HGNC Symbol;Acc:3201]	7	22,550,231	22,551,618	-1	pseudogene	2
293	AC009963.3		2	108,519,453	108,519,590	-1	pseudogene	1
294	AC007969.5		2	172,373,749	172,374,232	-1	pseudogene	2
295	CTC-575D19.1		5	168,043,317	168,044,059	1	pseudogene	2
296	RP11-169K16.7	ubiquinol-cytochrome c reductase hinge protein-like (UQCRHL), mRNA [Source:RefSeq DNA;Acc:NM_001089591]	1	16,133,679	16,134,194	-1	pseudogene	2
297	RP11-380N8.4		13	26,688,547	26,728,313	-1	pseudogene	1
298	AL049757.3		22	43,172,412	43,173,302	-1	pseudogene	1
299	RPL13AP6	ribosomal protein L13a pseudogene 6 [Source:HGNC	10	112,696,380	112,696,991	-1	pseudogene	1

		Symbol;Acc:23737]						
300	H2BFS	H2B histone family, member S [Source:HGNC Symbol;Acc:4762]	21	44,985,120	44,985,500	1	pseudogene	1
301	ST13P18	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein) pseudogene 18 [Source:HGNC Symbol;Acc:38861]	X	92,542,732	92,543,758	-1	pseudogene	1
302	RPS26P47	ribosomal protein S26 pseudogene 47 [Source:HGNC Symbol;Acc:36368]	13	101,192,089	101,192,528	-1	pseudogene	2
303	RP11-392P7.2		12	13,028,433	13,029,044	1	pseudogene	2
304	AC073072.7		7	22,813,265	22,813,612	-1	pseudogene	1
305	CDRT15P	CMT1A duplicated region transcript 15 pseudogene [Source:HGNC Symbol;Acc:33168]	17	14,095,430	14,095,681	-1	pseudogene	1
306	API5P1	API5 pseudogene 1 [Source:HGNC Symbol;Acc:595]	X	115,238,174	115,239,683	-1	pseudogene	1
307	RP11-422N19.3		9	79,013,615	79,014,502	1	pseudogene	1
308	NUS1P3	nuclear undecaprenyl pyrophosphate synthase 1 homolog (S. cerevisiae) pseudogene 3 [Source:HGNC Symbol;Acc:30934]	13	24,902,349	24,903,210	1	pseudogene	2
309	HMGN2P5	high mobility group nucleosomal binding domain 2 pseudogene 5 [Source:HGNC Symbol;Acc:33568]	15	30,022,917	30,023,189	1	pseudogene	1
310	NUS1P2	nuclear undecaprenyl pyrophosphate synthase 1 homolog (S. cerevisiae) pseudogene 2 [Source:HGNC	13	23,489,790	23,490,662	1	pseudogene	1

		Symbol;Acc:38473]						
311	AC144530.1		3	197,577,271	197,577,825	-1	pseudogene	1
312	RP11-36D19.5		10	82,006,975	82,007,439	-1	pseudogene	1
313	AC008132.12		22	18,821,207	18,821,951	1	pseudogene	1
314	RPS3AP6	ribosomal protein S3A pseudogene 6 [Source:HGNC Symbol;Acc:18630]	15	60,060,543	60,061,347	1	pseudogene	2
315	MORF4	mortality factor 4 [Source:HGNC Symbol;Acc:15773]	4	174,537,087	174,538,057	-1	pseudogene	2
316	RP11-3P17.3		3	161,146,915	161,147,398	-1	pseudogene	2
317	FABP5P7	fatty acid binding protein 5 pseudogene 7 [Source:HGNC Symbol;Acc:31070]	11	59,548,550	59,549,224	-1	pseudogene	2
318	FTH1P2	ferritin, heavy polypeptide 1 pseudogene 2 [Source:HGNC Symbol;Acc:3989]	1	228,823,162	228,823,574	1	pseudogene	2
319	RP3-497J21.1		6	167,114,582	167,116,327	-1	pseudogene	1
320	RPL24P2	ribosomal protein L24 pseudogene 2 [Source:HGNC Symbol;Acc:16597]	20	21,095,364	21,095,838	1	pseudogene	1
321	RP11-179G5.4		1	160,287,191	160,287,496	1	pseudogene	1
322	CHCHD2P6	coiled-coil-helix-coiled-coil-helix domain containing 2 pseudogene 6 [Source:HGNC Symbol;Acc:39590]	1	15,931,077	15,931,538	1	pseudogene	2
323	KRT8P17	keratin 8 pseudogene 17 [Source:HGNC Symbol;Acc:33369]	X	58,011,120	58,012,567	1	pseudogene	1
324	SUMO2P1	SMT3 suppressor of mif two 3 homolog 2 (S. cerevisiae) pseudogene 1 [Source:HGNC Symbol;Acc:13985]	6	29,603,837	29,604,120	-1	pseudogene	1

325	RP11-504H5.1		18	72,057,119	72,057,532	-1	pseudogene	1
326	RP11-332O19.5	rcRPE (LOC729020), mRNA [Source:RefSeq DNA;Acc:NM_001143909]	10	105,005,644	105,007,773	1	pseudogene	2
327	RP11-480I12.2		1	202,736,357	202,736,548	-1	pseudogene	1
328	RP11-678B3.2		7	53,929,926	53,931,032	1	pseudogene	1
329	RP1-101G11.2		22	34,985,077	34,987,270	-1	pseudogene	1
330	CTD-2124B20.1		18	6,462,143	6,463,009	-1	pseudogene	1
331	AC005822.1		17	16,520,162	16,521,195	-1	pseudogene	1
332	RP11-1198D22.1		5	70,516,183	70,555,065	-1	pseudogene	1
333	POU5F1P3	POU class 5 homeobox 1 pseudogene 3 [Source:HGNC Symbol;Acc:9222]	12	8,286,376	8,287,449	-1	pseudogene	2
334	NUS1P1	nuclear undecaprenyl pyrophosphate synthase 1 homolog (S. cerevisiae) pseudogene 1 [Source:HGNC Symbol;Acc:38472]	X	47,372,001	47,372,882	-1	pseudogene	1
335	AC018890.1		2	175,584,505	175,585,526	1	pseudogene	2
336	PA2G4P2	proliferation-associated 2G4 pseudogene 2 [Source:HGNC Symbol;Acc:16531]	20	12,360,704	12,361,903	1	pseudogene	1
337	AC006978.6		7	30,410,987	30,412,357	1	pseudogene	2
338	AC012379.1		15	49,170,346	49,170,937	1	pseudogene	1
339	COX7CP1	cytochrome c oxidase subunit VIIc pseudogene 1	13	49,761,615	49,761,806	-1	pseudogene	1

		[Source:HGNC Symbol;Acc:2293]						
340	FABP5P2	fatty acid binding protein 5 pseudogene 2 [Source:HGNC Symbol;Acc:31060]	13	52,540,878	52,541,285	1	pseudogene	1
341	RP11-334A14.2		1	53,458,868	53,459,374	-1	pseudogene	2
342	RP11-175B9.3		1	202,440,937	202,441,258	-1	pseudogene	2
343	RPL13AP5	ribosomal protein L13a pseudogene 5 [Source:HGNC Symbol;Acc:23736]	10	98,510,045	98,510,675	1	pseudogene	2
344	HNRNPA3P5	heterogeneous nuclear ribonucleoprotein A3 pseudogene 5 [Source:HGNC Symbol;Acc:39774]	13	66,362,064	66,362,896	-1	pseudogene	2
345	RANP1	RAN, member RAS oncogene family pseudogene 1 [Source:HGNC Symbol;Acc:21631]	6	30,453,717	30,454,367	1	pseudogene	2
346	BZW1P1	basic leucine zipper and W2 domains 1 pseudogene 1 [Source:HGNC Symbol;Acc:31378]	3	172,143,640	172,144,879	-1	pseudogene	1
347	AL021920.1		1	17,012,116	17,012,550	-1	pseudogene	1
348	RP11-373A9.1		9	72,832,111	72,832,584	1	pseudogene	1
349	RP11-446E9.1		8	56,962,597	56,963,766	-1	pseudogene	1
350	FABP5P1	fatty acid binding protein 5 pseudogene 1 [Source:HGNC Symbol;Acc:31059]	13	73,674,713	73,675,123	-1	pseudogene	2
351	AC018738.2		2	232,120,810	232,121,019	1	pseudogene	1
352	FTH1P11	ferritin, heavy polypeptide 1 pseudogene 11 [Source:HGNC Symbol;Acc:3981]	8	82,433,921	82,434,467	-1	pseudogene	1

353	RP11-603J24.7		12	56,374,216	56,374,819	-1	pseudogene	2
354	RP1-138A5.1		X	86,958,277	86,959,307	-1	pseudogene	2
355	RP13-996F3.1		15	82,664,416	82,748,784	-1	pseudogene	3
356	AC084031.1		2	231,589,858	231,590,118	-1	pseudogene	1
357	POU5F1P4	POU class 5 homeobox 1 pseudogene 4 [Source:HGNC Symbol;Acc:33310]	1	155,402,969	155,404,046	1	pseudogene	2
358	RP11-274E7.2		5	97,549,106	97,549,825	1	pseudogene	1
359	RP11-59K5.1		11	56,457,941	56,458,199	1	pseudogene	1
360	RPL9P7	ribosomal protein L9 pseudogene 7 [Source:HGNC Symbol;Acc:30335]	X	23,854,761	23,855,459	-1	pseudogene	2
361	RP11-464D20.2		17	60,593,682	60,594,128	-1	pseudogene	1
362	RP11-986E7.1		14	95,192,625	95,193,512	1	pseudogene	1
363	RP11-771F20.1		8	104,780,509	104,781,257	1	pseudogene	1
364	RP11-16F15.2		11	9,681,946	9,682,503	1	pseudogene	2
365	RP11-254B13.1		12	68,946,775	68,947,662	-1	pseudogene	2
366	FABP5P11	fatty acid binding protein 5 pseudogene 11 [Source:HGNC Symbol;Acc:19328]	22	17,075,979	17,076,375	-1	pseudogene	1
367	RPS2P5	ribosomal protein S2 pseudogene 5 [Source:HGNC Symbol;Acc:31386]	12	118,683,877	118,684,807	1	pseudogene	2
368	RP11-36C20.1		5	33,162,285	33,162,913	1	pseudogene	1
369	RP11-111F10.2		3	155,705,093	155,705,965	1	pseudogene	1

370	RP11-352G18.1		15	66,671,675	66,672,253	-1	pseudogene	1
371	RP11-179A9.1		14	95,267,186	95,267,800	-1	pseudogene	1
372	PMS2P11	postmeiotic segregation increased 2 pseudogene 11 [Source:HGNC Symbol;Acc:9125]	7	76,640,868	76,645,170	1	pseudogene	1
373	PSMC1P1	proteasome (prosome, macropain) 26S subunit, ATPase, 1 pseudogene 1 [Source:HGNC Symbol;Acc:30147]	3	68,684,856	68,686,175	1	pseudogene	1
374	RP11-490G8.1		12	95,861,173	95,861,637	-1	pseudogene	1
375	RPL7AP6	ribosomal protein L7a pseudogene 6 [Source:HGNC Symbol;Acc:19785]	14	70,352,057	70,352,869	-1	pseudogene	2
376	RP11-100N21.1		4	47,708,389	47,709,004	1	pseudogene	1
377	RP11-234A1.1		3	101,295,333	101,295,791	-1	pseudogene	1
378	RP11-439L12.1		14	90,833,041	90,833,523	-1	pseudogene	1
379	AC018462.3		2	62,373,548	62,374,288	1	pseudogene	1
380	RP11-734J24.1		8	34,731,677	34,732,151	1	pseudogene	1
381	RP11-408P14.1		4	66,439,177	66,440,046	1	pseudogene	1
382	RP11-330L19.1		15	64,885,178	64,885,525	1	pseudogene	1
383	RP11-627K11.1		12	31,404,926	31,405,531	-1	pseudogene	1
384	CTB-55B8.1		5	16,902,403	16,902,750	-1	pseudogene	1
385	RP11-598F17.1		3	96,068,398	96,069,783	1	pseudogene	1

386	RP11-193H5.2		1	238,090,131	238,090,905	1	pseudogene	1
387	CTB-33G10.1		19	50,222,755	50,223,338	-1	pseudogene	1
388	RP11-16F15.1		11	9,627,689	9,628,156	1	pseudogene	1
389	RP11-227J5.3		3	160,170,386	160,171,191	1	pseudogene	1
390	RP11-50D9.1		4	128,733,875	128,734,355	-1	pseudogene	1
391	RP11-120B7.1		5	107,929,310	107,929,668	1	pseudogene	1
392	PKD1P1	polycystic kidney disease 1 (autosomal dominant) pseudogene 1 [Source:HGNC Symbol;Acc:30065]	16	16,404,198	16,444,435	1	pseudogene	5
393	RP11-425L10.1		11	46,450,163	46,450,700	-1	pseudogene	2
394	RP11-526L22.1		5	76,878,201	76,878,947	1	pseudogene	1
395	RP11-466H18.1		11	16,996,240	16,996,560	-1	pseudogene	1
396	AC046134.1		3	139,328,511	139,328,780	-1	scRNA_pseudogene	1
397	RP11-20O24.4		1	109,534,878	109,535,489	1	pseudogene	2
398	TNXA	tenascin XA (pseudogene) [Source:HGNC Symbol;Acc:11975]	6	31,976,391	31,980,249	-1	pseudogene	1
399	AL691432.2		1	1,571,100	1,588,935	-1	pseudogene	5
400	ANP32C	acidic (leucine-rich) nuclear phosphoprotein 32 family, member C [Source:HGNC Symbol;Acc:16675]	4	165,118,159	165,118,863	-1	pseudogene	1
401	DSTNP2	destrin (actin depolymerizing factor) pseudogene 2 [Source:HGNC Symbol;Acc:34546]	12	6,994,062	6,994,848	1	pseudogene	1
402	AC003029.1		12	112,277,571	112,279,211	-1	pseudogene	1

403	ACTN3	actinin, alpha 3 [Source:HGNC Symbol;Acc:165]	11	66,313,866	66,330,800	1	polymorphic_pseudogene	3
404	AC004067.4		4	110,594,999	110,595,280	1	pseudogene	1
405	CTB-36O1.6		5	134,260,338	134,262,149	-1	pseudogene	1
406	AC091981.1		5	114,725,340	114,728,067	-1	pseudogene	1
407	RP11-548H18.2		4	120,312,986	120,316,490	-1	pseudogene	1
408	AC004066.1		4	106,405,855	106,407,237	1	pseudogene	1
409	RP11-65F13.1		5	93,019,034	93,019,373	-1	pseudogene	1
410	ZNF286B	zinc finger protein 286B [Source:HGNC Symbol;Acc:33241]	17	18,561,742	18,585,572	-1	pseudogene	2
411	CTD-2197O4.1		5	14,797,234	14,798,509	-1	pseudogene	1
412	RP11-721G13.1		4	79,102,193	79,103,200	-1	pseudogene	1
413	RP11-159F24.4		5	43,495,175	43,496,556	-1	pseudogene	1
414	RP11-284E20.1		4	46,725,793	46,726,625	-1	pseudogene	2
415	CTD-2165H16.1		5	14,652,050	14,653,438	-1	pseudogene	1
416	AC061975.3		17	26,554,332	26,554,718	1	pseudogene	1
417	GMPSP1	guanine monphosphate synthetase pseudogene 1 [Source:HGNC Symbol;Acc:39428]	4	8,176,148	8,176,721	1	pseudogene	1
418	CTD-2120M21.1		5	39,719,086	39,721,615	-1	pseudogene	1

419	RP11-365D23.1		1	214,778,813	214,782,183	-1	pseudogene	1
420	RNPS1P1	RNA binding protein S1, serine-rich domain pseudogene 1 [Source:HGNC Symbol;Acc:39525]	4	11,373,599	11,374,516	1	pseudogene	1
421	TUBB4Q	tubulin, beta polypeptide 4, member Q [Source:HGNC Symbol;Acc:12413]	4	190,903,678	190,906,026	-1	pseudogene	1
422	PPIAP11	peptidylprolyl isomerase A (cyclophilin A) pseudogene 11 [Source:HGNC Symbol;Acc:9263]	5	81,305,421	81,305,913	-1	pseudogene	1
423	RP11-497H16.8		5	69,790,927	69,796,730	-1	pseudogene	1
424	RP11-730G20.2		8	75,515,479	75,516,479	-1	pseudogene	1
425	RNF5P1	ring finger protein 5 pseudogene 1 [Source:HGNC Symbol;Acc:31053]	8	38,458,179	38,458,718	-1	pseudogene	1
426	CTB-79E8.3		5	172,083,525	172,083,716	-1	pseudogene	1
427	RP11-1415C14.3		5	69,515,745	69,580,686	-1	pseudogene	2
428	HMGN1P38	high mobility group nucleosome binding domain 1 pseudogene 38 [Source:HGNC Symbol;Acc:39422]	15	93,254,977	93,256,105	1	pseudogene	2
429	RP11-1K11.1		8	4,644,854	4,646,106	1	pseudogene	1
430	KRT8P3	keratin 8 pseudogene 3 [Source:HGNC Symbol;Acc:31056]	8	62,490,779	62,492,230	1	pseudogene	2
431	CSNK2A1P	casein kinase 2, alpha 1 polypeptide pseudogene	11	11,373,498	11,374,668	-1	pseudogene	1

		[Source:HGNC Symbol;Acc:2458]						
432	RP11-676M6.1		11	127,810,829	127,811,552	-1	pseudogene	1
433	RP11-113D6.8		11	18,210,547	18,211,168	1	pseudogene	1
434	AC005013.1	TLR4 interactor with leucine-rich repeats (TRIL), mRNA [Source:RefSeq DNA;Acc:NM_014817]	7	28,992,974	28,998,029	-1	pseudogene	1
435	AL137067.1		9	102,067,394	102,068,893	1	pseudogene	1
436	AL590369.1		9	132,500,629	132,501,902	-1	pseudogene	1
437	AL139385.1		13	111,291,555	111,292,340	1	pseudogene	1
438	AC005000.1		X	114,953,100	114,953,685	-1	pseudogene	1
439	AC093734.1		7	1,580,047	1,581,609	1	pseudogene	1
440	AC004889.2	CTAGE family, member 4 (CTAGE4), mRNA [Source:RefSeq DNA;Acc:NM_198495]	7	143,880,548	143,883,173	1	pseudogene	1
441	AC091304.4		15	28,560,798	28,561,102	1	pseudogene	1
442	AC138123.2		12	93,477,374	93,477,451	-1	pseudogene	1
443	RP11-785H5.1		12	11,323,963	11,324,164	1	pseudogene	1
444	AC093668.4		7	102,120,179	102,121,095	-1	pseudogene	1
445	AC007401.1		2	36,816,784	36,818,696	-1	pseudogene	1
446	AL021707.2		22	39,127,778	39,128,531	1	pseudogene	1
447	RP11-153M3.1		12	56,904,786	56,906,487	1	pseudogene	1
448	RP11-1136G11.6		12	53,547,056	53,548,268	1	pseudogene	1

Appendix II – Sequences of *ψPPM1K*

The sequences of *ψPPM1K* and *ψPPM1K*-derived esiRNAs.

Red-color: esiRNA1; pink-color: esiRNA2.

>ψPPM1K (locate: chr4 (bp 89,398,481-89,398,959); length: 474bp; GC content: 54.07%)

ATGGAGTCTCACTCTGTCACACAGGCTGGAGTACAGTGGTGCGGTCT
CGGCTCCCTGCAACCTCTGCCTCCCGGGTTCAAACGATTCTTCTGCCT
CAGCCTCCTGAGTAGCTGGGACTACAGGCGCCGTGCCAC**CACCCCTGGCT**
AATTTTTGTATTTTTTTTTTTTTTTTTGAGAAGGAGTCTCTCTCTGTCA
CCCAGGCTGGAGTGCAGTGGCACGATCTCGGCTCACTGCAAGCTGCG
CCTCCTGGGTTCACGCCATTCTCCTGCCTC**AGCCTCCCGAGTAGCTGGA**
ATTACAGGCGTCCGCCACCGTGCCCGGCTAATTTTTTTGTATTTTTAGT
AGAGATGGGGTTTTGCCATATTGGCCAGGCTGGTCTCGAACTCCTGAC
CTCAAGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGG
CATGAGCCACCGCACCTGGCCAGATCTTTGTATGTCTTAA

>ψPPM1K-specific precursor esiRNA1 (24-144nt)

GGCTGGAGTACAGTGGTGCGGTCTCGGCTCCCTGCAACCTCTGCCTC
CCGGGTTCAAACGATTCTTCTGCCTCAGCCTCCTGAGTAGCTGGACT
ACAGGCGCCGTGCCAC**CACCCCTGGCT**

>ψPPM1K-specific precursor esiRNA2 (170-273nt)

TTGAGAAGGAGTCTCTCTCTGTCACCCAGGCTGGAGTGCAGTGGCAC
GATCTCGGCTCACTGCAAGCTGCGCCTCCTGGGTTCACGCCATTCTCCT
GCCTC**AG**

Curriculum Vitae

Wen-Ling Chan (詹雯玲)

Centre of RNA Biology and Clinical Application
China Medical University Hospital
2, Yuh-Der Road,
Taichung ,Taiwan.
Tel: +886-4-22052121 ext.2008
Email: sophia.wlchan@msa.hinet.net

Date of Birth: October 29, 2073 **Nationality: Taiwan**

Education and Qualifications

2005-2012	Ph.D., Institute of Bioinformatics and Systems Biology, National Chiao Tung University
2002-2004	M.S., Computer Sciences, National Chung Hsing University

Work Experiences

2005/9~2012	Institute of Bioinformatics and Systems Biology, National Chiao Tung University (Part time of research assistant)
1999/2~2004/6	Department of Molecular Medicine, China Medical University Hospital (research assistant)
1991/8~1999/1	Department of Entomology, National Chung Hsing University (research assistant)

Skills

- Bioinformatics and Systems Biology
- Bio-Chip experimental techniques and data analyses
- Cell culture
- siRNA transfection

Interests and Activities

Non-coding RNA Research

- Pseudogene: Functional identification of transcribed pseudogenes in human.

- Long non-coding RNA (lncRNA): The relationship between very long non-coding RNA (vlncRNA) and human disease.
- miRNA: miRNA-target interaction

Drug Discovery

- Drug regulatory of alternative splicing mechanism
- Drug selection for cancer therapy

Bio-chip Analysis

- Genome chip: Identification of differentially expression profiles in variety of cancers.
- miRNA chip: Identification of bio-marker for cancer therapy.
- Exon chip: Drug modulates oncogenic RNA alternative splicing to devitalize human cancer cells.
- CGH chip: Identification of bio-marker for cancer therapy and diagnosis.
- SNP chip: Identification of disease-related SNP.

Deep Sequencing Data

- sRNA libraries of deep sequencing data analyses
- Genome re-sequencing of novel Helicobacter Pylori strains
- RNA-Seq analyses of chromatin interaction
- Exome Sequencing analyses of various cancers

Publications

***corresponding author**

1. **Chan WL**, Chang YS, Yang WK, *Huang HD, *Chang JG. Very long non-coding RNA and human disease. BioMedicine 2012 *In press* (http://dx.doi.org/10.1016/j.biomed.2012.10.001).
2. **Chan WL**, Yang WK, *Huang HD, *Chang JG. pseudoMap: a novel method and comprehensive resource for identification of siRNA-mediated mechanisms in human transcribed pseudogenes. Database: The Journal of Biological Databases and Curation *2*012 minor revision (IF 3.179).
3. **Chan WL**, You CY, Yang WK, Huang SY, Chang YS, Chiu CC, Yeh KT, *Huang HD, *Chang JG. Transcribed pseudogene *ψPPM1K* generates endogenous siRNA to suppress oncogenic cell growth in hepatocellular carcinoma. Nucleic Acids Res 2012 minor revision (IF 8.026).

4. Lin YH, Lee CC, **Chan WL**, Chang WH, Wu YC, Chang JG. 16-Hydroxycleroda-3,13-dien-15,16-olide deregulates PI3K and Aurora B activities that involve in cancer cell apoptosis. Toxicology 285: 72-80, 2011. (IF 3.241)
5. Hsu SD, Lin FM, Wu MY, Liang C, Huang WC, **Chan WL**, Tsai WT, Chen GZ, Lee CJ, Chiu CM, Chien CH, Wu MC, Huang CY, Tsou AP, Huang HD. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res 29: D163-169, 2011. (IF 7.479)
6. Chang JG, Yang DM, Chang WH, Chow LP, **Chan WL**, Lin HH, Huang HD, Chang YS, Hung CH, *Yang WK. Small Molecule Amiloride modulates oncogenic RNA alternative splicing to devitalize human cancer cells. PLoS One 6 (6): e18643, 2011. (IF 4.351).
7. **Chan WL**, Chang JG, Chen YF, Chan YK, *Chu YP. Identification of mouse mslp2 gene from EST databases by repeated searching, comparison, and assembling. Comput Biol 36 (1) :101-8, 2006.
8. Shih MC, Peck K, **Chan WL**, Chu YP, Chen JC, Tsai CH, *Chang JG. SARS-CoV infection was from at least two origins in the Taiwan area. Intervirology.48 (2-3):124-32, 2005 (IF 2.34).
9. Yeh KT, Yang MY, Liu TC, Chen JC, **Chan WL**, Lin SF, *Chang JG. Abnormal expression of period 1 (PER1) in endometrial carcinoma. J Pathol. 206(1):111-20, 2005 (IF 5.759).
10. Chen ST, Lin SY, Yeh KT, Kuo SJ, **Chan WL**, Chu YP, *Chang JG. Mutational, epigenetic and expressional analyses of caveolin-1 gene in breast cancers. Int J Mol Med. 14(4):577-82, 2004 (IF 1.854).
11. *Chang JG, Shih MC, Liu SC, Chen CM, **Chan WL**, Lee TP, Peng CT. Hb G-Honolulu [alpha30(B11)Glu-->Gln (alpha2)], Hb J-Meinung [beta56(D7)Gly-->Asp], and beta-thalassemia [codons 41/42 (-TCTT)] in a Taiwanese family. Hemoglobin. 26(3):325-8, 2002 (IF 1.304).
12. Shin MC, Chen CM, Liu SC, Huang CH, Lee TP, **Chan WL**, *Chang JG. Hb Ube-2 in a Taiwanese subject: an A-->G substitution at codon 68 of the alpha2-globin gene. Hemoglobin. 26(1):99-101, 2002 (IF 1.304).
13. *Chang JG, Shih MC, Liu SC, Chen CM, **Chan WL**, Peng CT. Hb G-Chinese: a G-->C substitution at codon 30 of the alpha2-globin gene creates a PstI cutting site. Hemoglobin. 26(1):95-7, 2002 (IF 1.304).
14. *Chang JG, Liu HC, Shih MC, Liu SC, **Chan WL**, Tsai FJ. Unstable Hb Perth in a Taiwanese subject: a T-->C substitution at codon 32

of the beta-globin gene creates an MspI site. Hemoglobin. 26(1):91-4, 2002 (IF 1.304).

15. *Chang JG, Shih MC, Liu SC, **Chan WL**, Peng CT. Hb Manitoba in a Taiwanese family: a C-->A substitution at codon 102 of the alpha2-globin gene. Hemoglobin. 25(4):437-9, 2001 (IF 1.304).