Original Article

Construction of Pooled Oocyte Expression Profiles of Rhesus Monkey and Mouse for Concurrent Meta-analyses

Kaj Chokeshaiusaha[1] Denis Puthier[2] Catherine Nguyen[2] Thanida Sananmuang[1*]

Abstract

Concurrent gene expression profiling meta-analysis of *in vitro* and *in vivo* matured oocytes among mammals can provide crucial knowledge to assist reproductive technologies. Due to the lack of methodology to prepare oocyte datasets for such analysis, we illustrated the procedures to merge *in vitro* and *in vivo* matured oocyte expression profiling datasets of rhesus monkey (*Macaca mulatta*) and mouse (*Mus musculus*). Datasets acquired from both species were pooled together based on types of their orthologous genes. To determine the feasibility of constructed pooled data, top orthologous genes differentially expressed between *in vitro* and *in vivo* oocytes were identified by Linear models and empirical Bayes methods with 500 generated learning datasets (FDR<0.01). Several clustering algorithms were then applied for oocyte sample clustering using the acquired differentially expressed genes. Gene enrichment analysis to determine biological processes associated with the differentially expressed genes was performed using DAVID Bioinformatics Resources 6.7. The results revealed successful construction of pooled oocyte expression profiles of monkey and mouse, and the pooled datasets used for subsequent analyses consisted of 10,214 one-to-one orthologous genes. With total selected 100 differentially expressed genes, oocyte clustering results revealed the correct clustering of *in vivo* and *in vitro* oocyte samples. Interestingly, enrichment analysis revealed association of several differentially expressed genes with maturation and developmental process of oocytes. Of note, the acquired results strongly suggested the feasibility of the prepared data, and its preparation's methodology. Hopefully, this approach would be beneficial for cross-species gene expression profiling analyses of several mammalian oocytes in the future.

Keywords: concurrent meta-analysis, mouse, oocytes, pooled expression profiles, rhesus monkey

[1]*Faculty of Veterinary Medicine, Rajamangala University of Technology Tawan-OK 43 Moo 6 Bangpra, Sriracha Chonburi 20110 Thailand*

[2]*Laboratoire TAGC/INSERM U1090, Parc Scientifique de Luminy case 928, 163, avenue de Luminy, Marseille cedex 09 13288, France*

***Correspondence**: t.sananmuang@gmail.com*

Introduction

Several assisted reproductive technologies (ARTs), including well-recognized *in vitro* fertilization (IVF) and intracytoplasmic sperm injection (ICSI) have been developed with great attempts to aid fertility of both livestock and endangered species for decades. Since the success rates of these techniques greatly depended on the quality of applied oocytes, numbers microarray expression profiling studies have been performed in order to identify important biomarkers associated with the oocyte quality. In fact, determination of molecular differences between *in vitro* and *in vivo* matured oocytes is among the most famous topic in oocyte gene expression profiling studies (Lee, et al., 2008; Mamo, et al., 2011; O'Shea, et al., 2012).

The competency of oocyte to develop into embryo greatly depends on the oocyte quality (Eichenlaub-Ritter and Peschke, 2002; Sananmuang, et al., 2011; O'Shea, et al., 2012; Sananmuang, et al., 2013). By means of this, a classical objective to optimize *in vitro* culture condition is to obtain mature oocytes with quality equivalent to those developed *in vivo*. According to gene expression profiling comparisons between *in vitro* and *in vivo* matured oocytes, several candidate genes and cell metabolisms were suggested for improving *in vitro* culture condition (Lee, et al., 2008; Mamo, et al., 2011; O'Shea, et al., 2012). However, the use of such knowledge only limited to animal species used in each of studies raising doubts and uncertainty to apply them among other mammalian species. Of note, oocyte gene expression profiling comparison requires deep understanding of transcriptions among analyzed species. With excellently clarified transcriptome knowledge, mouse (*Mus musculus*) and rhesus monkey (*Macaca mulatta*) thus become the excellent models to initiate such approach (Lu, et al., 2009).

Cross-species meta-analysis has been continuously recognized as a novel tool to compare gene expression profiling of various cell types among species including oocytes (Lu, et al., 2009). Since cross-species meta-analysis aim to compare microarray expression datasets across organisms, both conserved and unique differentially expressed genes can be identified among analyzed organisms. Cross-species gene expression profiling meta-analysis is archived by comparing the orthologous genes (orthologs), which are genes evolved from common ancestral genes by specification among animal species (Lu, et al., 2009). Since cross-species analysis relies on orthologous gene expression profiles across organism species, an increase of multi-species samples is feasible along with extraction of some common signatures among analyzed species. Despite such outstanding benefits, cross-species analysis is still formidable in practice due to the dynamic and noisy nature of expression profiling data. To deal with such limitations, three major strategies are normally adopted with varied deliberations to be considered (Lu, et al., 2009). These strategies are 1. Individual species analysis with post-processing approach, 2. Cross-species microarray hybridization, and 3. Concurrent analysis of pooled expression profiles.

Individual species analysis allows simple discovery of overlapped differentially expressed genes of each individual organism. Though considered the most flexible one, the approach usually provides inconsistent outcomes with considerable biases from probe effects (Bergmann, et al., 2004; Lu, et al., 2009). Cross-species microarray hybridization strategy, on the contrary is lesser biased according the use of same probes to detect orthologous gene expressions of all species under study. Unfortunately, the technique is considered non-standard by which greatly limits its application and popularity (Bar-Or, et al., 2007; Lu, et al., 2009). Finally, construction of concurrent microarray expression datasets for combined analysis should be considered as later developed strategy. It beneficially retains advantages of using species-specific datasets from various microarray platforms, while concurrently analyzed them together after reducing probe biases. Though these allow flexible experimental designs, the limited knowledge of gene orthology and diminished statistic test's sensitivity are major challenges in data preparation for such analysis (Jensen, et al., 2006; Lu, et al., 2007).

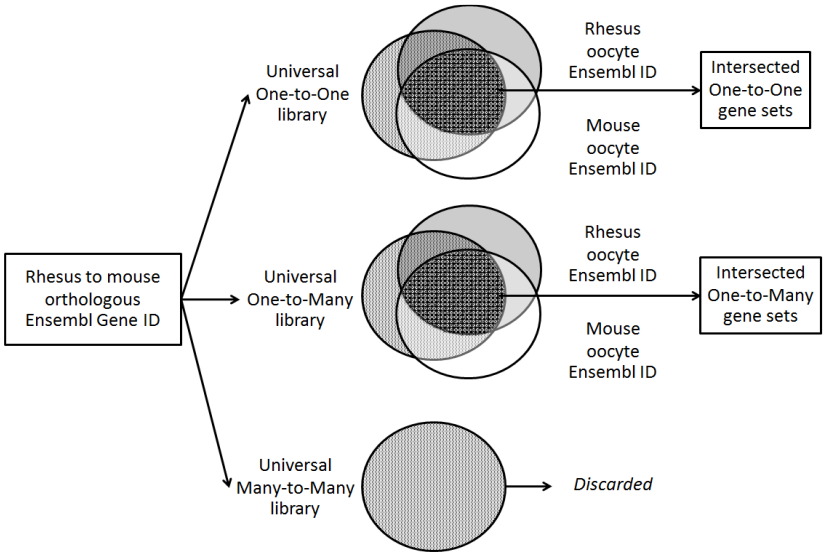


Figure 1 Orthologous gene extraction for pooling oocyte expression profiling datasets of rhesus monkey and mouse. Universal Ensembl Gene ID of rhesus monkey to mouse orthologous genes were retrieved from BioMart database and classified into one-to-one, one-to-many and many-to-many libraries. Orthologous gene lists in one-to-one and one-to-many libraries were intersected with rhesus and mouse oocyte genes of the pre-processed data corresponding to the animal species. Both of intersected gene sets were subsequently applied for further constructions of pooled datasets between the species.

Implement of concurrent microarray expression is usually performed between mouse and human (Ellwood-Yen, et al., 2003; Zheng-Bradley, et al., 2010). In fact, application of such analysis is rare in oocyte maturation study with extremely limited methodology evidenced. Seeing the great beneficial knowledge provided by the analysis of animal oocytes, we aimed to establish an approach to compare expression profiling of *in vitro* and *in vivo* matured oocytes among mammals. As previously mentioned, the reliability of the concurrent analysis greatly depends on data preparation. By means of the reason, our primary task was to forge reliable procedures to merge oocyte expression datasets acquired from different organisms together prior to further analyses.

In this study, we demonstrated methods to construct the pooled *in vivo* and *in vitro* matured oocyte expression datasets originated from rhesus monkey (*Macaca mulatta*) and mouse (*Mus musculus*). We focused on describing data preparation's procedures, and then applied some specific cross-species concurrent analyses with the constructed pooled datasets in order to determine the feasibility of the introduced preparation procedures.

Materials and Methods

Computer system: Bio-Linux-7 operating system (OS) was used in the current study (Field, et al., 2006). The computer had 10 GB hard-drive space, 16GB RAM, and Intel Pentium V processor.

R environment and software packages: R environment for statistical computing and graphics is freely available (http://www.r-project.org/). All software packages used in this study were available in Bioconductor (http:// www. bioconductor.org/), which was the open source software for bioinformatics.

Oocyte gene expression profiling datasets: Expression profiling datasets of *in vitro* and *in vivo* matured oocytes acquired from rhesus monkey (*Macaca mulatta*) and mouse (*Mus musculus*) were retrieved from Gene Expression Onimbus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) to put together total 18 samples (arrays). The details of all arrays used in this study were provided in Table 1 (further information about the oocyte samples could be retrieved directly from GEO using the sample ID). Briefly, *in vitro* matured mouse oocytes were acquired from germinal vesicle (GV) oocytes isolated from the follicles 48 hr post-PMSG injection and cultured for 16 hr. *In vivo* matured mouse oocytes were acquired from mice additionally stimulated with hCG 48 hr post-PMSG injection. In female monkeys, the first day of menses was noted as day 1 of the cycle. The monkeys were injected intramuscular with 37.5 IU recombinant macaque FSH (r-mFSH) twice daily for 7 days from day 1 of the cycle. *In vitro* matured monkey oocytes were acquired from GV oocytes isolated from the follicles on day 8 and cultured for 24 hr. *In vivo* matured monkey oocytes were acquired from female monkey additionally stimulated recombinant human chorionic gonadotropin (r-hCG) on day 8 for induction of oocyte maturation. Oocytes were collected 27-32 hr after r-hCG injection and were cultured until they reached metaphase II stage of nuclear maturation.

Table 1 Details of *in vitro* and *in vivo* matured oocyte samples (arrays) of rhesus monkey (*Macaca mulatta*) and mouse (*Mus musculus*) acquired from Gene Expression Onimbus (GEO) (http://www.ncbi.nlm.nih.gov/geo/).

Array number	Species	Sample ID	Series ID	Cell type	Platform
1	Monkey	GSM300525	GSE11895	*In vivo matured oocyte*	Affymetrix Rhesus Macaque Genome Array
2	Monkey	GSM300526	GSE11895	*In vivo matured oocyte*	Affymetrix Rhesus Macaque Genome Array
3	Monkey	GSM300527	GSE11895	*In vivo matured oocyte*	Affymetrix Rhesus Macaque Genome Array
4	Monkey	GSM300528	GSE11895	*In vivo matured oocyte*	Affymetrix Rhesus Macaque Genome Array
5	Monkey	GSM300529	GSE11895	*In vitro matured oocyte*	Affymetrix Rhesus Macaque Genome Array
6	Monkey	GSM300530	GSE11895	*In vitro matured oocyte*	Affymetrix Rhesus Macaque Genome Array
7	Monkey	GSM300531	GSE11895	*In vitro matured oocyte*	Affymetrix Rhesus Macaque Genome Array
8	Monkey	GSM300532	GSE11895	*In vitro matured oocyte*	Affymetrix Rhesus Macaque Genome Array
9	Mouse	GSM312251	GSE12432	*In vivo matured oocyte*	Affymetrix Mouse Genome 430 2.0 Array
10	Mouse	GSM312252	GSE12432	*In vitro matured oocyte*	Affymetrix Mouse Genome 430 2.0 Array
11	Mouse	GSM312253	GSE12432	*In vitro matured oocyte*	Affymetrix Mouse Genome 430 2.0 Array
12	Mouse	GSM312254	GSE12432	*In vivo matured oocyte*	Affymetrix Mouse Genome 430 2.0 Array
13	Mouse	GSM312255	GSE12432	*In vivo matured oocyte*	Affymetrix Mouse Genome 430 2.0 Array
14	Mouse	GSM312257	GSE12432	*In vivo matured oocyte*	Affymetrix Mouse Genome 430 2.0 Array
15	Mouse	GSM312258	GSE12432	*In vitro matured oocyte*	Affymetrix Mouse Genome 430 2.0 Array
16	Mouse	GSM312260	GSE12432	*In vitro matured oocyte*	Affymetrix Mouse Genome 430 2.0 Array
17	Mouse	GSM312262	GSE12432	*In vitro matured oocyte*	Affymetrix Mouse Genome 430 2.0 Array
18	Mouse	GSM312264	GSE12432	*In vivo matured oocyte*	Affymetrix Mouse Genome 430 2.0 Array

Table 2 Selected 100 genes differentially expressed between *in vitro* and *in vivo* matured oocytes in Ensembl gene ID

Rank	Rhesus	Mouse	Rank	Rhesus	Mouse	Rank	Rhesus	Mouse	Rank	Rhesus	Mouse	Rank	Rhesus	Mouse
1	ENSMMUG00000019876	ENSMUSG00000019942	21	ENSMMUG00000012578	ENSMUSG00000047261	41	ENSMMUG00000019599	ENSMUSG00000070291	61	ENSMMUG00000005502	ENSMUSG00000015291	81	ENSMMUG00000013534	ENSMUSG00000049323
2	ENSMMUG00000020337	ENSMUSG00000050312	22	ENSMMUG00000002144	ENSMUSG00000047228	42	ENSMMUG00000020960	ENSMUSG00000074643	62	ENSMMUG00000002555	ENSMUSG00000037600	82	ENSMMUG00000023126	ENSMUSG00000034192
3	ENSMMUG00000005528	ENSMUSG00000027615	23	ENSMMUG00000008585	ENSMUSG00000016626	43	ENSMMUG00000002912	ENSMUSG00000034758	63	ENSMMUG00000007483	ENSMUSG00000032968	83	ENSMMUG00000005283	ENSMUSG00000026572
4	ENSMMUG00000002810	ENSMUSG00000074607	24	ENSMMUG00000019535	ENSMUSG00000030717	44	ENSMMUG00000001799	ENSMUSG00000000378	64	ENSMMUG00000029171	ENSMUSG00000074336	84	ENSMMUG00000022962	ENSMUSG00000042262
5	ENSMMUG00000010652	ENSMUSG00000069378	25	ENSMMUG00000011305	ENSMUSG00000036019	45	ENSMMUG00000003043	ENSMUSG00000020156	65	ENSMMUG00000002981	ENSMUSG00000028738	85	ENSMMUG00000016251	ENSMUSG00000071470
6	ENSMMUG00000003673	ENSMUSG00000042251	26	ENSMMUG00000000713	ENSMUSG00000043372	46	ENSMMUG00000018785	ENSMUSG00000024271	66	ENSMMUG00000014433	ENSMUSG00000021957	86	ENSMMUG00000012911	ENSMUSG00000033126
7	ENSMMUG00000008830	ENSMUSG00000039670	27	ENSMMUG00000011541	ENSMUSG00000029685	47	ENSMMUG00000008783	ENSMUSG00000025938	67	ENSMMUG00000002623	ENSMUSG00000028070	87	ENSMMUG00000021256	ENSMUSG00000051502
8	ENSMMUG00000009546	ENSMUSG00000023353	28	ENSMMUG00000019339	ENSMUSG00000030519	48	ENSMMUG00000013601	ENSMUSG00000031935	68	ENSMMUG00000011095	ENSMUSG00000025612	88	ENSMMUG00000002404	ENSMUSG00000020827
9	ENSMMUG00000006076	ENSMUSG00000028576	29	ENSMMUG00000016212	ENSMUSG00000042608	49	ENSMMUG00000021011	ENSMUSG00000021007	69	ENSMMUG00000010292	ENSMUSG00000039206	89	ENSMMUG00000018791	ENSMUSG00000039616
10	ENSMMUG00000009858	ENSMUSG00000034765	30	ENSMMUG00000017370	ENSMUSG00000021917	50	ENSMMUG00000003134	ENSMUSG00000063760	70	ENSMMUG00000014341	ENSMUSG00000071573	90	ENSMMUG00000004379	ENSMUSG00000039384
11	ENSMMUG00000015230	ENSMUSG00000033739	31	ENSMMUG00000005939	ENSMUSG00000027834	51	ENSMMUG00000011548	ENSMUSG00000054967	71	ENSMMUG00000018922	ENSMUSG00000027330	91	ENSMMUG00000009307	ENSMUSG00000038085
12	ENSMMUG00000019820	ENSMUSG00000045629	32	ENSMMUG00000029445	ENSMUSG00000039910	52	ENSMMUG00000015041	ENSMUSG00000021715	72	ENSMMUG00000021229	ENSMUSG00000020514	92	ENSMMUG00000001572	ENSMUSG00000072582
13	ENSMMUG00000006295	ENSMUSG00000025355	33	ENSMMUG00000014844	ENSMUSG00000048138	53	ENSMMUG00000016004	ENSMUSG00000032561	73	ENSMMUG00000015093	ENSMUSG00000024924	93	ENSMMUG00000001500	ENSMUSG00000019761
14	ENSMMUG00000008281	ENSMUSG00000038025	34	ENSMMUG00000001848	ENSMUSG00000037211	54	ENSMMUG00000012415	ENSMUSG00000043969	74	ENSMMUG00000023565	ENSMUSG00000060012	94	ENSMMUG00000012002	ENSMUSG00000040661
15	ENSMMUG00000002220	ENSMUSG00000057777	35	ENSMMUG00000006205	ENSMUSG00000040731	55	ENSMMUG00000002363	ENSMUSG00000023170	75	ENSMMUG00000016292	ENSMUSG00000027525	95	ENSMMUG00000006860	ENSMUSG00000048399
16	ENSMMUG00000008018	ENSMUSG00000039158	36	ENSMMUG00000013209	ENSMUSG00000054582	56	ENSMMUG00000012353	ENSMUSG00000024083	76	ENSMMUG00000013834	ENSMUSG00000020910	96	ENSMMUG00000005639	ENSMUSG00000047554
17	ENSMMUG00000003239	ENSMUSG00000032323	37	ENSMMUG00000017119	ENSMUSG00000021147	57	ENSMMUG00000002820	ENSMUSG00000040651	77	ENSMMUG00000018279	ENSMUSG00000036291	97	ENSMMUG00000000784	ENSMUSG00000022070
18	ENSMMUG00000013607	ENSMUSG00000024646	38	ENSMMUG00000017390	ENSMUSG00000049907	58	ENSMMUG00000018144	ENSMUSG00000072295	78	ENSMMUG00000003468	ENSMUSG00000015312	98	ENSMMUG00000016406	ENSMUSG00000026775
19	ENSMMUG00000000491	ENSMUSG00000035439	39	ENSMMUG00000022693	ENSMUSG00000021906	59	ENSMMUG00000018067	ENSMUSG00000015879	79	ENSMMUG00000023143	ENSMUSG00000041237	99	ENSMMUG00000005998	ENSMUSG00000020099
20	ENSMMUG00000000188	ENSMUSG00000010097	40	ENSMMUG00000014529	ENSMUSG00000060044	60	ENSMMUG00000017567	ENSMUSG00000022868	80	ENSMMUG00000012423	ENSMUSG00000070866	100	ENSMMUG00000011678	ENSMUSG00000062309

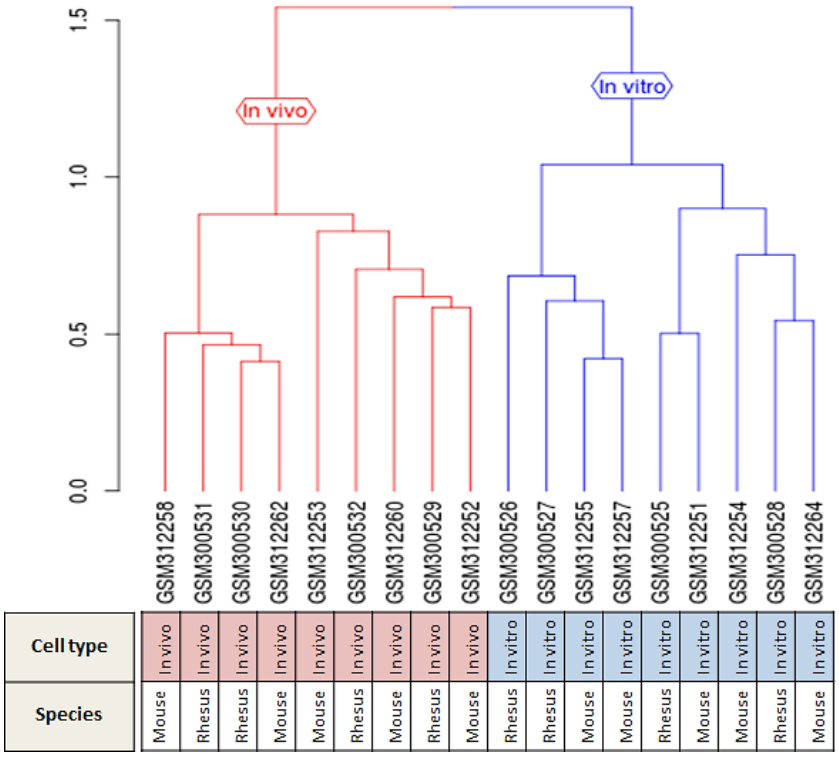
Table 3 Connectivity, Dunn and Silhouette scores calculated from each tested oocyte cluster number (from 2 to 8 clusters) acquired from each algorithm

Algorithms	Validation	Scores of each tested cluster number (2 to 8)						
		2	3	4	5	6	7	8
hierarchical[a]	Connectivity	3.8	13.2476	17.2774	26.194	28.9714	29.7881	31.8881
	Dunn	1.0539	0.6047	0.6178	0.5247	0.5766	0.6141	0.6328
	Silhouette	0.4872	0.3186	0.3063	0.1745	0.1553	0.1563	0.1492
kmeans[b]	Connectivity	3.8	8.8845	14.9071	23.8238	27.2738	32.4774	36.0702
	Dunn	1.0539	0.5622	0.5306	0.4172	0.4172	0.5247	0.5247
	Silhouette	0.4872	0.2339	0.2584	0.1249	0.0771	0.0963	0.0394
diana[c]	Connectivity	3.8	13.0167	15.0524	22.2595	28.0452	28.4952	31.6202
	Dunn	1.0539	0.6042	0.6173	0.626	0.6991	0.6991	0.7204
	Silhouette	0.4872	0.3386	0.3201	0.169	0.1679	0.1572	0.1339
fanny[d]	Connectivity	3.8	7.7798	21.9333	NA	NA	29.4393	28.4143
	Dunn	1.0539	0.6003	0.4825	NA	NA	0.5734	0.6178
	Silhouette	0.4872	0.2919	0.1994	NA	NA	0.0643	0.0867
som[e]	Connectivity	3.8	8.4679	26.2821	32.7988	31.4321	36.2964	35.2202
	Dunn	1.0539	0.4234	0.4902	0.4902	0.4702	0.5089	0.558
	Silhouette	0.4872	0.2065	0.1379	0.0736	0.0717	0.0832	0.1065
model[f]	Connectivity	3.8	8.8845	14.9071	18.175	19.9417	28.8583	31.1774
	Dunn	1.0539	0.5622	0.5306	0.5306	0.627	0.5247	0.5247
	Silhouette	0.4872	0.2339	0.2584	0.2462	0.2622	0.1346	0.1739
sota[g]	Connectivity	3.8	13.2476	16.081	20.1107	20.9274	22.2607	31.1774
	Dunn	1.0539	0.6047	0.6047	0.6178	0.6178	0.6178	0.5247
	Silhouette	0.4872	0.3186	0.2656	0.2748	0.2861	0.2984	0.1739
pam[h]	Connectivity	3.8	13.0167	22.0845	25.775	26.9333	28.5524	31.6024
	Dunn	1.0539	0.6042	0.6042	0.6047	0.6047	0.6265	0.6404
	Silhouette	0.4872	0.3386	0.1775	0.125	0.1129	0.1233	0.1246

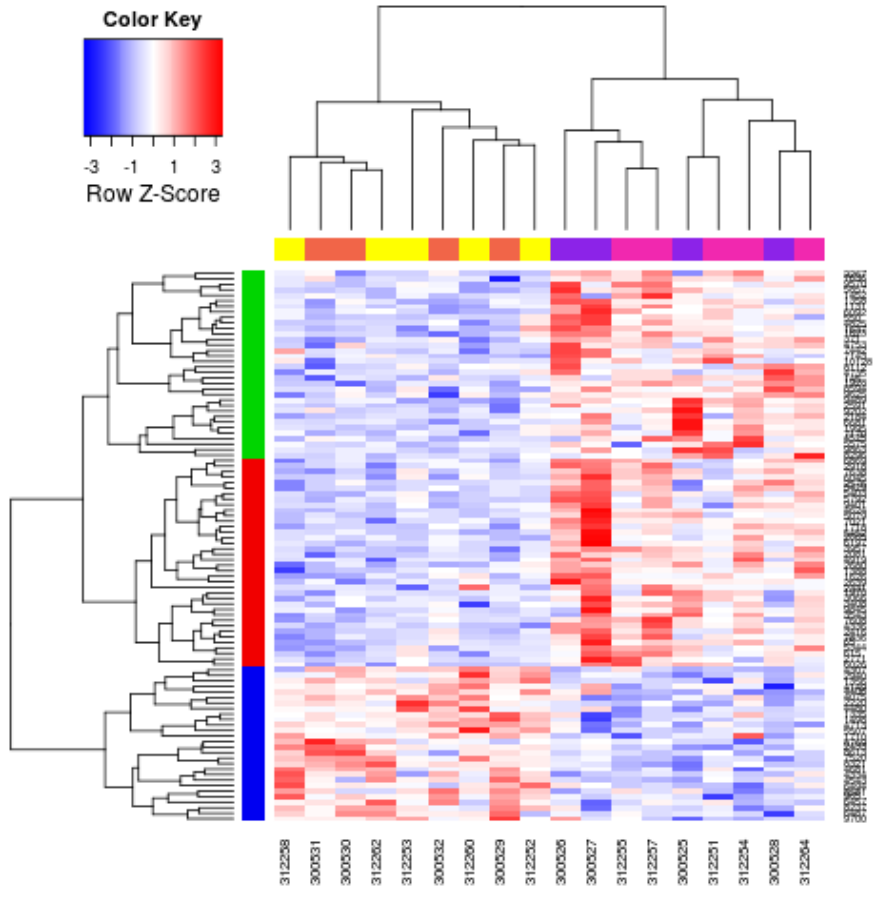
Note - The best cluster number should have highest connectivity score, but lowest Silhouette and Dunn scores.
[a]Agglomerative hierarchical clustering, [b]K-means, [c]Divisive hierarchical algorithm, [d]Fuzzy clustering, [e]Self-organizing maps, [f]Model based clustering, [g]Self-organizing tree, and [h]Partitioning around medoids

Table 4 Selected biological processes of cellular functions and pathways associated with the genes differentially expressed between *in vitro* and *in vivo* matured oocytes identified by DAVID tool

Organisms	Category	Term	% Count	P-Value	Genes
Mouse	INTERPRO	IPR001763:Rhodanese-like	3	0.00458	ENSMUSG00000039384, ENSMUSG00000027330, ENSMUSG00000034765
Mouse	SMART	SM00450:RHOD	3	0.004714	ENSMUSG00000039384, ENSMUSG00000027330, ENSMUSG00000034765
Mouse	SP_PIR_KEYWORDS	hydrolase	15	0.013345	ENSMUSG00000020910, ENSMUSG00000033126, ENSMUSG00000039384, ENSMUSG00000034765, ENSMUSG00000021917, ENSMUSG00000026775, ENSMUSG00000040661, ENSMUSG00000032561, ENSMUSG00000042251, ENSMUSG00000072582, ENSMUSG00000027330, ENSMUSG00000051502, ENSMUSG00000062309, ENSMUSG00000039206, ENSMUSG00000025355
Mouse	KEGG_PATHWAY	mmu04010:MAPK signaling pathway	4	0.024279	ENSMUSG00000015312, ENSMUSG00000039384, ENSMUSG00000027330, ENSMUSG00000034765
Mouse	SP_PIR_KEYWORDS	phosphoprotein	41	0.02607	ENSMUSG00000023353, ENSMUSG00000038085, ENSMUSG00000049323, ENSMUSG00000037211, ENSMUSG00000039384, ENSMUSG00000020099, ENSMUSG00000019942, ENSMUSG00000039616, ENSMUSG00000024083, ENSMUSG00000022868, ENSMUSG00000015312, ENSMUSG00000025612, ENSMUSG00000040651, ENSMUSG00000027330, ENSMUSG00000015291, ENSMUSG00000038025, ENSMUSG00000033739, ENSMUSG00000039158, ENSMUSG00000019761, ENSMUSG00000027525, ENSMUSG00000040731, ENSMUSG00000040661, ENSMUSG00000021715, ENSMUSG00000074643, ENSMUSG00000030519, ENSMUSG00000047554, ENSMUSG00000045629, ENSMUSG00000043372, ENSMUSG00000020827, ENSMUSG00000000378, ENSMUSG00000041237, ENSMUSG00000072582, ENSMUSG00000028576, ENSMUSG00000034758, ENSMUSG00000028070, ENSMUSG00000047261, ENSMUSG00000062309, ENSMUSG00000021957, ENSMUSG00000031935, ENSMUSG00000022070, ENSMUSG00000039206
Monkey	KEGG_PATHWAY	mcc04010:MAPK signaling pathway	3	0.030087	ENSMMUG00000009858, ENSMMUG00000003468, ENSMMUG00000004379
Mouse	GOTERM_BP_FAT	GO:0003006~reproductive developmental process	5	0.033962	ENSMUSG00000022868, ENSMUSG00000071470, ENSMUSG00000048138, ENSMUSG00000027330, ENSMUSG00000039910
Mouse	GOTERM_CC_FAT	GO:0034361~very-low-density lipoprotein particle	2	0.036198	ENSMUSG00000024924, ENSMUSG00000074336
Mouse	GOTERM_CC_FAT	GO:0034385~triglyceride-rich lipoprotein particle	2	0.036198	ENSMUSG00000024924, ENSMUSG00000074336
Mouse	KEGG_PATHWAY	mmu04110:Cell cycle	3	0.037711	ENSMUSG00000015312, ENSMUSG00000027330, ENSMUSG00000019942
Mouse	GOTERM_CC_FAT	GO:0005576~extracellular region	12	0.037929	ENSMUSG00000022868, ENSMUSG00000071573, ENSMUSG00000032968, ENSMUSG00000028070, ENSMUSG00000047228, ENSMUSG00000032561, ENSMUSG00000024924, ENSMUSG00000027834, ENSMUSG00000033739, ENSMUSG00000074336, ENSMUSG00000042251, ENSMUSG00000025355
Mouse	GOTERM_BP_FAT	GO:0000279~M phase	5	0.04212	ENSMUSG00000071470, ENSMUSG00000027330, ENSMUSG00000019942, ENSMUSG00000022070, ENSMUSG00000035439
Mouse	INTERPRO	IPR008343:MAP kinase phosphatase	2	0.045227	ENSMUSG00000039384, ENSMUSG00000034765
Mouse	UP_SEQ_FEATURE	region of interest:Small GTPase-like	2	0.048375	ENSMUSG00000023353, ENSMUSG00000049907
Monkey	KEGG_PATHWAY	mcc04115:p53 signaling pathway	2	0.081696	ENSMMUG00000019876, ENSMMUG00000003468



2A)



2B)

Figure 2 Hierarchically clustered *in vitro* and *in vivo* matured oocytes. All *in vitro* and *in vivo* matured oocytes acquired from both rhesus monkey and mouse were separately clustered by use of 100 selected differentially expressed genes (2A). To virtualize differences in gene regulations between these clusters, a heatmap of hierarchical clustering of the differentially expressed genes in all oocyte samples was shown with displayed fold-value color key. The rows represented genes, and the columns represented oocyte samples. Red in the heatmap denoted gene upregulation while blue denoted gene downregulation. The order of oocyte samples (arrays) in the heatmap was the same as presented in Fig 2A, but with omission of "GSM" in the names (2B).

Raw dataset pre-procession: To prepare mouse and monkey data for subsequent pooling procedures, the raw datasets of monkey and mouse were separately pre-processed using probe-wise background correction and between-array Variance Stabilization Normalization (VSN) on the perfect match (PM) values only. Probe set summaries were calculated with the medianpolish method of Robust multi-array average (RMA) algorithm. The process was simply achieved using 'justvsn' function provided by package 'vsn' (Huber, et al., 2002). Quality of the pre-processed data was determined using package 'arrayQualityMetrics' (Kauffmann, et al., 2009) as described in our previous report (Chokeshai-u-saha and Sananmuang, 2014). Probe sets specific to multiple genes were removed from analysis, and subsequently consolidated to corresponding Ensembl Gene ID (Birney, et al., 2004; Stabenau, et al., 2004) based on their maximal interquartile ranges. The acquired expression sets of each species were scaled to median value prior to pooling step.

Pooling of rhesus monkey and mouse expression profiling datasets: After pre-procession, pooling of scaled pre-processed expression profiling datasets of rhesus monkey and mouse was performed based on their orthologous genes (orthologs) retrieved from BioMart database using package 'biomaRt' (Durinck, et al., 2009). To merge the datasets of the two species, Ensembl Gene ID among orthologs of rhesus monkey and mouse were subtracted to create 3 gene libraries as follows: one-to-one, one-to-many, and many-to-many libraries using mouse as reference species (Fig. 1). Each library contained rhesus monkey entry gene list and their corresponding mouse orthologs. For one-to-one library, each of rhesus genes was paired with a corresponding mouse ortholog. For one-to-many library, each of rhesus genes was paired with more than one corresponding mouse orthologs. We paid no attention to many-to-many library due to its redundancy. Orthologs acquired from intersection between each library's gene list and pre-processed data's gene list of each species were then classified, and used for constructing pooled datasets (Fig. 1). Only the expression values of intersected orthologs acquired from datasets of each species were retrieved, then pooled together based on the ortholog type. Of note, only one-to-one ortholog pooled datasets were applied for further oocyte sample clustering analysis.

Differentially expressed gene selection: The candidate genes differentially expressed between *in vitro* and *in vivo* matured oocytes were determined from the one-to-one ortholog pooled datasets for subsequent clustering and enrichment analyses. Five hundred learning datasets were generated from the one-to-one ortholog pooled datasets using five-fold cross validation method. Linear models for differential gene expression analysis with Benjamini & Hochberg correction (*False discover rate* < 0.01) was applied with each of the learning datasets (Smyth, 2004) using 'limma' method in 'GeneSelection' function of package 'CMA' (Slawski, et al., 2009). Considering all differentially expressed genes derived from every learning dataset, top 100 indexed genes from ranking were selected for subsequent oocyte sample clustering.

Oocyte sample clustering: Efficiency of the selected differentially expressed genes to discriminate oocyte cell types, several sample clustering algorithms were applied with the one-to-one ortholog pooled datasets using package 'clValid' (Brock, et al., 2008). There algorithms were as follows: Agglomerative hierarchical clustering, K-means, Divisive hierarchical algorithm, Fuzzy clustering, Self-organizing maps, Model based clustering, Self-organizing tree, and Partitioning around medoids. Compactness and connectedness of the clustering result were determined by connectivity, Silhouette and Dunn scores, accordingly. The best clustering result should have highest connectivity score, but lowest Silhouette and Dunn scores.

Gene enrichment analysis: To identify biological processes of cellular functions and pathways associated with the genes differentially expressed between *in vitro* and *in vivo* matured oocytes, DAVID bioinformatics resources were used for enrichment analysis (Huang da, et al., 2009). The acquired differentially expression gene lists of each organism was uploaded to DAVD tool (http:// david. abcc. ncifcrf.gov/home.jsp). ENSEMBL_GENE_ID was selected as the identifier. The important biological themes were selected from functional annotation charts of both organisms.

Results

The pooled gene expression profiling data between rhesus monkey and mouse oocytes were successfully constructed, and applicable for differential gene expression analysis: Oocyte expression profiling datasets of both rhesus monkey and mouse were pre-processed and pooled together as previously described in material and methods. We constructed two pooled datasets in this study. The first one contained with only one-to-one orthologous genes (10,214 genes), and the other contained with both one-to-one and one-to-many orthologous genes (11,356 genes). The complete pooled datasets were not provided in the manuscript. Differentially expressed genes between *in vitro* and *in vivo* oocytes were determined in one-to-one ortholog pooled datasets as described in material and methods. The acquired differentially expressed genes (in Ensembl Gene ID) of both species used for subsequent oocyte clustering analysis and gene enrichment analysis were provided in Table 2.

Correct clusters of in vitro and in vivo oocytes were revealed by all clustering algorithms: Several clustering algorithms were applied with one-to-one pooled oocyte expression datasets using the selected 100 differentially expressed genes (Table 2). Connectedness and compactness scores were acquired among clustering algorithms (Table 3). However, the clustering results indicated 2 clusters always performed the best by all algorithms (connectivity score = 3.8, Silhouette score = 0.49 and Dunn score = 1.05). One sample cluster contained all *in vivo* matured

oocytes, while the other contained all *in vitro* matured oocytes. To demonstrate an example of the acquired results, dendrogram (Fig 2A) and heatmap of agglomerative hierarchical clustering algorithm (Fig 2B) with complete linkage of oocyte samples were illustrated. The heatmap result clearly indicated two sets of genes responsible for separating *in vitro* from *in vivo* matured oocytes (Fig 2B).

Differences in developmental processes between in vitro and in vivo matured oocytes were implied by enrichment analysis: Enrichment analysis was performed as described in materials and methods with default setting. The differentially expressed genes were found associated with reproductive developmental process such as cell-cycle, Mitosis phase, MAPK signaling pathway, and some specific enzymes ($p < 0.05$) (Table 4).

Discussion

Most researchers imply knowledge of oocyte maturation among experimental organisms based on evolutionary relationship regardless of biological similarity determination in term of statistics. Such disregard can bring about great amount of unexpected outcome, especially when the optimized oocyte culture systems were applied across different species. While some intensive cross-species gene expression profiling meta-analyses of animal oocytes were conducted to deal with the issue (O'Shea, et al., 2012), the studies usually applied individual species analysis strategy by which produced certain technical biases. Since concurrent microarray expression analysis could reduce the biases and to allow other outstanding signatures, we aimed to introduce it as an alternative strategy for the future cross-species oocyte maturation studies. However, the lack of methodology to merge the oocyte expression datasets across species was a major obstacle to complete the task. We thus demonstrated such procedures using rhesus monkey and mouse as organism models in this study.

To pool the monkey and mouse datasets together, several combined pre-processions were actually tested with the raw data (Chokeshai-u-saha and Sananmuang, 2014). According to the quality of the pre-processed data (results not shown), probe-wise background correction and between-array Variance Stabilization Normalization (VSN) on the perfect match (PM) values with median polish summarization were selected. After pre-procession, we pooled and categorized the data by their orthologous gene types (one-to-one, one-to-many and many-to-many). We found that pooling data by ortholog categories allowed easy manipulation of the pooled datasets both for dataframe construction and extraction of orthologous gene names.

In the current study, we selected only the pooled one-to-one ortholog datasets for further analyses in order to avoid requirement to duplicate gene expression values for analyses. As previously described in the introduction, the use of pooled expression profiles comes with several advantages (Lu, et al., 2009). Generation of learning datasets for subsequent differential gene expression analysis (Table 2) and oocyte sample clustering (Fig 2) were possible in this study. According to the clustering results, correct clustering of both *in vitro* and *in vivo* oocyte samples verified the feasibility of our prepared data (Table 3 and Fig 2A). Concordance with the clustering results, several differences in developmental processes between *in vitro* and *in vivo* matured oocytes were also implied by subsequent enrichment analysis ($p < 0.05$) (Table 4). For in-depth understanding of these differences, intensive analyses and narrations of both acquired differentially expressed genes and enriched terms should be performed in a separate study.

In conclusion, the current study successfully introduced a methodology to merge microarray expression profiling of oocytes acquired from different organisms. In addition, we successfully demonstrated the use of some concurrent analyses to validate the prepared data. The proposed data preparation procedures were simple, yet flexible to apply with various organisms of interest. It should be noted that number of informative orthologs among the analyzed organisms was a major limitation of such approach. Since it determined the left-over gene numbers after pooling data, it could greatly affect the sensibility of differential gene expression analysis subsequently performed. Finally, we hoped that our provided methodology would be useful for future cross-species gene expression profiling researches of oocytes.

Acknowledgements

The study was supported by ASEAN Community Project of Rajamangala University of Technology Tawan-ok Bangpra, Chonburi, Thailand. We also would like to thank Benoît Ballester, the researcher of TAGC for sharing his experience in analyzing large-scale expression data, discussion and technical help.

References

Bar-Or C, Czosnek H, and Koltai H 2007. Cross-species microarray hybridizations: a developing tool for studying species diversity. Trends Genet. 23(4): 200-207.

Bergmann S, Ihmels J, and Barkai N 2004. Similarities and differences in genome-wide expression data of six organisms. PLoS Biol. 2(1): E9.

Birney E, Andrews TD, Bevan P, Caccamo M, Chen Y, Clarke L, Coates G, Cuff J, Curwen V, Cutts T, *et al.* 2004. An overview of Ensembl. Genome Res. 14(5): 925-928.

Brock G, Pihur V, Datta S, and Datta S 2008. clValid: An R Package for Cluster Validation. JStatSoftware. 25(4): 1-22.

Chokeshai-u-saha K, and Sananmuang T 2014. Optimization of Pre-processing Procedures for Canine Mammary Tumor Oligonucleotide Microarray Meta-analysis. The Thai Journal of Veterinary Medicine. 44(Sup 1): 125-128.

Durinck S, Spellman PT, Birney E, and Huber W 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 4(8): 1184-1191.

Eichenlaub-Ritter U, and Peschke M 2002. Expression in *in-vivo* and *in-vitro* growing and maturing oocytes: focus on regulation of expression at the translational level. Hum Reprod Update. 8(1): 21-41.

Ellwood-Yen K, Graeber TG, Wongvipat J, Iruela-Arispe ML, Zhang J, Matusik R, Thomas GV, and Sawyers CL 2003. Myc-driven murine prostate cancer shares molecular features with human prostate tumors. Cancer Cell. 4(3): 223-238.

Field D, Tiwari B, Booth T, Houten S, Swan D, Bertrand N, and Thurston M 2006. Open software for biologists: from famine to feast. Nat Biotechnol. 24(7): 801-803.

Huang da W, Sherman BT, and Lempicki RA 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 4(1): 44-57.

Huber W, von Heydebreck A, Sultmann H, Poustka A, and Vingron M 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics. 18(Sup 1): S96-104.

Jensen LJ, Jensen TS, de Lichtenberg U, Brunak S, and Bork P 2006. Co-evolution of transcriptional and post-translational cell-cycle regulation. Nature. 443(7111): 594-597.

Kauffmann A, Gentleman R, and Huber W 2009. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. Bioinformatics. 25(3): 415-416.

Lee YS, Latham KE, and Vandevoort CA 2008. Effects of *in vitro* maturation on gene expression in rhesus monkey oocytes. Physiol Genomics. 35(2): 145-158.

Lu Y, Huggins P, and Bar-Joseph Z 2009. Cross species analysis of microarray expression data. Bioinformatics. 25(12): 1476-1483.

Lu Y, Mahony S, Benos PV, Rosenfeld R, Simon I, Breeden LL, and Bar-Joseph Z 2007. Combined analysis reveals a core set of cycling genes. Genome Biol. 8(7): R146.

Mamo S, Carter F, Lonergan P, Leal CL, Al Naib A, McGettigan P, Mehta JP, Evans AC, and Fair T 2011. Sequential analysis of global gene expression profiles in immature and in vitro matured bovine oocytes: potential molecular markers of oocyte maturation. BMC Genomics. 12: 151.

O'Shea LC, Mehta J, Lonergan P, Hensey C, and Fair T 2012. Developmental competence in oocytes and cumulus cells: candidate genes and networks. Syst Biol Reprod Med. 58(2): 88-101.

Sananmuang T, Tharasanit T, Nguyen C, Phutikanit N, and Techakumphu M 2011. Culture medium and embryo density influence on developmental competence and gene expression of cat embryos. Theriogenology. 75(9): 1708-1719.

Sananmuang T, Phutikanit N, Nguyen C, Manee-In S, Techakumphu M, and Tharasanit T 2013. *In vitro* culture of feline embryos increases stress-induced heat shock protein 70 and apoptotic related genes. J Reprod Dev. 59(2): 180-188.

Slawski M, Boulesteix A, and Bernau C 2009. "CMA: Synthesis of microarray-based classification." [Online.] Available: http:// www. bioconductor.org/packages/release/bioc/html/CMA.html. Accessed October 15, 2014.

Smyth GK 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 3(1): 1544-6115.

Stabenau A, McVicker G, Melsopp C, Proctor G, Clamp M, and Birney E 2004. The Ensembl core software libraries. Genome Res. 14(5): 929-933.

Zheng-Bradley X, Rung J, Parkinson H, and Brazma A 2010. Large scale comparison of global gene expression patterns in human and mouse. Genome Biol. 11(12): R124.

บทคัดย่อ

การสังเคราะห์โปรไฟล์การแสดงออกของยีนร่วมกันระหว่างเซลล์ไข่ของลิงวอกและเซลล์ไข่ของหนูเพื่อการวิเคราะห์ข้อมูลเชิงอภิมาณ

กาจ โชคชัยอุสาหะ[1] เดนิส พูทิเยร์[2] แคเธอรีน เหงียน[2] ธนิดา สนั่นเมือง[1*]

ในปัจจุบันองค์ความรู้ที่ได้รับจากการเปรียบเทียบโปรไฟล์การแสดงออกของยีนระหว่างเซลล์ไข่ที่เจริญพัฒนาภายนอกกับเซลล์ไข่ที่เจริญพัฒนาภายในร่างกายระหว่างชนิดของสัตว์เลี้ยงลูกด้วยนม ด้วยวิธีวิเคราะห์อภิมาณมีบทบาทอย่างยิ่งต่อการพัฒนาเทคโนโลยีทางการสืบพันธุ์ในสัตว์ อย่างไรก็ตามการแสดงกระบวนการในการสังเคราะห์โปรไฟล์เพื่อการวิเคราะห์ดังกล่าวยังคงมีอยู่อย่างจำกัด การศึกษาครั้งนี้จึงมีจุดมุ่งหมายเพื่อนำเสนอและทดสอบกระบวนการสังเคราะห์โปรไฟล์ดังกล่าว โดยใช้ข้อมูลโปรไฟล์การแสดงออกยีนของเซลล์ไข่ที่เจริญพัฒนาภายนอกและภายในร่างกายที่ได้จากหนูและลิงวอกเป็นต้นแบบการศึกษา โดยอาศัยการจับคู่ยีนที่มีคุณลักษณะเป็นออโธโลกัสยีน จากนั้นจึงสังเคราะห์โปรไฟล์จำลองขึ้นอีก 500 โปรไฟล์ เพื่อวิเคราะห์หายีนที่มีระดับการแสดงออกที่แตกต่างกันระหว่างเซลล์ไข่ที่เจริญพัฒนาขึ้นภายนอกเทียบกับภายในร่างกายด้วยไลเนียร์โมเดลจำนวน 1,000 ยีน (FDR<0.01) และใช้รายชื่อยีนดังกล่าววิเคราะห์จัดกลุ่มตัวอย่างเซลล์ไข่ทั้งหมดด้วยอัลกอลิธึมประเภทต่างๆ พร้อมทั้งวิเคราะห์หาความสัมพันธ์ระหว่างยีนดังกล่าวกับกระบวนการเมตาบอลิซึมต่าง ๆ ของเซลล์ไข่ด้วย DAVID Bioinformatics Resources 6.7 ผลจากการศึกษาแสดงให้เห็นว่ากระบวนการสังเคราะห์ยีนโปรไฟล์ระหว่างหนูและลิงวอกในการศึกษาครั้งนี้สามารถสังเคราะห์ยีนโปรไฟล์รวมได้สำเร็จ โดยยีนโปรไฟล์ดังกล่าวประกอบด้วยออโธโลกัสยีนแบบหนึ่งต่อหนึ่งจำนวน 10,214 ยีน เมื่อวิเคราะห์ยีนที่มีระดับการแสดงออกแตกต่างกันระหว่างเซลล์ไข่ที่เจริญพัฒนาภายนอกและภายในร่างกายพบว่ายีนดังกล่าวสามารถใช้ในการจัดกลุ่มแยกเซลล์ไข่ที่เจริญพัฒนาภายนอกและภายในร่างกายในหนูและลิงวอกได้อย่างถูกต้อง โดยผลการวิเคราะห์ยังบ่งชี้ความสัมพันธ์ระหว่างกลุ่มยีนดังกล่าวกับกระบวนการเจริญและพัฒนาของเซลล์ไข่ที่แตกต่างกันอีกด้วย กล่าวโดยสรุปกระบวนการสังเคราะห์โปรไฟล์การแสดงออกของยีนร่วมกันระหว่างหนูและลิงวอกในการศึกษาครั้งนี้ประสบความสำเร็จ โดยโปรไฟล์รวมที่สร้างขึ้นจากกระบวนการที่นำเสนอสามารถนำไปวิเคราะห์ต่อเนื่องและให้ผลการวิเคราะห์ที่สอดคล้องกับคุณสมบัติทางชีววิทยาที่แตกต่างกันระหว่างเซลล์ไข่ที่เจริญพัฒนาภายนอกและภายในร่างกาย กระบวนการดังกล่าวจึงสามารถนำไปประยุกต์ใช้กับเซลล์ไข่ของสัตว์เลี้ยงลูกด้วยนมชนิดอื่น ๆ ได้ต่อไปในอนาคต

คำสำคัญ: การวิเคราะห์อภิมาณ หนู เซลล์ไข่ โปรไฟล์การแสดงออกของยีนร่วมกัน ลิงวอก

[1]คณะสัตวแพทยศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก วิทยาเขตบางพระ 43 หมู่ 6 ตำบลบางพระ อำเภอศรีราชา จังหวัดชลบุรี 20110 ประเทศไทย

[2]ศูนย์ความเป็นเลิศด้านจีโนมิกส์และคลินิก สถาบันด้านสาธารสุขและวิจัยทางการแพทย์แห่งชาติฝรั่งเศส ยูนิต 1090 อุทธยานทางวิทยาศาสตร์ลูมินี 928 163 ถนนลูมินี เมืองมาร์เซย์ 13288 ประเทศฝรั่งเศส

****ผู้รับผิดชอบบทความ** E-mail: t.sananmuang@gmail.com*