International Journal of Current Research in Life Sciences
Vol. 4, No. 08, pp. 321-327, August, 2015

www.ijcrls.com

Full Length Research Article

Gene Expression Profile of Ovarian Cancer Dataset that helps to Predict Potential Biomarkers

Divya, V., Anitha P. Muttagi, Seema J Patel, Gurumurthy, H. and Prashantha Nagaraja

GM Institute of Technology, Department of Biotechnology, Davanagere, Bangalore-560092

Accepted **22ed** ***July*** **2015;** ***Published Online*** **31st** ***August*** **2015**

ABSTRACT

The viral pathogens such as human papillomavirus (HPV), cytomegalovirus (CMV) and Chlamydia trachomatis are a significant risk factor for developing women mucinous epithelial ovarian cancer. The clinical and morphological distinct of ovarian cancer subtypes of frequent concurrence of endometriosis results in poor prognosis. The lack of significant markers associated with diagnosis in early stage of infection. The aim of the study is conducted oligonucleotide microarray to compare image analysis and normalization algorithms to analyze host pathogen interacting genes and proteins that reside on network of disease susceptibility. The expression is calculated on the ratio of expression levels between virus-infected tissues and normal tissues are brought great expectations for finding biomarkers that would improve patient's treatment in the early stage of infection. The intensity of gene chip is processed and normalization is carried out using R statistical software. We have identified significant clusters of highly enriched gene markers that extent in epithelial malignancies and predicted the large expression data of candidate molecular biomarkers.

Key words: ovarian cancer, viral cancers, human papilloma virus, cytomegalovirus, Microarray, image analysis, statistical analysis, hierarchical clustering, k-means clustering

INTRODUCTION

Ovarian cancer is a most leading cause of death from gynecological malignancy in all over the world (Greenlee *et al*., 2001). Approximately, 70% of women are diagnosed with ovarian adenocarcinoma are derived from ovarian surface epithelia (OSE) (Scully *et al*., 2004; Auersperg *et al*., 2001). Based on histological subtypes of epithelial ovarian carcinomas shows 40% of population is affected with viral infection such as human papillomavirus (HPV), Chlamydia trachomatis and cytomegalovirus (CMV) (Atalay *et al*., 2007). The closer look at HPV virus infected with cervical cancer is link with ovarian cancer is still unclear (HPV, 2011). The women with sexually active has infected with HPV family members HPV-16 and HPV-18. The high risk of HPV infected subtypes produce two types of oncogens designed E6 and E7 proteins induce transformation by interference with endogenous cell cycle regulatory proteins, including P53, retinoblastoma (Rb) and breast cancer type 1 susceptibility protein (BRAC1) (Fishman *et al*., 2002). The HPV virus is clearly link with different cancers such as head and neck cancers (Giordano *et al*., 2008), still there is lot of investigations is going on to identify the genes, functions and pathways that helps to identify potential drug targets. The virus such as Chlamydia trachomatis is affected with women having sexually transmitted disease (Claman *et al*., 1997).

Corresponding author: Prashantha Nagaraja,
GM Institute of Technology, Department of Biotechnology, Davanagere, Bangalore-560092.

There are several serotypes that may cause urogenital infection which evolve in chronic infection results in spreading to female pelvic blockage to cause infertility (Den Hartog, 2005; Dieterle and Wollenhaupt, 1996). The pathogenesis of the c.trachomatis is unknown, but the bacterium may affect non immune cells that produce proinflammatory response against host cell (Idahl *et al*., 2007; Tiitinen *et al*., 2006). The genomic function and structure activity remains unknown, the bacterium express high levels of Chlamydial heat shock protein 60 (cHSP60) affects immune system and cause tubal factor infertility (Madeleine *et al*., 2007). There are several other pathogens such as mycobacterium tuberculosis is also associated with squamous cell carcinoma of the cervix (Koskela *et al*., 2000).

The role of persistent infection, leading to chronic inflammation, in the pathogenesis of ovarian cancer has received very little consideration, although a history of pelvic inflammatory disease (PID) is in a case-control study correlated to higher risk for ovarian cancer (Risch *et al*., 1995). Recent histopathological studies of different genomes involved in disease causing are a major task in disease identification. There are different histotypes helps to identify the differentially expressed genes that significantly associated with pathogenesis in both cancerous and non-cancerous tissues that is suitable prediction of drug identification or vaccines preparation to cure the disease.

MATERIALS AND METHODS

Raw Data selection and pre-processing of microarray data

In order to determine the pathogenic genes involved in ovarian cancer and the role of gene expression in disease progression can be analyzed using microarray data. The Comparative gene expression profiling analyses were carried out to determine disease mechanism and the role of signaling pathways, (i) Gene expression measurements (ii) definitions of signaling pathways and (iii) protein drug targets prediction. We have evaluated all published case control studies and diseased datasets were selected using various repositories such as GEO (Gene expression Omnibus), Array express (EBI database), and PUMAdb (Princeton University Microarray database). In order to determine the Chlamydia trachomatis in ovarian cancer dataset such as GSE41075 (Vicetti Miguel *et al*., 2013) and Human Papilloma virus infected dataset GSE49288 (Kim and Shin, 2013) is used for studies.

The GSE41075 dataset contains 22 samples of which 10 controls trans-cervical, endometrial biopsy specimens of upper and lower genital tract infection and 12 women of C. Trachomatis endometrial infection, cells were processed for microarray analysis using Affymetrix Human genome U133A Gene Chip. The GSE49288 dataset has 39 cervical cancer samples were grouped into 4 sets based on physical examination and differential expression of gene profiles using Agilent two-color experiment. The Affymetrix and Agilent datasets is pre-processed using RMA and MAS5 algorithms. However, the signal intensity of MM probe can often be larger than the PM probe implying that MM probe is detecting a true signal as well as background signal. Probe set results were further evaluated using R and BioConductor software Probes were considered differentially expressed if they had a fold change value of ≥ 3 and a p-value $< .005$ (Student's t-test). The sequence clusters were created from the UniGene database and then refined by analysis and comparison with a number of other publicly available databases.

Identification of Differential gene expression data analysis

The preprocessed dataset is used for differential gene expression studies using limma package. The RMA function assigned the factorial design expression that transforming log2 values. To assign column names of eset creates contrast matrix to perform all pairwise comparisons to compute estimated coefficients and standard errors of a given datasets. Computes moderated t-statistics and log-odds of differential expression by empirical Bayes shrinkage of the standard errors towards a common value. Generates list of top 10 ('number=10') differentially expressed genes sorted by B-values ('sort.by=B') for each of the three comparison groups ('coef=1') in this sample set. The summary table has logFC is the log2-fold change, the AveExpr is the average of expression value across all arrays and channels, the moderated t-statistic (t) is the logFC to its standard error, the P.Value is the associated p-value, the adj. P. Value is the p-value adjusted for multiple testing and the B-value (B) is the log-odds that a gene is differentially expressed (the-higher-the-better). Usually one wants to base gene selection on the adjusted P-value rather than the t- or B-values. Filters out candidates that have P-values < 0.05 in each group ('coef=1') and provides the number of candidates for each list. Same as above, but with complex filter: P-value < 0.01 AND at least 2-fold change AND expression value A > 10. The genes are sorted based on p-value of differential expressions under the primary condition. The argument 'primary=1' selects the first contrast column in the 'results' matrix as a primary condition.

Functional enrichment analyses

In order to obtain the functional enrichment of the differentially expressed genes on the cell level, we used the GO (Gene Ontology) database to classify the gene function and location information. We performed GO cluster analysis by using the cluster Profiler package, then deduced the affection of these differentially expressed genes to the cells by cluster the cells within the molecular functions and biological processes. The Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov) were used to identify over-represented biological functions and pathways among the differentially expressed genes.

Gene Network prediction and Drug target prediction

A gene co-citation network of differentially expressed genes that allows gene network prediction and visualization helps to identify potential drug targets. The topology study of gene network is further analyzed using Fast Greedy community structure prediction algorithm is implemented in the Cytoscape plug-in Glay. The Glay helps to identify coherent subnetworks of functional annotated and enriched data by DAVID to reveal over-represented biological functions. The gene expression signatures is calculated the distance matrix for the disease pairs based on the overlap between sets of differentially expressed genes used for potential drug targets.

RESULTS

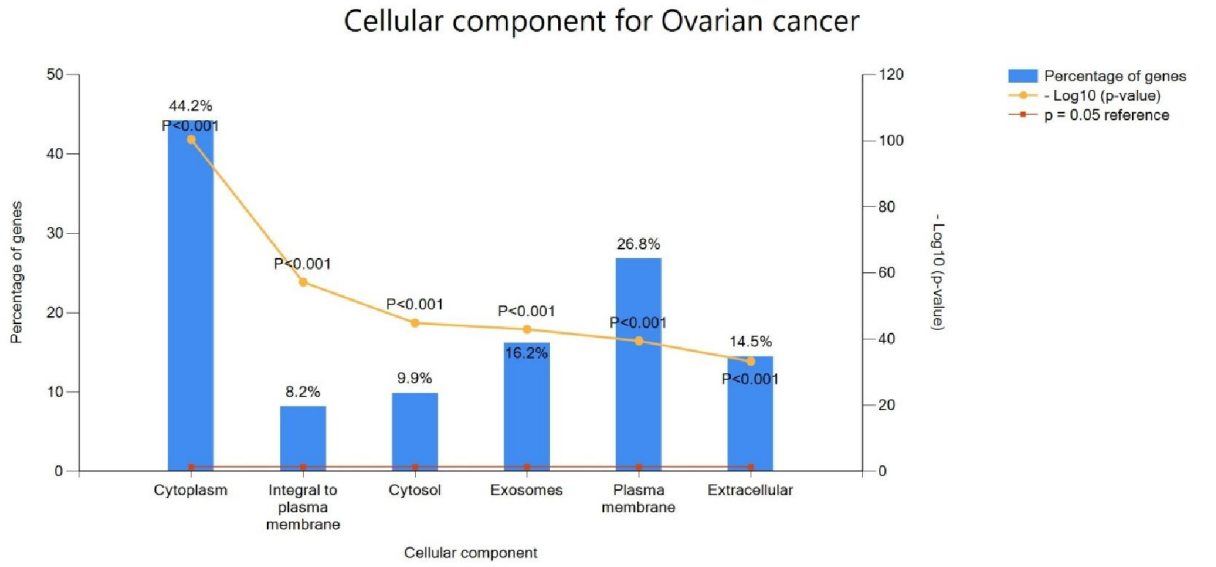
The pre-processed data of Chlamydia Trachomatis is used for differential gene expression analysis. There are 24789 genes in a dataset is and used for pair wise comparisons with different series of samples. The differential gene expression is predicted based on log2 fold changes, standard errors, t-statistics and p-values. The analysis of significance is based on probe intensity that is interpreted by ordinary t-statistic except that the standard errors have been moderated across genes, i.e., shrunk towards a common value, using a simple Bayesian model. Using functional annotation, we have predicted 1643 differentially expressed genes, out of these 276 up regulated and 1367 down regulated genes expressed in Chlamydia Trachomatis infection in ovarian cancer. There are 158 genes is significantly associated with protein expression of which 26 protein functional genes is used for potential drug targets. The network prediction of differently expressed genes such as RPL24, CYP2A6, CYP2E` of Cytochrome P45 subfamily, CST3, DDX5, RPS27A, NONO, JUN, TP53, EGR1, NCOR1, HNRNPU and EGFR genes is predicted in subnetworks of close association and is used for potential drug targets, the overall results is predicted in Table: 1, Figure: 1a-h.

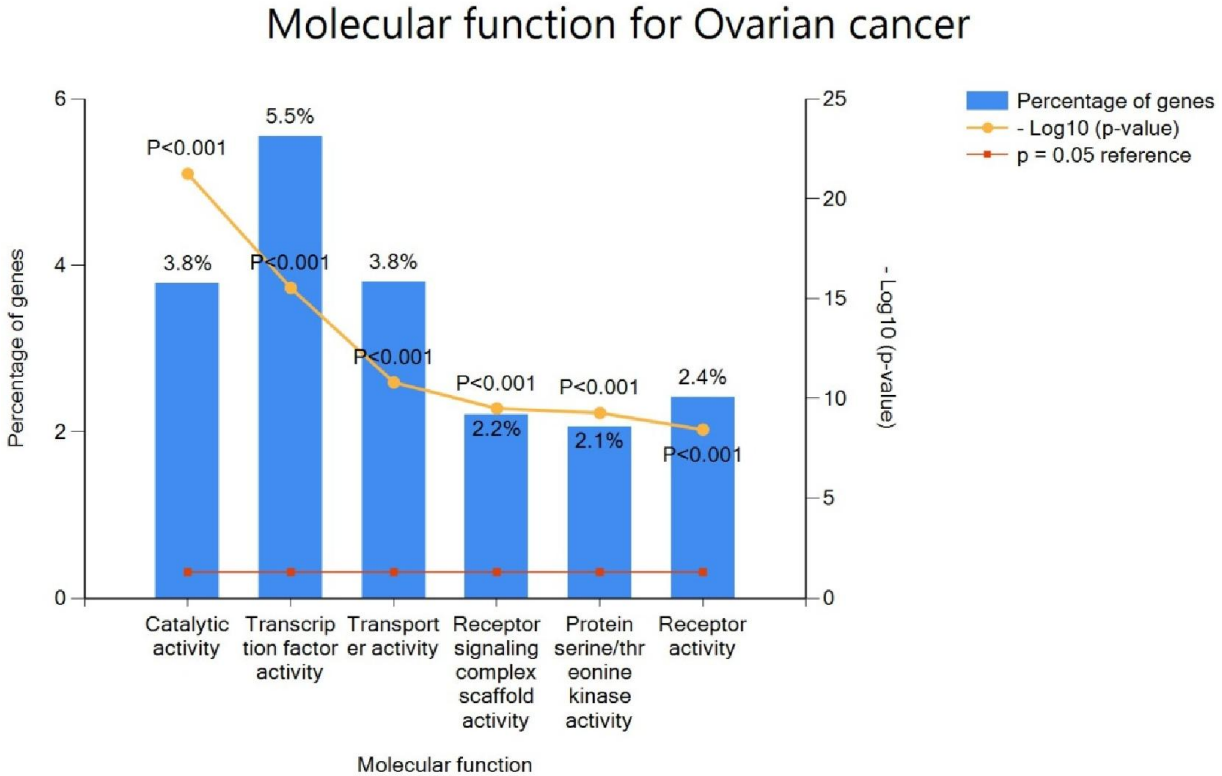
The HPV infected dataset is preprocessed using Agilent platform of RMA and MAS5 algorithms that used for differential gene expression analysis. The overall significant differences of gene signatures were listed in table 2a-d. We

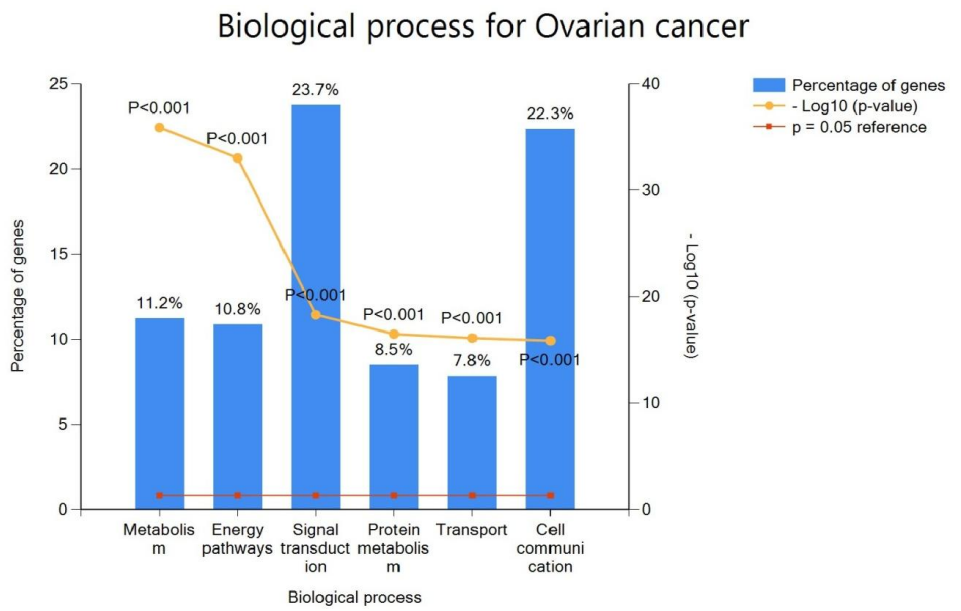
Table: 1. significant gene signatures is predicted from Chlamydia Trachomatis is differentially expressed in ovarian and cervical cancer

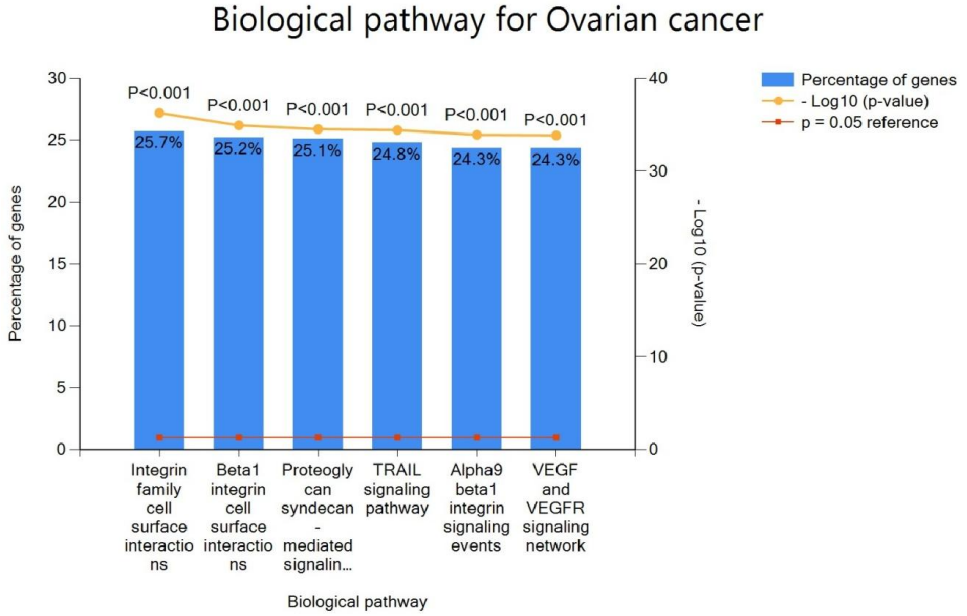
Description	P-value	FDR q-value	Enrichment
Nuclear gene expression	1.68E-6	2.18E-60	3.67
SRP-dependent cotranslational protein targeting to membrane	3.26E-5	2.11E-50	11.84
cotranslational protein targeting to membrane	2.01E-5	8.69E-50	11.62
protein targeting to ER	4.46E-5	1.45E-49	15.9
protein localization to endoplasmic reticulum	4.51E-5	1.17E-49	10.65
establishment of protein localization to endoplasmic reticulum	3.77E-5	8.16E-49	15.46
protein targeting to membrane	1.15E-4	2.13E-46	12.49
viral process	3.89E-4	6.30E-46	3.24
symbiosis, encompassing mutualism through parasitism	3.89E-4	5.60E-46	3.24
multi-organism cellular process	1.63E-4	2.12E-45	3.21
interspecies interaction between organisms	2.34E-4	2.77E-44	3.01
establishment of protein localization to membrane	2.79E-4	3.02E-42	6.44
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	6.17E-4	6.16E-42	15.11
Translation	6.22E-4	5.76E-42	7.26
viral transcription	1.97E-4	1.71E-40	15.31
mRNA metabolic process	4.26E-4	3.45E-40	4.32
protein targeting	4.14E-4	3.16E-38	7.79

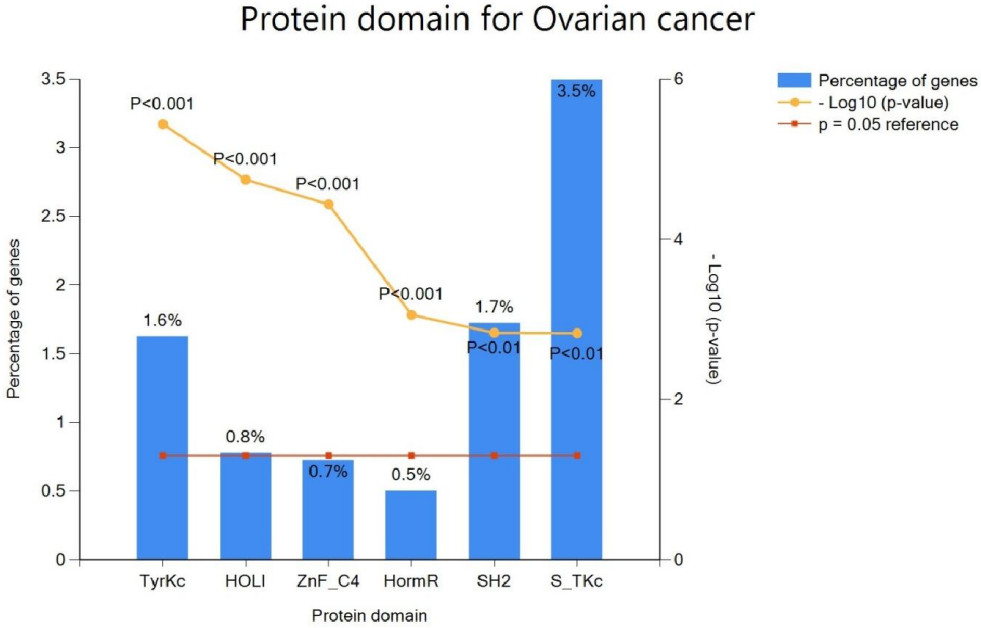
Fig.1a-h. Differentially expressed gene signatures in protein functions of ovarian cancer

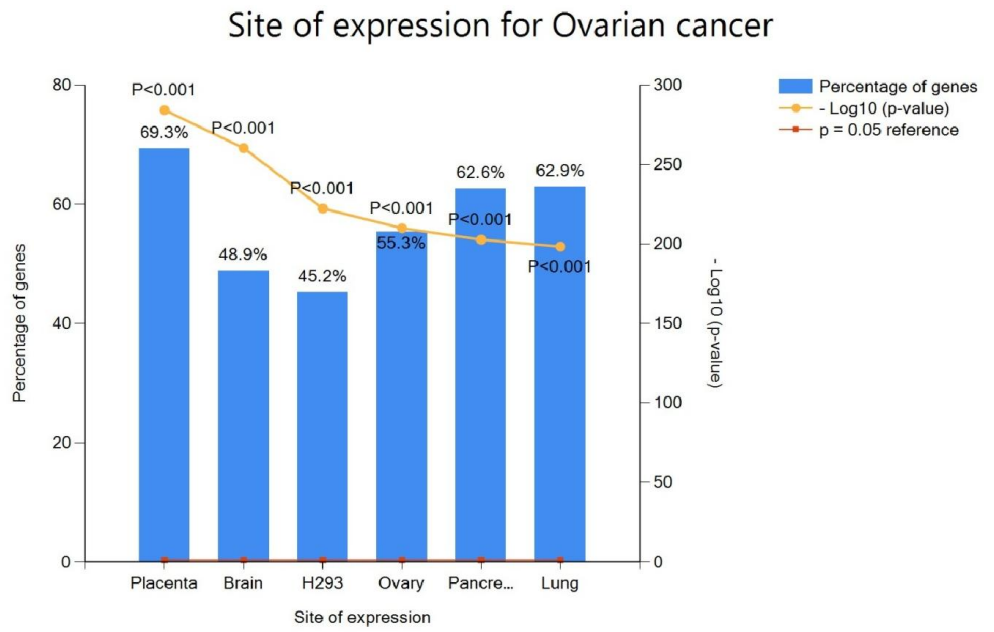


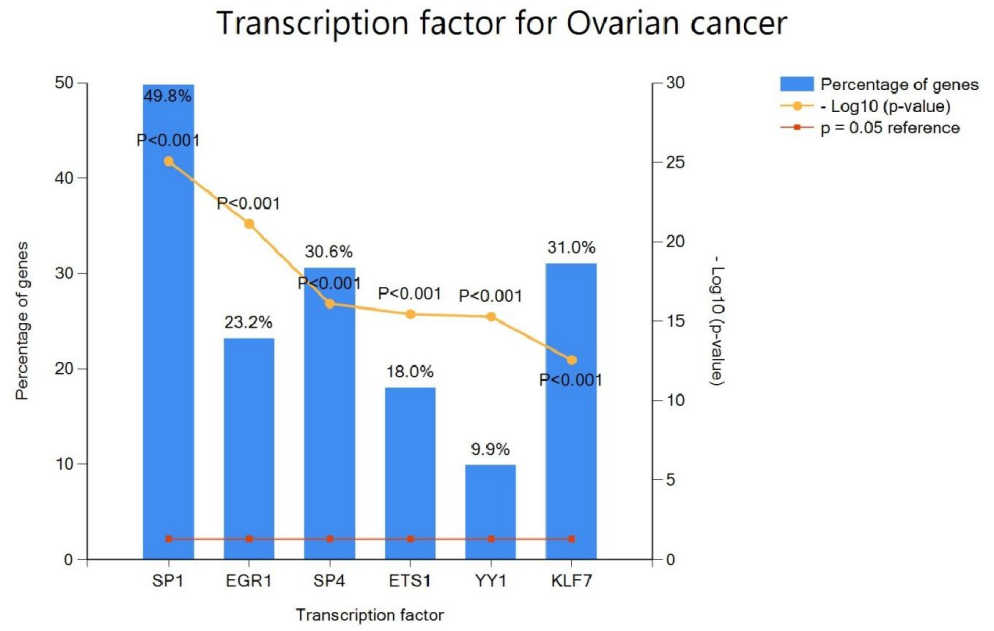



Biological process for Ovarian cancer
Percentage of genes
- Log10 (p-value)
p = 0.05 reference
P<0.001
P<0.001
P<0.001
P<0.001
P<0.001
P<0.001
11.2%
10.8%
23.7%
8.5%
7.8%
22.3%
Metabolism
Energy pathways
Signal transduction
Protein metabolism
Transport
Cell communication
Biological process


Biological pathway for Ovarian cancer
Percentage of genes
- Log10 (p-value)
p = 0.05 reference
P<0.001
P<0.001
P<0.001
P<0.001
P<0.001
P<0.001
25.7%
25.2%
25.1%
24.8%
24.3%
24.3%
Integrin family cell surface interactions
Beta1 integrin cell surface interactions
Proteoglycan syndecan-mediated signalin...
TRAIL signaling pathway
Alpha9 beta1 integrin signaling events
VEGF and VEGFR signaling network
Biological pathway


Protein domain for Ovarian cancer
Percentage of genes
- Log10 (p-value)
p = 0.05 reference
P<0.001
P<0.001
P<0.001
P<0.001
P<0.01
P<0.01
1.6%
0.8%
0.7%
0.5%
1.7%
3.5%
TyrKc
HOLI
ZnF_C4
HormR
SH2
S_TKc
Protein domain


Site of expression for Ovarian cancer
Percentage of genes
- Log10 (p-value)
p = 0.05 reference
Percentage of genes
- Log10 (p-value)
P<0.001
69.3%
48.9%
45.2%
55.3%
62.6%
62.9%
Placenta
Brain
H293
Ovary
Pancre...
Lung
Site of expression


Transcription factor for Ovarian cancer
Percentage of genes
- Log10 (p-value)
p = 0.05 reference
Percentage of genes
- Log10 (p-value)
P<0.001
49.8%
23.2%
30.6%
18.0%
9.9%
31.0%
SP1
EGR1
SP4
ETS1
YY1
KLF7
Transcription factor

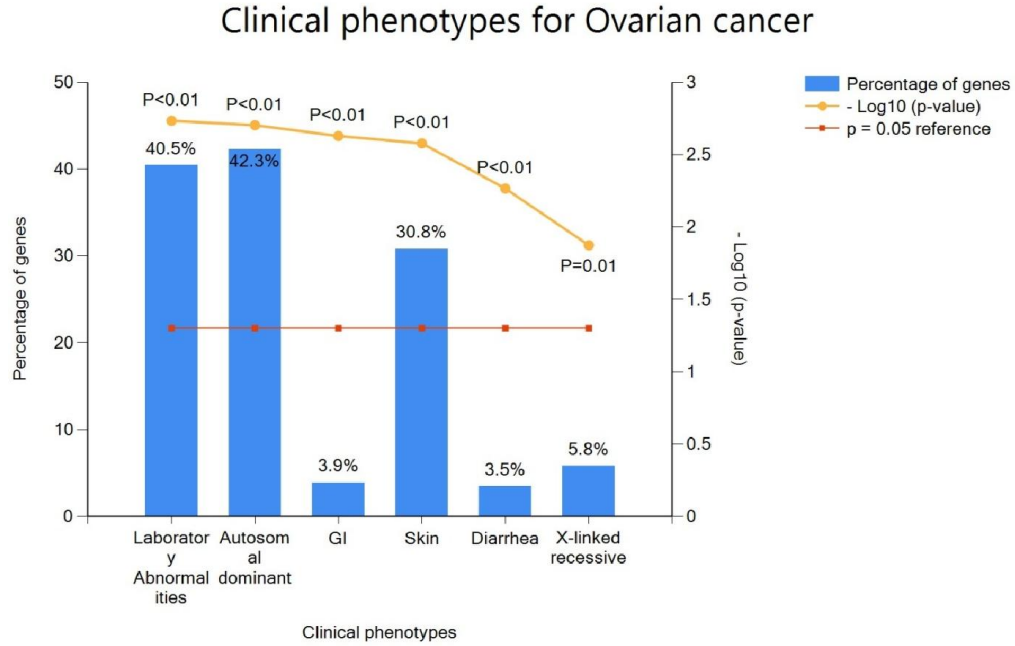

Clinical phenotypes for Ovarian cancer
Percentage of genes
- Log10 (p-value)
p = 0.05 reference
Percentage of genes
- Log10 (p-value)
P<0.01
40.5%
42.3%
3.9%
30.8%
3.5%
5.8%
P=0.01
Laboratory Abnormalities
Autosomal dominant
GI
Skin
Diarrhea
X-linked recessive
Clinical phenotypes

Table 2a-d. Differentially expressed gene signatures are predicted from HPV virus dataset in ovarian and cervical cancer

Group_1

Systematic Name	logFC	AveExpr	t	P.Value	adj.P.Val	B
AU146383	-4.190478	13.07684	-73.10817	1.84E-42	7.55E-38	84.57788
AF130077	-3.937033	13.64195	-64.96314	1.57E-40	3.23E-36	80.81721
NM_000384	-3.602157	12.05400	-59.49027	4.31E-39	5.90E-35	77.92578
NM_001922	-4.217703	12.61427	-59.02669	5.78E-39	5.94E-35	77.66547
NM_001633	-3.477487	13.49116	-57.77211	1.30E-38	1.06E-34	76.94805
THC2606976	-4.193509	13.14557	-57.14937	1.95E-38	1.33E-34	76.58473
NM_000042	-3.546585	13.55756	-55.42191	6.16E-38	3.61E-34	75.55090
NM_001643	-3.319628	13.13852	-55.01945	8.09E-38	4.16E-34	75.30439
NM_000509	-3.613461	12.15498	-53.25116	2.75E-37	1.26E-33	74.19485
THC2724353	-3.578358	12.44191	-51.63890	8.71E-37	3.58E-33	73.14401

Group_2

Systematic Name	logFC	AveExpr	t	P.Value	adj.P.Val	B
AU146383	-4.069740	13.07684	-59.07701	5.60E-39	2.30E-34	76.377542
AF130077	-3.806848	13.64195	-52.26526	5.55E-37	1.14E-32	72.532514
THC2606976	-4.123639	13.14557	-46.75888	3.57E-35	4.83E-31	68.911794
NM_001922	-3.981425	12.61427	-46.36182	4.91E-35	4.83E-31	68.630142
NM_000384	-3.357627	12.05400	-46.13870	5.87E-35	4.83E-31	68.470556
NM_001633	-3.300054	13.49116	-45.61666	8.98E-35	6.15E-31	68.093443
NM_001643	-3.185937	13.13852	-43.93530	3.64E-34	2.14E-30	66.842100
NM_000042	-3.341126	13.55756	-43.44237	5.54E-34	2.85E-30	66.464219
NM_000509	-3.431692	12.15498	-42.07884	1.82E-33	8.30E-30	65.391600
THC2724353	-3.420932	12.44191	-41.07591	4.45E-33	1.83E-29	64.575999

Group_3

Systematic Name	logFC	AveExpr	t	P.Value	adj.P.Val	B
AU146383	-4.179398	13.07684	-70.05434	9.18E-42	3.77E-37	82.77293
AF130077	-3.864762	13.64195	-61.26885	1.42E-39	2.92E-35	78.54435
NM_001922	-4.190444	12.61427	-56.34448	3.31E-38	4.54E-34	75.80832
NM_000384	-3.460444	12.05400	-54.90780	8.73E-38	8.97E-34	74.95241
NM_001633	-3.419427	13.49116	-54.57894	1.09E-37	8.99E-34	74.75256
THC2606976	-4.045935	13.14557	-52.97509	3.35E-37	2.29E-33	73.75608
NM_001643	-3.305887	13.13852	-52.64217	4.24E-37	2.49E-33	73.54457
NM_000042	-3.454238	13.55756	-51.86117	7.42E-37	3.81E-33	73.04193
NM_000509	-3.555429	12.15498	-50.34041	2.26E-36	1.03E-32	72.03632
NM_000596	-3.871317	13.25031	-48.71193	7.74E-36	3.18E-32	70.91819

Group_4

Systematic Name	logFC	AveExpr	t	P.Value	adj.P.Val	B
AU146383	-4.07096	13.076840	-39.3965	2.10E-32	8.62E-28	61.549129
AF130077	-3.81418	13.641952	-34.9106	1.83E-30	3.76E-26	57.764220
NM_001922	-4.17110	12.614272	-32.3803	2.91E-29	3.99E-25	55.348703
NM_000384	-3.46389	12.054003	-31.7326	6.11E-29	6.27E-25	54.693505
NM_001633	-3.40523	13.491166	-31.3803	9.19E-29	7.55E-25	54.330422
NM_001643	-3.30521	13.138527	-30.3868	2.98E-28	2.04E-24	53.279967
NM_000042	-3.43094	13.557564	-29.7401	6.53E-28	3.83E-24	52.574541
NM_000509	-3.54431	12.154984	-28.9731	1.69E-27	8.68E-24	51.714773
NM_000596	-3.84755	13.250311	-27.9512	6.22E-27	2.84E-23	50.528154
THC272435	-3.46241	12.441919	-27.7159	8.45E-27	3.17E-23	50.248103

have classified 4 different groups which are top ranked with functional annotation and enrichment studies. We have identified 19848 upregulated, 1860 down regulated genes that significantly expressed in cervical tissues.

Using significant association studies shows only 100 genes that eventually predicts functional characters of proteins. We have predicted AU146383 probe is highly expressed HMEC-1 endothelial cell line response of transcriptional factor-induced pluripotent stem cells of ovarian tissues. The other probe such as AF130077, IGFBP1, AFP, AGT, ABCC2, ITIH3, ApoB and ApoH genes is mainly involved in HPV transcriptional regulation and immune suppression on host cells. These selected proteins are mainly used for potential drug targets against cervical and ovarian cancer.

Conclusion

On ongoing challenges of drug targets identification of different cancer types based on type of pathogenesis, disease progression, risk factors and signs and symptoms is evidence to slow down the development of novel anticancer agents. We have used computational methods to identify drug targets based on gene expression of ovarian and cervical cancer associated pathogens such as Chlamydia trachomatis and HPV. Using different statistical analysis to identify differential expressions of genes involved in ovarian and cervical cancer furthermore used genome-wide scale of functional annotation and functional enrichment studies to identify potential drug targets based on protein expression. We have identified the top level expressed genes such as RPL24, CYP2A6, and CYP2E

`of Cytochrome P45 subfamily, CST3, DDX5, RPS27A, NONO, JUN, TP53, EGR1, NCOR1, HNRNPU and EGFR is mainly targeted to Chlamydia trachomatis infection in cervical and ovarian cancer infection. The HPV targeted genes such as AU146383, AF130077, IGFBP1, AFP, AGT, ABCC2, ITIH3, ApoB and ApoH is best drug targets against cervical and ovarian cancer tissues. We believe that the application of our integrated approach has the potential to provide a list of drug target candidates for other human diseases.

Acknowledgement

We gratefully acknowledge the assistance of Dr. C. N Prashantha for supporting project work, workflows, technical assistance and report preparation. We also grateful to Seema J Patel and Anitha P.Muttagi as internal guide to conduct seminars and guiding project. We also grateful to Department of Biotechnology, GMIT and Scientific Bio-Minds to providing continuous support to successfully complete the project.

Conflict of interest

There is no conflict of interest

REFERENCES

"Human Papillomavirus (HPV)." 2011. cdc.gov. Centers for Disease Control and Prevention. 11 March.

Anttila, T. Saikku, P. Koskela, P. *et al*. 2001. "Serotypes of Chlamydia trachomatis and risk for development of Ovarian squamous cell carcinoma," *Journal of the American Medical Association*, vol. 285, no. 1, pp. 47–51, 2001.

Atalay, F. *et al*. 2007. "Detection of human papillomavirus DNA and genotyping in patients with epithelial ovarian carcinoma." *Journal of Obstetrics and Gynaecological Research* 33.6:832-888.

Auersperg, N, Wong, A.S, Choi, K.C., Kang, S.K., Leung, P.C. 2001. Ovarian surface epithelium: biology, endocrinology, and pathology. Endocr Rev 2001, 22:255-288.

Claman, P. Honey, L. Peeling, R. W. Jessamine, P. and Toye, B. 1997. "The presence of serum antibody to the chlamydial heat shock protein (CHSP60) as a diagnostic test for tubal factor infertility," Fertility and Sterility, vol. 67, no. 3, pp. 501–504.

den Hartog, J. E. Land, J. A. Stassen, F. R. Kessels, A. G. and Bruggeman, C. A. 2005. "Serological markers of persistent C.trachomatis infections in women with tubal factor subfertility," Human Reproduction, vol. 20, no. 4, pp. 986–990.

Dieterle, S. and Wollenhaupt, J. 1996. "Humoral immune response to the chlamydial heat shock proteins hsp60 and hsp70 in Chlamydia-associated chronic salpingitis with tubal occlusion," Human Reproduction, vol. 11, no. 6, pp. 1352–1356.

Fishman, D.A., Bozorgi, K. 2002. The scientific basis of early detection of epithelial ovarian cancer: the national ovarian cancer early detection program (NOCEDP). Cancer Treat Re, 107:3-2

Giordano, G. *et al*. 2008. "Human papilloma virus (HPV) status, p16INK4a, and p53 overexpression in epithelial malignant and borderline ovarian neoplasm." Pathology Research and Practice 204.3:163-174.

Greenlee, R.T., Hill-Harmon, M.B., Murray, T and Thun, M. 2001. : Cancer statistics, 2001. *CA Cancer J Clin* 51: 15-36.

Idahl, A. Abramsson, L. Kumlin, U. Liljeqvist, J. A. and J. I. Olofsson, 2007. "Male serum Chlamydia trachomatis IgA and IgG, but not heat shock protein 60 IgG, correlates with negatively affected semen characteristics and lower pregnancy rates in the infertile couple," *International Journal of Andrology*, vol. 30, no. 2, pp. 99–107.

Kim, J., Shin, H., 2013, Differentially expressed gene profiles according to physical status of human papillomavirus integration in cervical cancer, GEO database, Jul 27, 2013.

Koskela, P. Anttila, T. Bjørge, T. *et al*. 2000. "Chlamydia trachomatis infection as a risk factor for invasive Ovarian cancer," *International Journal of Cancer,* vol. 85, no. 1, pp. 35–39.

Madeleine, M. M. Anttila, T. Schwartz, S. M. *et al*. 2007. "Risk of Ovarian cancer associated with Chlamydia trachomatis antibodies by histology, HPV type and HPV cofactors," *International Journal of Cancer*, vol. 120, no. 3, pp. 650–655,.

Risch, H. A. and Howe, G. R. 1995. "Pelvic inflammatory disease and the risk of epithelial ovarian cancer," Cancer Epidemiology Biomarkers and Prevention, vol. 4, no. 5, pp. 447–451.

Scully, R.E., Clement, P.B., Young, R.H. 2004. Ovarian Surface Epithelial-Stromal Tumors. In STERNBERG'S Diagnostic Surgical Pathology. Edited by Mills SE. Philadelphia, PA: Lippincott Williams & Wilkins;:2543-2578.

Tiitinen, A. Surcel, H. Halttunen, M. *et al*., 2006. "Chlamydia trachomatis and chlamydial heat shock protein 60-specific antibody and cell-mediated responses predict tubal factor infertility," 2006.

Vicetti Miguel RD1, Harvey SA, LaFramboise WA, Reighard SD, Matthews DB, Cherpes TL. 2013.. Human female genital tract infection by the obligate intracellular bacterium Chlamydia trachomatis elicits robust Type 2 immunity. PLoS One. 2013;8(3):e58565.

****