## SUBMITTED AS A WEB SERVER ISSUE IN NAR

Title:

# ArrayXPath: mapping and visualizing microarray geneexpression data with biomedical ontologies and integrated biological pathway resources using Scalable Vector Graphics

*Hee-Joon Chung<sup>1</sup>, Chan Hee Park<sup>1</sup>, Mi Ryung Han<sup>1</sup>, Seokho Lee<sup>1</sup>, Jung Hun Ohn<sup>1</sup>, Jihun Kim<sup>1,2\*</sup>* 

<sup>1</sup>Seoul National University Biomedical Informatics (SNUBI) and <sup>2</sup>Human Genome Research Institute, Seoul National University College of Medicine, Seoul 110-799, Republic of Korea.

\* To whom correspondence should be addressed at Seoul National University Biomedical Informatics (SNUBI), Seoul National University College of Medicine, Seoul 110-799, Korea. Tel: +82-2-740-8320; Fax: +82-2-747-4830; Email: juhan@snu.ac.kr

Running title: Integrating microarray data with biomedical ontology and pathway

#### ABSTRACT

**Summary**: ArrayXPath (http://www.snubi.org/software/ArrayXPath/) is a web-based service for mapping and visualizing microarray gene-expression data with integrated biological pathway resources using Scalable Vector Graphics (SVG). Deciphering the crosstalk among pathways and integrating biomedical ontologies and knowledge bases may help biological interpretation of microarray data. ArrayXPath is empowered by integrating gene-pathway, disease-pathway, drug-pathway, and pathway-pathway correlations with integrated Gene Ontology (GO), Medical Subject Headings (MeSH), and OMIM Morbid Map-based annotations. We applied Fisher's exact test and relative risk to evaluate the statistical significance of the correlations. ArrayXPath produces Javascript-enabled SVGs for web-enabled interactive visualization of gene expression profiles integrated with gene-pathway-disease interactions enriched by biomedical ontologies.

#### **INTRODUCTION**

Cluster analysis is one of the most powerful methods for the exploratory analysis of gene expression data. Genes expression clusters based on similarity measures between expression profiles have positional associations along the chromosomes (1, 2), exhibit common cis-regulatory elements in their upstream regions (3), and are coordinated by shared sets of regulators (4). Gene expression clusters can be assigned to the well-known functional categories of the MIPS classification (5), the GO terms (6) or pathway resources (7) using annotations from public databases (3, 8, 9).

ArrayXPath (7) is a web-based application that (i) receives a clustered gene-expression profile of any microarray platform in a tab-delimited text format; (ii) automatically resolves the microarray probe identifiers (i.e., GenBank accession number, UniGene ID, LocusLink ID, official gene symbol, SwissProt ID, or TrEMBL ID); (iii) searches major public pathway resources (i.e., GenMAPP, KEGG, BioCarta and PharmGKB Pathways); (iv) maps the different identifier sets between microarray probes and pathway nodes; (v) tests the statistical significance of the association between gene expression clusters and pathways (hence providing an automated annotation of clusters with the ranked pathways); (vi) visualizes expression levels onto pathways, and (vii) allows web-based user navigation through multiple clusters and pathways enriched with animation features, using Javascript-enabled SVG.

Although biological pathways can provide key information about the organization of biological systems, relatively small number (i.e.  $\sim$ 3,000) of genes compared to the estimated number (i.e. > 30,000) of genes for our species, as reported in our previous work (7), do appear in major pathway resources, resulting insufficient coverage for genome-wide expression data analysis. Although GO-based annotations give lesser information than pathway-based ones, increasingly more gene products are being annotated by GO terms, resulting much higher coverage. As of 2005 February, we found that 13,949 LocusLink IDs have at least one GO annotations. Therefore, integrating lesser-knowledge-higher-coverage GO-based annotations can complement more-knowledg-lower-coverage pathway-based annotations for microarray data analysis.

Gene-pathway correlation alone may not be sufficient for deciphering the genomic secret of normal and pathological physiology. Integrating not only biological (i.e. GO) but also clinical ontologies like the disease nomenclature system supported by MeSH can provide further information for genotype-to-phenotype association. OMIM Morbid Map provides valuable gene-disease correlations. Integrating drug-pathway correlation from PharmGKB Pathways (10) can also improve ArrayXPath, which is intrinsically an automatic annotation machine.

We found that pathways had significant and informative crosstalk. Many genes appear in multiple pathways. Systematic analysis and interactive visualization of the complex crosstalk structures among pathways, pathway nodes (i.e. gene products), and gene expression clusters may help understanding gene-pathway correlations. Fig. 1(a) shows the concept diagram of new ArrayXPath that integrates the quinta-partite graph structure of cluster, gene, disease, pathway and GO-term associations from multiple resources.

Here we present an improved version of ArrayXPath. In addition to the functionalities described above, this is a software that (viii) tests the statistical significance of the association between gene expression clusters and GO-based annotations to complement pathway-based microarray data analysis; (ix) allows users to search disease-related pathways; (x) visualizes the global crosstalk of biological pathways by measuring and mapping the similarity distances superimposed by the local crosstalk of the subset of pathways matched to input gene-expression clusters, and (xi) visualizes the detailed local crosstalk through gene-cluster, gene-pathway, gene-disease and gene-GO associations using interactive multi-partite graph representations in SVG. OMIM knowledge base and drug-pathway correlations from PharmGKB Pathways are also tightly integrated to AXP.

## **INPUT AND OUTPUT**

#### Input

Input to ArrayXPath is a common tab-delimited text file for a clustered gene expression profile:  $<Probe ID>-<Cluster ID>-[<Expression level at condition_i>]$ . The first column must contain either GenBank accession number, UniGene ID, LocusLink ID, SwissProt ID, TrEMBL ID or an official gene symbol. The second column contains the cluster ID. The third to *i*th columns are optional and contain expression levels. ArrayXPath does not perform cluster analysis *per se*. The input format is designed primarily for a *partitional* clustering algorithm (i.e., *K*-means or Self-Oraganizing Maps) but a clustering result from a hierarchical algorithm (i.e., dendrogram) may be applied by choosing a threshold carefully. One can search disease-related pathways and their correlations by entering a disease name.

#### Output

ArrayXPath produces a list of the best-matching pathways and GO terms for each cluster with statistical significance scores of non-random association. Relevant pathways are listed in ascending order of p-values (and multiple-comparison corrected q-values) (11). ArrayXPath provides a summary statistic for the overall mapping between input clusters and all pathways and GO terms matched.

If one chooses a pathway among the list, ArrayXPath outputs a Javascript-enabled SVG file, color-coded both by expression level and by cluster membership at each pathwaynode level. If one chooses a cluster, ArrayXPath outputs cluster-pathway-disease diagram with significantly associated GO terms and OMIM information (Fig. 1(c)). The cluster-centric view visualizes the related genes, pathways and diseases by measuring the shared membership of gene products. The whole quinta-partite associations (Fig. 1(a)) can be interactively navigated by choosing one of the cluster, pathway or disease nodes from the graph in SVG.

Each node in pathway graph and correlation multi-partite graph is enriched with a hyperlink to an automated summary page for the corresponding gene product(s)

provided by our integrated database: GRIP (Genome Research Informatics Pipeline, http://grip.snubi.org/) (7).

# METHODS

## Pathway integration and resolving diverse identifiers

ArrayXPath searches publicly available major pathway resources including KEGG, GenMAPP, BioCarta and PharmGKB Pathways. We have created a repository of metainformation by parsing SBML files for KEGG and HTML files for GenMAPP (http://www.genmapp.org/MAPPSet-Human/MAPP index.htm) and BioCarta (http://www.biocarta.com/genes/allPathways.asp), manually encoding and by PharmGKB pathways (http://www.pharmgkb.org/search/pathway/pathway.jsp). A variety of gene-product identifiers including GenBank accession number, UniGene ID, LocusLink ID, EC number, official gene symbol, SwissProt ID and TrEMBL ID are inconsistently used for the pathway nodes as well as microarray probes, resulting in enormous ambiguity in integrating data from different resources. By integrating major databases including GenBank, UniGene, LocusLink Homologene, SwissProt, Ensemble, UCSC Golden Path and NetAffyx (http://www.affymetrix.com/analysis/index.affx), ArrayXPath automatically matches the probe identifiers of microarray data to the identifiers of pathway nodes. When a pathway node is a composite type, i.e. consists of more than one element, ArrayXPath separately matches and visualizes each probe identifier to the corresponding individual element of the composite object.

Table 1 shows the distribution of the pathway nodes identified from KEGG, GenMAPP, BioCarta and PharmGKB Pathways for *Homo sapiens*. We found 1,942 redundant nodes representing genes and proteins for the 45 GenMAPP pathways. Among the 1,454 non-redundant elements, ArrayXPath successfully assigned 1,391 gene products (95.7%) to either official gene symbols (n = 1,329; 91.4%), LocusLink ID (n = 39; 2.7%), or SwissProt ID (n = 23; 1.6%). Only 63 (4.3%) remain unresolved because of intractable ambiguity. KEGG has 256 non-composite (i.e. simple) and 121 composite elements (i.e. enzymes), containing 256 and 505 gene products, respectively. Among the 256 simple-type elements, 21 appear as members of composite type elements. Overall, KEGG has 740 unique elements and ArrayXPath successfully assigned all of them (100%) either to official gene symbol (n = 720, 97.3%) or LocusLink (n = 20, 2.7%). PharmGKB Pathways added 133 official gene symbols and 1 LocusLink ID.

Overall, ArrayXPath identified 3,025 gene products for the four major pathways. We created a pre-computed association table of these elements to all resolvable IDs and to official gene symbols for reliable mapping of incoming microarray-probe identifiers.

## Search pathways by disease name in MeSH (PathMeSH)

ArrayXPath allows one to search disease-related pathways. The OMIM Morbid Map (<u>http://www.ncbi.nlm.nih.gov/Omim/getmorbid.cgi</u>) contains official gene symbol, alias gene symbol and cytogenetic location of disease-related genes with OMIM ID and the related disease name. The C category in the 15 branches of MeSH contains disease

names and their entry terms with hierarchical structure. We extracted the gene- and disease-related information from OMIM and MeSH. We mapped the disease names by using exact keyword match method provided by MeSH. We mapped the disease-related genes onto the integrated pathway resources by using our integrated database, GRIP, as described above. Among the 3,259 official gene symbols resolvable from the Morbid Map, we found that 2,395 genes had disease names that could be mapped to MeSH disease names through headings or entry terms. We successfully mapped 1,928 genes onto both MeSH disease names and pathway nodes (i.e. gene products). It means that about 64% (i.e. 1,928/3,025) (Table 1) of the non-redundant nodes in all pathway resources of our species have at least one link to human pathophysiology through standard disease name(s) in MeSH.

If the input disease name is matched to the corresponding MeSH heading or entry term, ArrayXPath outputs the list of the pathways containing the disease-related gene product (Fig. 1(b)). ArrayXPath determines the statistical significance of the association between a pathway and a disease name in terms of the non-random proportion of matched entities. ArrayXPath applies Fisher's exact test by constructing a  $2 \times 2$  contingency table containing the two pathway memberships (within and without the pathway) as column variables and the disease memberships (within and without the disease) as row variables. We used Fisher's exact test because a large sample approximation is inappropriate in the pathway case (a  $2 \times 2$  table often contains a cell with expected values < 5).

# Visualization of the correlational structure of biological pathways

Visualizing the crosstalk among pathways may reveal important biological understanding (12). We created pairwise similarity matrix of pathway distances by calculating the ratio of the number of the genes in the intersection divided by that in the union of each pair of pathways. Multi-dimensional scaling of the similarity matrix and drawing the edges of the pathway pairs above certain similarity threshold create a global crosstalk graph among all biological pathways (Fig. 2(a)). ArrayXPath interactively visualizes the local crosstalk of the pathway crosstalk graph (Fig. 2(b)).

ArrayXPath interactively visualizes the detailed local crosstalk. The shared membership of gene products in gene-expression clusters, pathways and disease names can be captured by multi-partite graph representations in SVG (Fig. 1(a), 1(c)). By selecting a node from one view, one can interactively navigate the different views of the whole associations.

# Mapping GO-based annotations

Among the 33,108 LocusLink IDs (as of 2005 February), we identified 3,025 gene products (i.e.  $\sim$ 9%) in major pathway resources (Table 1). Integrating GO-based annotations covering 13,949 LocusLink IDs (as of 2005 February) can help filling the gaps for pathway-based microarray data analysis. ArrayXPath provides both implicit and explicit GO annotations (Fig. 1(c)) (13). While explicit annotation provides the GO

terms directly mapped to the members of gene expression clusters, implicit annotation considers all ancestor terms from the GO hierarchical tree structure, providing general understanding. We applied hypergeometric distribution to evaluate the statistical significance of the associations.

#### DISCUSSION

ArrayXPath is a web-based service for mapping and visualizing microarray gene expression clusters with biomedical ontologies and major biological pathway resources using SVG. It permits one to input a clustered gene expression data in a tab-delimited text format via an Internet connection.

We found that integrating biomedical ontologies including the GO-based annotations, disease names supported by MeSH, and the genotype-to-phenotype information from OMIM Morbid Map greatly improve the capability of ArrayXPath to interpret gene expression profiles. Integrated analysis and interactive visualization of the global and local crosstalks among pathways can facilitate system-level understanding of microarray gene-expression data. Although we evaluated the statistical significance of each association in the present study, combined analysis may improve the inference. It is required in the future study to develop a computational method to reconstruct the whole correlational structure and extract more biology from gene-expression microarray data. Standard web-based integration of a wide range of bioinformatics modules and heterogeneous genomic data will obviously help advance biological science.

#### Acknowledgements

Authors thank all SNUBI members and external testers for their effort to input and evaluate the pathway diagrams. This study was supported by a grant from Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (0412-MI01-0416-0002).

#### REFERENCES

- 1. Roy, P.J., Stuart, J.M., Lurd, J. and Kim, S.K. (2002) Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. *Nature*, **418**, 975–979.
- 2. Lercher, M.J., Urrutia, A. and Hurst, L.D. (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet.*, **31**, 180–183.
- 3. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. and Church, G.M. (1999) Systematic determination of genetic network architecture. *Nat. Genet.*, **22**, 281–285.
- 4. Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D. and Friedman, N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, **34**, 166–176.
- 5. Mewes,H.W., Frishman,D., Guldener,U., Mannhaupt,G., Mayer,K., Mokrejs,M., Morgenstern,B., Munsterkotter,M., Rudd,S. and Weil,B. (2002) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.*, **30**, 31–34.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, 25, 25–29.

- Chung,H.J., Kim,M., Park,C.H., Kim,J. and Kim,J.H. (2004) ArrayXPath: mapping and visualizing microarray gene expression data with integrated biological pathway resources using Scalable Vector Graphics. *Nucleic Acids Res.*, 32, W464-W464.
- 8. Dennis, G.Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, R60.
- 9. Zeeberg, B.R., Feng, W., Wang, G., Wang, M.D., Fojo, A.T., Sunshine, M., Narasimhan, S., Kane, D.W., Reinhold, W.C., Lababidi, S. *et al.* (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.*, **4**, R28.
- 10. Hewett, M., Oliver, D.E., Rubin, D.L., Easton, K.L., Stuart, J.M., Altman, R.B., Klein, T.E. (2002) PharmGKB: the Pharmacogenetics Knowledge Base. *Nucleic Acids Res.* **32**, 163-165.
- 11. Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA.*, **100**, 9440–9445.
- 12. Zheng, C.J., Zhou, H., Xie, B., Han, L.Y., Yap, C.W. and Chen, Y.Z. (2004) TRMP: a database of therapeutically relevant multiple pathway. *Bioinformatics*, **20**, 2236-2241.
- 13. Robinson, P.N., Wollstein, A., Böhme, U. and Beattie, B. (2004) Ontologizing gene-expression microarray data: characterizing clusters with Gene Ontology, *Bioinformatics*, **20**, 979-981.

Figure 1. ArrayXPath functions. (a) Gene-cluster, gene-pathway, gene-disease and gene-GO associations (solid lines) are the building blocks of the quinta-partite graph representation used by ArrayXPath integration. Dotted lines explain how the associations are created. GO and MeSH have their own hierarchical organizations and clusters can be organized by profile similarity measures (broken circular arrows). (b) PathMeSH returns a list of disease-related pathways with statistical significance scores by integrating pathway resources, MeSH disease names, and OMIM Morbid Map. (c) When one chooses a cluster, ArrayXPath outputs the cluster-centric view of the associations of related genes, pathways and diseases through the shared membership of gene products. The whole quinta-partite associations can be interactively navigated by choosing cluster, pathway or disease node from the graph in SVG. ArrayXPath also provides GO-based annotation and OMIM information to complement pathway-based analysis of gene expression clusters.

Figure 2. Pathway crosstalk. (a) Calculating pairwise similarity matrix between each pair of pathways and applying multi-dimensional scaling method created the global crosstalk graph of major biological pathways. Yellow nodes represent BioCarta, green nodes GenMAPP, red nodes KEGG, and blue nodes PharmGKB Pathways (see methods). (b) ArrayXPath interactively visualizes the local crosstalk (i.e. purple lines) of the pathways associated with the selected clusters, superimposedly on the global pathway crosstalk graph.

Pathway		Gene / Protein			ID resolution					Metabolite		Embedded		Free text	
		simple	compl	redundant	Total	OGS	LL	SP	UR			patl	nway	desc	ription
			ex												
		(256)*		(637)											
KEGG	70	(505)*	(121	) (469)	740	720	20	0	0	1,896	(2,624)	0	(0)	121	(275)
		740		(1,106)											
GenMAPP	45	1,454		(1,942)	1,391	1,329	39	23	63	83	(97)	4	(4)	130	(372)
BioCarta	346	1,584		(8,976)	1,584	1,580	4	0	0	0	(0)	50	(141)	18	(53)
PharmGKB	9	134		(189)	134	133	1	0	0	11	(25)	1	(1)	23	(26)
Overall	470	3,025		(12,900)	3,088	2,938	64	23	63	1,990	(2,746)	55	(146)	55	(146)

Table 1. Distribution of pathway-node identifiers among the major pathway resources

\* There were 21 elements redundant in the simple (256) and composite (505) elements so that 740 unique elements were found in the 70 KEGG human pathways.

Numbers in parentheses are redundant counts. KEGG has 121 composite elements containing 505 identifiable gene products.

OGS: official gene symbol, LL: LocusLink, SP: SwissProt, UR: unresolved

Number: Table 1.

First author: Hee Joon Chung



(b)

Image: Second						
tp://blosing.anubl.org 8880/PathMoSH/pathway_based_vtew.jpp  PathMoSH [PathMosyH [PathMosyH [PathMosyM [PathMosyM [PathMosyM [PathMosyM [PathMosyM [PathMosyM [PathMosyM ] PathMosyM ] PathMosyM [PathMosyM ] PathMosyM [PathMosyM ] PathMosyM [PathMosyM ] PathMosyM [PathMosyM ] PathMosyM ]	😳 - 📕 🖉 🞧 🔑 감석 🏋 물거찾기 😽 미디어 🚱 🔯 - 8	🖕 🖂 📒 🚳				
ArrayXPath       IPathMeSH [Pathway-mapping [Help [ChromoViz [Redistration [Contact         PathMeSH         Search keyword is Breast Neoplasme that is disease name.         PathMeSH [Ind Sg pathways and search 18 genes.         This Vew is pathway-based gene information. Another view is gene-based pathway information.         Gene list [rel] (AM] (BRCA] (BRCA2] (CDH1] (CHR2] (ESR1] (RBICC1] (TPSI) (TSG101] (BACH1] (BRIP1] (CDS1] [PHB] (PPMID] (RADSA] (SLC22AIL] (DRCC3]         Select ordering       © Pvalue       • Relative Risk         Pathway       Gene Symbol Drug       OMIM       Cytoband       P-value Relative R         BicCarta/Hs_Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility       ATM       607505       11q22.3       0.0000001       55.0183         BicCarta/Hs_Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility       ATM       607505       11q22.3       0.0000000       45.2080         BicCarta/Hs_Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility       ATM       607565       11q22.3       0.0000000       45.2080         BicCarta/Hs_Cell Cycle: G2/M Checkpoint       ATM       607565       11q22.3       0.0000000       45.2080         BicCarta/Hs_Cell Cycle: G2/M Checkpoint       ATM       607565       11q22.3       0.0000000       45.2080         BicCarta/Hs_Cell Cycle: G2/M Checkpoint       ATM       607565       11q22.	ttp://blosvg.snubl.org:8080/PathMeSH/pathway_based_view,jsp					×
PathMeSH         Search keyword is Breast Neoplasms that is disease name.         PathMeSH lind SQ pathways and search 18 genes.         PathMeSH lind SQ pathways and search 18 genes.         This Vew is pathway-based gene information. Another view is gene-based pathway information.         Gene list (RR) [ATM] [BRCA1] [BRCA2] [CDH1] [CHE2] [ESR1] [RBICC1] [TP53] [TISG101] [BACH1] [BRIP1] [CDS1] [PHB] [PPMID] [RAD541] [SLC22AIL] [DRCC3]         Solect ordering	ArrayXPath	I <u>PathMeSH</u> I <u>Pa</u>	thway mapping	I <u>Help</u> I <u>ChromoVi</u>	z I <u>Registration</u>	IContact us
Search keyword is Breast Neoplasms that is disease name. PathMeSH find S2 pathways and search 18 genes. This view is pathway-based gene information. Another view is gene-based pathway information.         Search keyword is Breast Neoplasms that is disease name.           Gene list [PH] (AIM] (BRCA1] (BRCA2] (CDH1] [CHEK2] [ESR1] [RBICC1] [TP6S] [TSG101] (BACH1] (BRIP1] [CDS1] [PHB] (PPMID] [RAD54L] [SLC22A1L] [DRCC3]         Steect ordering              • Pealue             • Relative Risk             • Pathway             • O Relative Risk             • InficA1             • 007505             • 11q22.3             0.000001             • 55.0183             • 0000001             • 55.0183             • 000000             • 00000	PathMeSH					
United to the constraint of the constraint	Search keyword is <b>Breast Neoplasms</b> that is disease name. PathMeSH find 32 pathways and search 18 genes. This view is pathway-based gene information. Another view is ge	ene-based pathway informat	lion.			
Solar ci undoring         P Value         Relative Risk           Pathway         Gene Symbol         Drug         OMIM         Cytoband         P-value         Relative	ABU (ATM) [BRCA1] [BRCA2] [CDH1] [CHEK2] [ESR1] [RB1CC1] [TP	53] [TSG101] [BACH1] [BRIP1]	[CDS1] [PHB]	[PPM1D] [RAD54L]	[SLC22A1L] [XRC	C3]
Pathway         Gene Symbol         Drug         OMIM         Cytoband         P-value         Relative R           BioCarta//Hs_Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility         ATM         007095         11q22.3         0.000001         \$5.0183           BICCA1         113705         17q21         0.000001         \$5.0183         0.000001         \$5.0183           BICCA1         113705         17q21         0.000001         \$5.0183         0.000001         \$5.0183           BIOCarta//Hs_ATM Signaling Pathway         ATM         607595         11q22.3         0.000000         \$5.0183           BIOCarta//Hs_CEI Cycle: G2/M Checkpoint         ATM         607595         11q22.3         0.000000         \$5.0183           BIOCarta//Hs_CEI Cycle: G2/M Checkpoint         TPS3         191170         17p13.1         0.000000         \$5.0183           BIOCarta//Hs_CEI Cycle: G2/M Checkpoint         BICA1         113705         11q22.3         0.000000         \$5.0183           BIOCArta//Hs_CEI Cycle: G2/M Checkpoint         BIOCA         119170         17p13.1         \$5.0183           BIOCArta//Hs_CEI Cycle: G2/M Checkpoint         BIOCA         11922.3         0.000000         \$5.0183           BIOCARTA//Hs_CEI Cycle: G2/M Checkpoint         BIOCA         11922	Select ordering	k				
Partivacy         Gene Symbol         Drug         OMM         Cytoband         P-value         Relative R           BinCarta//Hs_Role of BRCAL_BRCA2 and ATR in Cancer Susceptibility BinCA1         ATM         607505         11q22.3         0.0000001         55.0183           BRCA1         113705         17q21         BRCA2         600165         15q12.3         0.0000001         55.0183           BRCA2         600165         15q12.3         0.0000001         55.0183           BRCA2         600173         22q12.1         17953         191170         17p13.1           BloCarta//Hs_ATM Signaling Pathway         ATM         607595         11q22.3         0.0000000         45.2030           BRCA1         115705         17q21         191170         17p13.1         17p13.1           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         ATM         607595         11q22.3         0.0000000         35.72611           BRCA1         115705         17q21         17p13.1         17p13.1         0.0000000         35.72611           BRCA1         115705         17q21         17p13.1         0.0000000         35.72611           BRCA1         115705         17q21         17p13.1         17p13.1         17p13.1						
BisCarta//Hs_Role of BRCAL_BRCA2 and ATR in Cancer Susceptibility         ATM         607595         11q22.3         0.0000001         55.0183           BISCAL         113705         17q21         19412.3         0.0000001         55.0183           BISCAL         113705         17q21         19412.3         0.0000001         55.0183           BISCAL         113705         17q21         19412.3         0.0000000         45.0183           CHEN2         604373         22q12.1         1753         191170         17p13.1         0.0000000         45.2030           BISCAL         113705         17q21         0.0000000         85.0183         191170         17p13.1           JISCarta//Hs_ATM Signaling Pathway         ATM         607595         11q22.3         0.0000000         45.2030           BISCAL         115705         17q21         17p13.1         17p13.1         17p13.1           JISCarta//Hs_Cell Cycle: G2/M Checkpoint         ATM         607595         11q22.3         0.0000000         35.72611           BISCA1         113705         17q21         17q21         17q21         17q21         17q21           BISCA1         113705         17q21         17q21         17q21         17q21         17q21	<sup>2</sup> athway	Gene Symbol Drug	омім	Cytoband	P-value I	Relative Ris
BICA1         11705         17021           BRCA2         600185         13q12.3           CIER2         600193         22q12.1           TP53         191170         17p13.1           BiCCats//Hs_ATM Signaling Pethway         ATM         607595         11q22.3         0.0000000         45.2030           BiCCats//Hs_ATM Signaling Pethway         ATM         607595         11q22.3         0.0000000         45.2030           BiCCats//Hs_Cell Cycle: 62/M Checkpoint         TTP53         191170         17p13.1         17p13.1           BiCCats//Hs_Cell Cycle: 62/M Checkpoint         607595         11q22.3         0.0000000         35.72611           BiCCA1         113705         17q21         0.0000000         35.72611           BiCCA1         113705         17q21         0.0000000         35.72611           BiCA1         113705         17q21         0.00	BioCarta//Hs_Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility	ATM	607585	11q22.3	0.0000001	55.018315
BBC/2         B01/6         13a12.3           CHEK2         604973         22a12.1           TP53         191170         17o13.1           BioCarta//Hs_ATM Signaling Pethway         ATM         607595         11a22.3         0.000030         45.2030           BioCarta//Hs_Cell Cycle: G2/M Checkpoint         ATM         607595         11a22.3         0.000000         45.2030           BinCArt         113705         17a21         17p13.1         17p13.1         0.000000         95.7261           BinCArt         113705         17a21         0.000000         95.7261           Encert         119170         17p13.1         17p13.1         17p13.1		BRCAL	113705	17q21		
CLEV2 TP53         604973 191170         22012.1           BloCarta//Hs_ATM Signaling Pathway         ATM BRCA1         607595         11022.3         0.0000000         45.2030           CHEV2 TP53         191170         17021         2012.1         17021         0.0000000         45.2030           SteCarta//Hs_Cell Cycle: G2/M Checkpoint         ATM BRCA1         607595         11022.3         0.000000         95.7261           BRCA1         113705         17021         0.000000         95.7261           ERCA1         113705         17021         0.000000         17013		00010	600195	19-12.2		
TP53         191170         17p13.1           BioCarta//Hs_ATM Signaling Pathway         ATM         607995         11q22.3         0.000000         45.030           BioCarta//Hs_Cill Cycle: G2/M Checkpoint         17p3         117p13.1         17p13.1         17p13.1           BioCarta//Hs_Cell Cycle: G2/M Checkpoint         6000000         17p13.1         17p13.1         6000000         45.030           BioCarta//Hs_Cell Cycle: G2/M Checkpoint         600000         17p13.1         0.000000         35.7261           BioCarta//Hs_Cell Cycle: G2/M Checkpoint         600000         6000000         35.7261           BioCarta//Hs_Cell Cycle: G2/M Checkpoint         604373         22q12.1           TP53         1		BHCA2	000103	10412-0		
ATM         607595         11q22.3         0.000000         45,030           BRCA1         113076         17q21         17q21         17q21         17q21           CirEIC2         604373         22q12.1         17p13.1         17p13.1         0.0000000         45,030           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         ATM         607595         11q22.3         0.000000         35,7261           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         BDCA1         113705         17q21         0.000000         35,7261           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         BDCA1         113705         17q21         0.000000         35,7261           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         BDCA1         113705         17q21         0.000000         35,7261           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         BDCA1         113705         17q21         0.000000         35,7261           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         BDCA1         113705         11q22.3         0.000000         35,7261           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         BDCA1         113705         17q21         17q21           Checkpoint         TES3         191170         17p13.1         17q1         17q1         17q1		CHEK2	604373	22q12.1		
BRCA1         113(205         17a(21)           ChEC2         604373         22a(12.1)           TP53         19170         17p13.1           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         ATM         607595         11a(22.3)         0.0000000         35.72611           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         BLCA1         113705         11a(22.3)         0.0000000         35.72611           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         BLCA1         113705         11a(22.3)         0.0000000         35.72611           BLCA1         113705         11a(22.3)         0.0000000         35.72611           DFEC2         804373         22a(12.1)         11523         191170         17p13.1		CHEK2 TP53	604373 191170	22q12.1 17p13.1		
CHER2         604373         22q12.1           TP53         191170         17p13.1           BloCarta//Hs_Cell Cycle: G2/M Checkpoint         ATIM         607595         11q22.3         0.000000         35,7261           BloCA1         113705         17q21         17g13.1         17g13.1         17g13.1	lioCarta//Hs_ATM Signaling Pathway	CHEK2 TP53	604373 191170 607585	13q12.3 22q12.1 17p13.1	0.0000030	45.203007
TP53         191170         17p13.1           NoCarta//Hs_Cell Cycle: G2/M Checkpoint         ATM         607595         11q22.3         0.0000000         95.7261           BRCA1         113705         17q21           CHEK2         804373         22q12.1           TE53         191170         17p13.1	NoCarta//Hs_ATM Signaling Pathway	ATM BRCAL	604373 191170 607585 113705	19412.3 22q12.1 17p13.1 11q22.3 17q21	0.0000030	45.203007
ATM         607595         11q22.3         0.000000         35,7261           BRCAL         113705         17q21           CHEK2         804373         22q12.1           TES2         191170         17p13.1	NoCarta//Hs_ATM Signaling Pathway	CHEK2 TP53 ATM BRCA1 CHEK2	604373 191170 607585 113705 604373	10q12.3 22q12.1 17p13.1 11q22.3 17q21 22q12.1	0.0000030	45.203007
BRCA1         113705         17q21           CHER2         B04373         22q12.1           TP53         191170         17p13.1	NoCerta//Hs_ATM Signaling Pethway	ATM BRCAL CHEK2 TP53 ATM CHEK2 TP53	604373 191170 607585 113705 604373 191170	10412.3 22q12.1 17p13.1 11q22.3 17q21 22q12.1 17p13.1	0.0000030	45.203007
CHER2         804373         22q12.1           TEES         191170         17p13.1	NoCarta//Hs_ATM Signaling Pathway NoCarta//Hs_Cell Cycle: G2/M Checkpoint	ATM BRCA1 CHEK2 TP53 ATM BRCA1 CHEK2 TP53 ATM	604373 191170 607585 113705 604373 191170 607585	13412.3 22q12.1 17p13.1 11q22.3 17q21 22q12.1 17p13.1	0.0000030	45.203007
<u>TP53</u> 191170 17p13.1	lloCarta//Hs_ATM Signaling Pathway lloCarta//Hs.Cell Cycle: G2/M Checkpoint	ATM BRCAL CHEK2 TP53 ATM BRCAL TP53 ATM BRCAL	604373 191170 607585 113705 604373 191170 607585 113705	13412.3 22q12.1 17p13.1 11q22.3 17q21 22q12.1 17p13.1 11q22.3 17q21	0.0000030	45.203007
	lioCarta//Hs_ATM Signaling Pathway lioCarta//Hs_Cell Cycle: G2/M Checkpoint	ATM BRCAL CHEK2 TP53 ATM BRCAL CHEK2 TP53 ATM BRCAL CHEK2	604373 191170 607585 113706 604373 191170 607585 113705 604373	11q22.3 11q22.3 11q22.3 17q21 22q12.1 17p13.1 11q22.3 17q21 22q12.1	0.0000030	45.203007
3ioCarta//Hs_Regulation of cell cycle progression by PIK3 ATM 607595 11o22.3 0.0000089 75.4250	NoCarta//Hs_ATM Signaling Pathway NoCarta//Hs_Call Cycle: G2/M Checkpoint	ATM BRCAL CHEK2 TP53 ATM BRCAL CHEK2 TP53 ATM BRCAL CHEK2 TP53	60/373 191170 607585 113705 604373 191170 607585 113705 604373 191170	114223 226[2] 176131 17421 226[2] 226[2] 17421 176131 114223 17421 226[2] 274[2] 274[2] 274[2] 176131	0.0000030	45.203007 35.726190
CHEK2 604373 22012.1	NoCarta//Hs_ATM Signaling Pathway NoCarta//Hs_Cell Cycle: G2/M Checkpoint	ATM BRCAL CHEK2 TP53 ATM BRCAL CHEK2 TP53 ATM BRCAL CHEK2 TP53 ATM	604373 191170 607585 113705 604373 191170 607585 113705 804373 191170 607595	114223 226[2] 176131 114223 17421 226[2] 176131 114223 17421 226[2] 176131 114223	0.0000030	45.203007 35.726190 75.425000
TP53 191170 17p18.1	NoCarts//Hs_ATM Signaling Pethway NoCarts//Hs_Cell Cycle: G2/M Checkpoint NoCarts//Hs_Regulation of cell cycle progression by PK2	ATM CHEK2 IP53 ATM BRCA1 CHEK2 IP53 ATM BRCA1 CHEK2 IP53 ATM CHEK2 IP53	60173 60173 191170 607585 113705 604373 191170 607585 614373 191170 607505 604373	114223 226[2] 170131 114223 17421 226[2] 17621 17621 114223 17621 226[2] 114223 17621 114223 226[2]	0.0000030	45.203007 35.726190 75.425000

Number: Figure 1.

First author: Hee Joon Chung

(c)



Number: Figure 1.

First author: Hee Joon Chung





#### Number: Figure 2. First author: Hee Joon Chung