

## ORIGINAL ARTICLE

## PATHOME: an algorithm for accurately detecting differentially expressed subpathways

S Nam<sup>1,8</sup>, HR Chang<sup>2,8</sup>, K-T Kim<sup>3</sup>, M-C Kook<sup>4</sup>, D Hong<sup>1</sup>, CH Kwon<sup>1</sup>, HR Jung<sup>2</sup>, HS Park<sup>2</sup>, G Powis<sup>5</sup>, H Liang<sup>6</sup>, T Park<sup>7</sup> and YH Kim<sup>2</sup>

The translation of high-throughput gene expression data into biologically meaningful information remains a bottleneck. We developed a novel computational algorithm, PATHOME, for detecting differentially expressed biological pathways. This algorithm employs straightforward statistical tests to evaluate the significance of differential expression patterns along subpathways. Applying it to gene expression data sets of gastric cancer (GC), we compared its performance with those of other leading programs. Based on a literature-driven reference set, PATHOME showed greater consistency in identifying known cancer-related pathways. For the WNT pathway uniquely identified by PATHOME, we validated its involvement in gastric carcinogenesis through experimental perturbation of both cell lines and animal models. We identified HNF4 $\alpha$ -WNT5A regulation in the cross-talk between the AMPK metabolic pathway and the WNT signaling pathway, and further identified WNT5A as a potential therapeutic target for GC. We have demonstrated PATHOME to be a powerful tool, with improved sensitivity for identifying disease-related dysregulated pathways.

*Oncogene* (2014) 33, 4941–4951; doi:10.1038/onc.2014.80; published online 31 March 2014

## INTRODUCTION

To obtain system-level descriptions of biological and cellular events, transcriptome profiling, including recent RNA-sequencing approaches,<sup>1</sup> has generated large, unmanageable high-throughput data sets.<sup>2</sup> To make sense of these gene expression data sets, researchers have developed various gene set analysis (GSA) tools, such as DAVID<sup>3</sup> and GSEA,<sup>4</sup> in combination with prior knowledge-based databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG),<sup>5</sup> Gene Ontology,<sup>6</sup> BioCarta, PANTHER,<sup>7</sup> MetaCyc,<sup>8</sup> Molecular Signatures Database<sup>4</sup> and RegulonDB.<sup>9</sup>

Despite these diligent research efforts, the biologically meaningful interpretation of findings from high-throughput gene expression data remains a bottleneck.<sup>2</sup> The persistence of this congestion arises from the challenge of exploring the complex relationships between cellular components,<sup>10</sup> especially in the context of functional molecular pathways. Pathway information as it relates to a phenotype of interest (for example, a disease) necessarily implies that a key molecular target should be considered within the framework of its network. A network focus enables us to more effectively infer key transcriptional changes related to the specific phenotype by examining multiple downstream (or cross-talk) effectors of the target. However, the current GSA<sup>11</sup> tools utilize over-representation analysis,<sup>12</sup> which reports the enrichment of functional groups (for example, gene sets) for the genes of interest. They compromise the connectivity in favor of computational simplicity that is based on cellular components and not their connectivity.<sup>3,4</sup> In other words, current tools do not analyze the wiring diagram of the interactions (for example, activation, inhibition) in a functional molecular network.<sup>13</sup>

We have developed PATHOME (*pathway* and *transcriptome* information), a novel computational algorithm for identifying differentially expressed subpathways. Methodologically, PATHOME has two benefits: It analyzes the regulation information between nodes in the biological pathways and is applicable to a small number of samples. PATHOME is not a permutation-based approach that requires more samples in order to obtain a null distribution for a statistical test. We demonstrated the utility of PATHOME by applying it to gene expression data of gastric cancer (GC), thereby identifying tumor-related dysregulated pathways and novel therapeutic targets. Based on a reference set of known cancer-related pathways, PATHOME showed greater sensitivity and robustness than other leading methods in detecting differential molecular signals. For the WNT signaling pathway revealed only by PATHOME, we validated its involvement in gastric carcinogenesis through experimental studies of both cell lines and animal models. Our results further revealed a potential therapeutic target, WNT5A. Thus, PATHOME represents a powerful tool for inferring biologically interpretable patterns from gene expression data.

## RESULTS

## Overview of PATHOME algorithm and study design

PATHOME takes the gene expression profiles of two comparison groups (for example, cancer vs non-cancer tissue) and related biological pathways from prior knowledge. In this study, we used the KEGG pathway database as the source of prior knowledge. PATHOME first decomposes the pathways into linear paths (subpathways) from the top nodes to leaf nodes, and then

<sup>1</sup>Cancer Genomics Branch, National Cancer Center of Korea, Goyang-si Gyeonggi-do, Republic of Korea; <sup>2</sup>New Experimental Therapeutics Branch, National Cancer Center of Korea, Goyang-si Gyeonggi-do, Republic of Korea; <sup>3</sup>Molecular Epidemiology Branch, National Cancer Center of Korea, Goyang-si Gyeonggi-do, Republic of Korea; <sup>4</sup>Department of Pathology, National Cancer Center of Korea, Goyang-si Gyeonggi-do, Republic of Korea; <sup>5</sup>Cancer Center, Sanford-Burnham Medical Research Institute, La Jolla, CA, USA; <sup>6</sup>Department of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, TX, USA and <sup>7</sup>Department of Statistics, Seoul National University, Kwanak-gu Seoul, Republic of Korea. Correspondence: Professor H Liang or Professor T Park or Professor YH Kim, New Experimental Therapeutics Branch, National Cancer Center of Korea, Goyang-si Gyeonggi-do 410-769, Republic of Korea.

E-mail: hliang1@mdanderson.org or tspark@stats.snu.ac.kr or yhkim@ncc.re.kr

<sup>8</sup>These authors contributed equally to this study.

Received 24 September 2013; revised 11 February 2014; accepted 14 February 2014; published online 31 March 2014

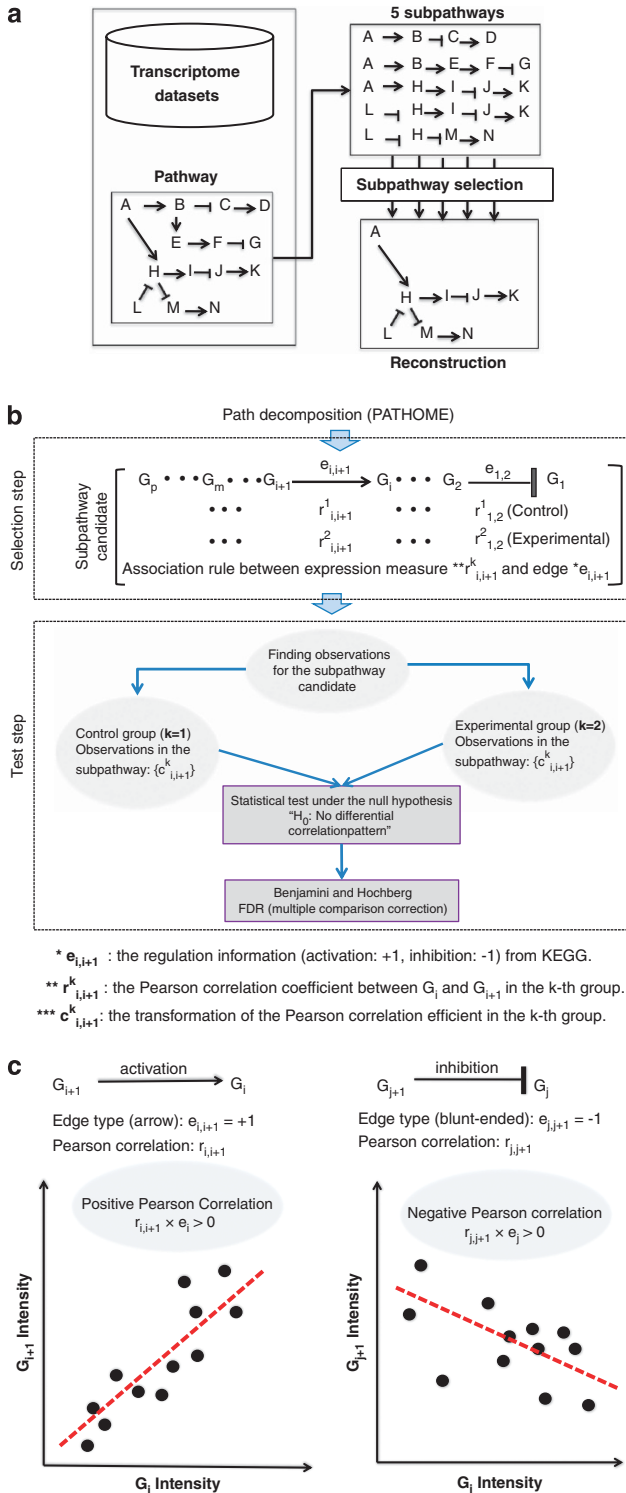
employs simple statistical tests to evaluate the significance of differential expression patterns along the subpathways (Figures 1a and b). The interaction property between pathway members (for example, activation or inhibition) is also considered (Figure 1c). A detailed description of the PATHOME algorithm is provided in Materials and methods.

To evaluate the performance and demonstrate the utility of PATHOME, we applied it to publicly available gene expression data sets of Asian GC (Table 1). Figure 2 summarizes our study strategy,

which consists of two stages: a discovery stage and a validation stage. The primary aim of the discovery stage is to compare the performance of PATHOME with those of DAVID<sup>3</sup> and GSEA,<sup>4</sup> two leading GSA tools based on a reference set of known cancer-related pathways. The data sets we used in this stage were one Korean GC data set (GSE13861)<sup>14</sup> and one Japanese GC data set (GSE15081).<sup>15</sup> A second aim of this two-stage strategy was to select potential therapeutic targets in GC. Thus, based on differential subpathways inferred from the discovery stage, we further identified robust gene signatures by using independent data sets at the validation stage. The data sets we used in this stage were another Korean GC data set (GSE36968)<sup>16</sup> and one Chinese GC data set (GSE27342).<sup>17</sup> Finally, through hierarchical clustering analysis, transcription factor-binding site (TFBS) analysis and expression pattern analysis from ArrayExpress,<sup>18</sup> we identified the target gene/pathway for experimental validation.

Performance comparison of PATHOME with other algorithms

The comparison sample groups in the first gene expression data set (GSE13861)<sup>14</sup> were 65 primary gastric adenocarcinoma frozen tissue samples and 19 normal appearing gastric tissue samples. With PATHOME, we identified 113 810 subpathways that belong to 27 KEGG pathways (Supplementary Table 1) at a false discovery rate (FDR) < 0.05. To the same data set, we applied DAVID and



**Table 1.** Summary of gastric cancer-related transcriptome-wide expression data sets used in this study

Data set	Ethnic group	Comparison group	Profiling platform
GSE13861 <sup>14</sup>	Korean	65 Tumor vs 19 non-tumor	Illumina Human WG-6 v3.0
GSE15081 <sup>15</sup>	Japanese	18 Relapse vs 38 relapse-free	Human Oligo Chip 30K
GSE36968 <sup>16</sup>	Korean	24 Tumor vs 6 non-tumor	SOLiD Single-read RNA-seq
GSE27342 <sup>17</sup>	Chinese	80 Tumor vs 80 non-tumor	Affymetrix Human Exon 1.0 ST array

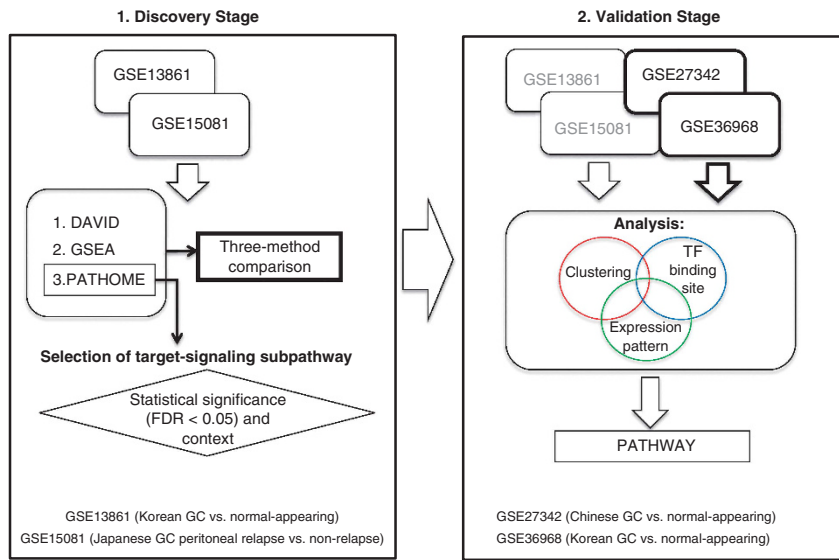
**Figure 1.** Overview of PATHOME. (a) Gene expression data and pathways are input to identify significant subpathways. The pathways are first decomposed to linear paths (subpathways) from top nodes to leaf nodes. Two types of regulation information are considered: activation (arrow-headed) and inhibition (blunt end-headed). Nodes A and L refer to root nodes, and nodes D, G, K, N refer to leaf nodes. Each subpathway is statistically tested for significance of its selection. In the example, two subpathways are selected and subsequently reconstructed for visualization. (b) PATHOME consists of two steps: a selection of candidate subpathways and a test of the selected subpathways. The selection step determines whether a given subpathway complies with the association rule (see c) between edge information and gene expression. If the rule is satisfied in the subpathway, statistical tests will be performed in the test step. (c) The rule associates regulation (edge) information of adjacent entries with their expressions in terms of the sign information of Pearson product-moment correlation coefficients. The graphical edge type comes from regulation information between the adjacent entries (here,  $e_{i,i+1}$  and  $e_{j,j+1}$  from pairs  $G_i - G_{i+1}$  and  $G_j - G_{j+1}$ , respectively) from KEGG pathway. The edge type is coded to +1 (activation) or -1 (inhibition) according to its head shape (arrow or blunt-ended). We used the sign information of Pearson correlation product coefficient (here,  $r_{i,i+1}$  and  $r_{j,j+1}$ ) as an edge type surrogate in the expression data. When the sign information of the coefficient and that of the edge type are the same, we say that the expression and the prior regulation information are agreeable in an experimental group.

GSEA, using default parameter settings (the results are shown in Supplementary Tables 2 and 3, at FDR < 0.3). The comparison groups in the second gene expression data set (GSE15081)<sup>15</sup> were the samples from 18 patients with GC who experienced peritoneal relapse and 38 who did not experience peritoneal relapse. At FDR < 0.05, PATHOME reported 126 095 subpathways that belong to 15 KEGG pathways (Supplementary Table 4). The significant pathways identified by DAVID and GSEA are shown in Supplementary Tables 5 and 6 (FDR < 0.3), respectively. To evaluate the performance of PATHOME, we used a set of known cancer-related pathways<sup>19</sup> as a reference standard for

comparing the three methods (see Table 2). PATHOME used a lower significance cutoff (FDR < 0.05) compared with that of the two methods (FDR < 0.3 for both methods). Despite the lower cutoff, our method detected more differential cancer-related pathways.

#### Selection of potential therapeutic targets in the WNT signaling pathway

Although other tools assess the overall enrichment of genes of interest in a given pathway, the strategy of PATHOME is to thoroughly inspect all possible paths in the pathway. Path



**Figure 2.** Overview of the study design. The study consists of two stages (a discovery stage and a validation stage). The discovery stage employed two independent gastric cancer data sets for methodological comparison. In this stage, we compared our method to two GSA tools, GSEA and DAVID. From target-signaling subpathways inferred from the discovery stage, the validation stage aims to identify a small number of genes with reproducible expression patterns through independent data sets, and further using clustering analysis and transcription factor-binding site (TFBS) analysis, and expression pattern analysis from ArrayExpress.

**Table 2.** Performance comparison of three methods based on terms of cancer-related pathways identified in Vogelstein and Kinzler<sup>19</sup>

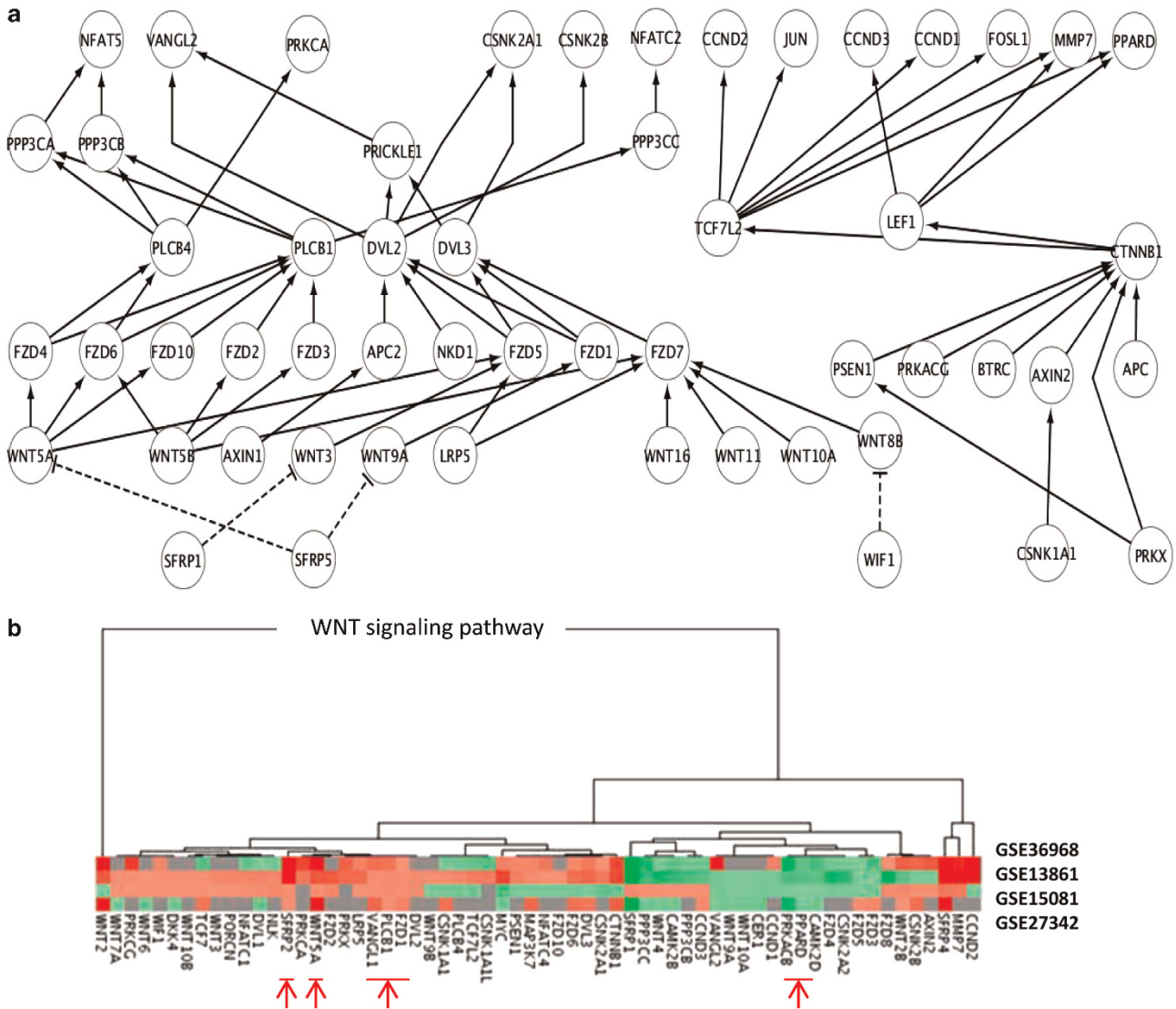
Reference standard <sup>19</sup>	KEGG pathway	Title	GSE13861 data set <sup>14</sup>			GSE15081 data set <sup>15</sup>		
			PATHOME	DAVID	GSEA	PATHOME	DAVID	GSEA
HIF1	HSA04150	mTOR signaling	X	X	X	X	X	X
	HSA05200	Pathways in cancer	0	X	X	X	X	X
	HSA05211	Renal cell carcinoma	X	X	X	X	X	X
P53	HSA04115	P53 signaling	X	X	X	X	X	X
RB (cell cycle)	HSA04110	Cell cycle	X	X	0	X	X	X
Apoptosis	HSA04210	Apoptosis	X	X	X	X	X	X
GLI	HSA04340	Hedgehog signaling	X	X	X	X	X	X
APC	HSA04310	Wnt signaling	0	X	X	0	X	X
RTK	HSA04012	ERBB signaling	X	X	X	X	X	X
SMAD	HSA05200	Pathways in cancer	0	X	X	X	X	X
	HSA04350	TGF-β signaling	X	X	X	X	X	X
PI3K	HSA04012	ERBB signaling	X	X	X	X	X	X
	HSA05200	Pathways in cancer	0	X	X	X	X	X
	HSA04150	mTOR signaling	X	X	X	X	X	X
	HSA04010	MAPK signaling	0	X	X	0	X	X
	HSA04910	Insulin signaling	0	X	X	X	X	X
	HSA04510	Focal adhesion	0	0	X	0	X	X
	HSA04062	Chemokine signaling	0	X	X	0	X	X
HSA04370	VEGF signaling	X	X	X	0	X	X	

Abbreviations: APC, antigen-presenting cell; ERBB, v-erb-b avian erythroblastic leukemia viral oncogene; GLI, glioma-associated oncogene; HIF1, hypoxia inducible factor; MAPK, mitogen-activated protein kinase; mTOR, mammalian target of rapamycin; PI3K, phosphoinositide-3-kinase; RB, retinoblastoma; RTK, protein-tyrosine kinase; TGF, transforming growth factor; VEGF, vascular endothelial growth factor. Note: X (not detected), 0 (detected).

decomposition in Figure 1a allows PATHOME to detect a pathway for which the enrichment analyses of other tools cannot report significance.

To further demonstrate the utility of our method, applied PATHOME to the identification of potential therapeutic targets from within the results we had obtained. For this purpose, PATHOME reported significant subpathways relating to WNT signaling, MAPK signaling, insulin signaling, focal adhesion and chemokine signaling (Table 2) in East Asians. Among these identified pathways, we selected the WNT pathway as identified uniquely by PATHOME for further cell line and animal studies for accuracy validation. Although the two discovery data sets were constructed from different comparisons of GC patients, the PATHOME results strongly suggested that subpathways of the WNT pathway were involved in the development of primary GC and peritoneal cancer relapse. Using Cytoscape,<sup>20</sup> we combined the significant subpathways related to the WNT pathway from the two data sets and visualized the combined subpathways of 62 genes (Figure 3a and Supplementary Table 7). At the validation

stage, we investigated whether the expression patterns in the combined subpathways inferred from the discovery stage were reproducible in another two independent data sets: our published RNA-seq GC data set of Korean patients (GSE36968: 24 primary cancer tissue samples vs 6 non-cancer samples)<sup>16</sup> and another microarray data set (GSE27342: 160 samples of paired gastric tumor and adjacent normal tissues).<sup>17</sup> We found that the expression changes in the 62 genes of the combined subpathways showed great concordance in terms of up- or downregulation (relative to non-cancer samples) between GSE13861 and GSE36968 (Fisher's exact text,  $P < 0.007$ , Supplementary Table 8) or between GSE13861 and GSE27342 (Fisher's exact text,  $P < 0.018$ , Supplementary Table 8), highlighting the reproducibility of differential WNT subpathways identified. We arbitrarily assigned the four data sets into either the discovery stage or the validation stage. Because of the considerable concordance among the data sets (Supplementary Table 8), this assignment yet enabled us to work in the current study design in spite of the heterogeneity of the origins of the data sets.



**Figure 3.** Selection of significant targets in the combined WNT signaling subpathways. **(a)** The combined differential subpathways of 62 genes, visualized by Cytoscape. **(b)** Identification of the commonly up- or downregulated gene clusters in the hierarchical clustering analysis of the four expression data sets (GSE36968, GSE13861, GSE15081 and GSE27342) in terms of the signed fold-changes for the 62 genes. The eight genes with consistent changes across the four data sets are marked by red arrows (color coding: gray, missing data; red, upregulation; green, downregulation)

To pinpoint individual genes for further experimental valuation, we performed hierarchical clustering analysis, TFBS analysis and expression pattern analysis from ArrayExpress.<sup>18</sup> The hierarchical clustering analysis (Figure 3b) found that eight genes showed up- or downregulation consistently across all four data sets used in both stages (Figure 3b, indicated by the red arrows), among which five genes (*WNT5A*, *VANGL1*, *SFRP2*, *FZD1* and *PLCB1*) were upregulated. We next focused on these five upregulated genes (as potential drug-inhibition candidates) and evaluated their TFBSs. Using the ENCODE transcription factor ChIP-Seq data<sup>21</sup> and the evolutionarily conserved TFBSs,<sup>22</sup> we examined the TFBSs within upstream 2-kb regions from the transcription start sites of the five genes. Interestingly, *VANGL1* has two binding sites for hepatocyte nuclear factor 4 alpha (*HNF4a*) according to the ChIP-Seq data; whereas *WNT5A* and *PLCB1* have binding sites for *HNF4a* in terms of conserved TFBSs (Supplementary Figure 1A). Thus, *VANGL1*, *WNT5A* and *PLCB1* are potential target genes in the WNT pathway through *HNF4a*-mediated transcriptional regulation. To further confirm the potential regulation of *HNF4a* with these three genes at the transcriptional level, we examined the expression patterns of the four genes (*HNF4a* and *VANGL1*, *WNT5A*, *PLCB1*) in various experimental data sets in a public expression warehouse, ArrayExpress.<sup>18</sup> We found that 27 out of 166 data sets reported by ArrayExpress showed co-expression of the four genes (Supplementary Figure 1B), which was statistically significant (binomial distribution,  $P < 1.2 \times 10^{-14}$ ). Supplementary Figure 3 shows that gene expression of *WNT5A*, *FZD1*, *PLCB1* and *VANGL1* was detected in eight GC cell panels.

Our previous study showed that the loss of the energy-sensing protein AMPK is a critical event in the initiation of GC; and that both mRNA and protein levels of *HNF4a* can be inhibited by metformin-mediated AMPK $\alpha$  activation in GC cells.<sup>16</sup> Thus, the above analyses revealed *HNF4a*-*WNT5A* as having a cross-talk function between the AMPK $\alpha$  energy-sensing pathway and the WNT signaling pathway. We hypothesized that a metformin effect on *HNF4a* by an AMPK $\alpha$ -mediated mechanism could transcriptionally downregulate the WNT pathway. To confirm this link, we selected *WNT5A* for further experimental validation (described in the following section).

#### Involvement of WNT pathway and associated antitumor activity in GC

We first examined the basal protein expression level of several WNT family members (including *WNT5A*) and WNT-related downstream genes in 14 gastric tumor cell line panels (Figure 4a). Although *WNT1* and *WNT3* protein expression levels varied between the GC cell lines, *WNT5A* showed a similar protein expression level between gastric cell lines through immunoblotting measurement. In addition, we examined the basal protein expression level of *WNT5A* in xenograft tumors of 16 gastric tumor cell lines, and found similar protein expression levels between each xenograft model through immunohistochemistry (Figure 4b).

As described above, *WNT5A* contains several highly conserved *HNF4a*-binding sites in its promoter region. Upon small interfering RNA (siRNA)-*HNF4a* treatment, the *WNT5A* gene expression level by real-time reverse transcription-PCR (RT-PCR) was downregulated by greater than 50% and 85% on days 2 and 3, respectively (Figure 4c), and the *TCF4* gene expression level was 25% downregulated on day 2. Meanwhile, the *WNT5A* protein level showed 21% and 52% inhibition on days 2 and 3, respectively (Figure 4d); whereas there was no difference in the protein level of  $\beta$ -catenin, and *HIF1a* was downregulated by 70% (Figure 4d). These results indicate that *HNF4a* as a transcription factor could regulate the transcription of *WNT5A*. However, different quantitation of knockdown and downregulation of gene expression level and protein level for *HNF4a* and *WNT5A* is observed in our experiments.

To investigate the functional relevance of *WNT5A* in gastric tumors, we performed metformin treatment in the tumor cell lines. Upon metformin treatment, both the NCI-N87 and AGS gastric tumor cell lines showed anti-proliferation activities (Figure 4e). While there were small changes in *WNT1* and *WNT3* protein levels, we observed 42% inhibition of *WNT5A* and 58% inhibition of  $\beta$ -catenin protein expression level, with indifferences seen for glycogen synthase kinase (GSK)3 protein levels in both the NCI-N87 and AGS cell lines (Figure 4e), also, detected loss of transcription factor (TCF) and lymphoid enhancer-binding factor (LEF) protein expression level on days 2 and 3 (Figure 4e). The WNT canonical signaling pathway acts by regulating cell differentiation, cell growth and cell proliferation. We observed the regulation of cyclin D1 and cyclin D2 protein levels through *WNT5A* (Figure 4e). Moreover, we detected minimal effects on the phosphorylated-calcium/calmodulin-activated protein kinase (pCAMKII) protein level in WNT non-canonical signaling pathway, as we observed 5 out of 14 gastric cell lines to have confirmed pCAMKII protein detection. Potentially in GC, *WNT5A* does not activate a CAMKII-dependent signaling cascade. However, we observed different level of  $\beta$ -catenin protein expression inhibition in comparison between metformin and knockdown of *HNF4a* through siRNA. But in both metformin and siHN4a, decreased level of *HIF1a* protein expression level was observed in our previous<sup>16</sup> and present studies, where  $\beta$ -catenin potentiates transcription of *HIF1a*. Further studies are warranted to confirm this supposition.

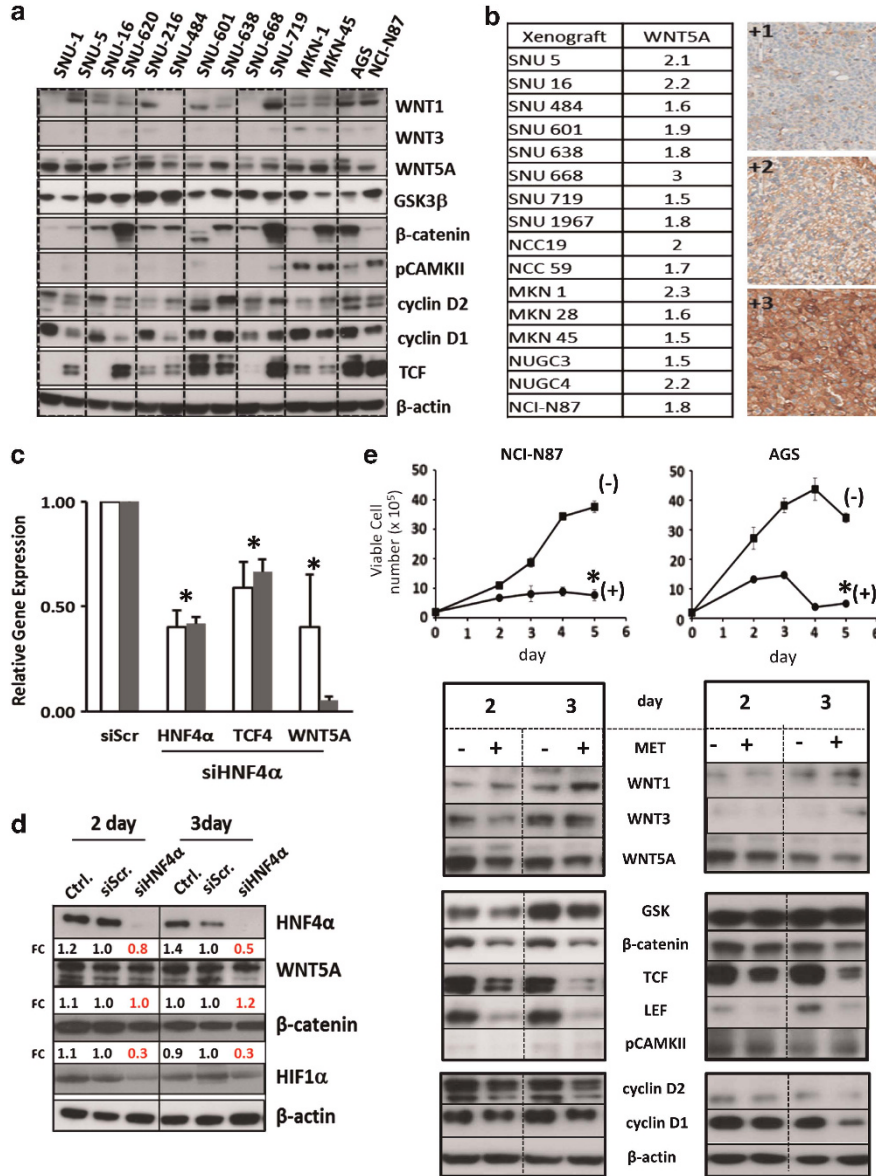
We silenced *WNT5A* expression using siRNA knockdown on the two GC cell lines (NCI-N87 and AGS) and performed cell proliferation assay. As shown in Figures 5a and b, siRNA-*WNT5A* affected cell growth through apoptosis. As shown in Figure 5c, siRNA-*WNT5A* affected protein translation of cell growth regulator cyclin D1 in both cell lines.

To further evaluate our findings in a more biologically relevant context, we performed xenograft experiments in mouse models. Consistently, we found that the metformin treatment introduced strong antitumor activity (Figure 6a, NCI-N87 and MKN-45 cell lines, respectively). Importantly, we found the regulatory relationship to be consistent with that observed in the cell line studies. Upon metformin treatment vs non-treatment, the protein level of *WNT5A* decreased in eight out of nine NCI-N87 xenograft animals and in six out of eight MKN-45 xenograft animals. Also, the detected  $\beta$ -catenin protein level decreased more than 45% in the NCI-N87 and 25% in the MKN-45 animal models compared with that in the mice without metformin treatment, taking the mean of the ratio of protein expression levels (Figures 6b and c). Furthermore, immunohistochemistry revealed a loss of *WNT5A* protein expression in mice treated with metformin (Figure 6d). In summary, these *in vivo* studies confirmed the involvement of *WNT5A* signaling in the WNT pathway in gastric tumorigenesis through *HNF4a*, highlighting the potential of *WNT5A* as a therapeutic target in GC.

#### DISCUSSION

We developed a novel algorithm, PATHOME, for sensitively detecting differentially expressed pathways. Biological pathways consist of nodes and edges representing gene information and regulation information (for example, activation, inhibition), respectively. Current GSA methods for pathway (or gene set) analysis, such as DAVID,<sup>3</sup> GSEA,<sup>4</sup> GoMiner,<sup>23</sup> Onto-Tools<sup>24</sup> and GeneTrail,<sup>25</sup> consider nodes only. In contrast, our method is able to consider both types of information stored in the pathways, thereby improving the sensitivity for detecting differential signals.

Disease occurs as a consequence of the dysregulation of a complex interdependency among biological components.<sup>10</sup> The identification of biologically targetable components requires the consideration of interdependency in terms of measurable



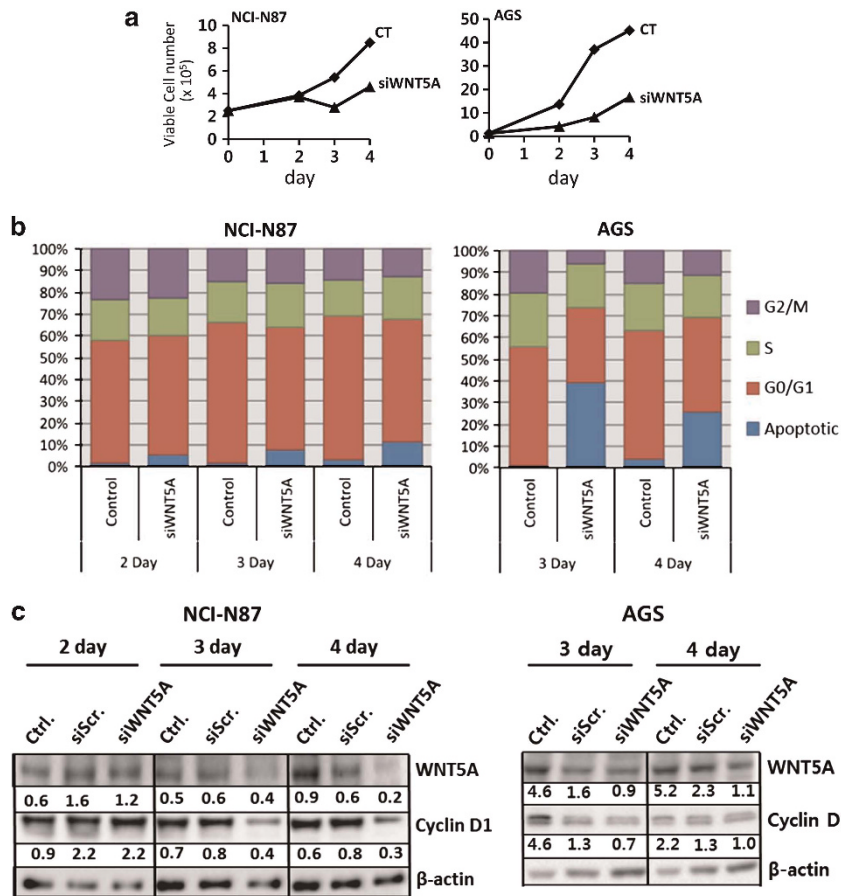
**Figure 4.** WNT5A is a downstream target of HNF4 $\alpha$  in gastric cancer (GC). **(a)** Immunoblotting detection of basal protein levels of WNT family members and downstream genes of WNT pathway in GC cell lines. **(b)** Immunohistochemistry measurement of WNT5A in GC cell line xenograft mouse models (+1 = low; +2 = mid and +3 = high). **(c)** Decreased gene expression level of HNF4 $\alpha$ , WNT5A and TCF4, RT-PCR measurement on day 2 (white bar) and day 3 (black bar), in the NCI-N87 cell line silenced with HNF4 $\alpha$ . **(d)** Immunoblotting of HNF4 $\alpha$ , WNT5A and HIF1 $\alpha$  on days 2 and 3 in the NCI-N87 cell line silenced with HNF4 $\alpha$ . **(e)** Growth inhibition was observed in GC cell lines (NCI-N87 and MKN-45) upon metformin (square = non-treated (-), circle = metformin (+)). Cells were counted from the day of metformin treatment (day 0). \* $P < 0.05$ . Immunoblotting of WNT family members and downstream genes treated with metformin in NCI-N87 and AGS GC cell lines for 5 days, and detection of cell cycle regulators. Immunoblotting was quantified using ImageQuant software normalized to  $\beta$ -actin (Molecular Dynamics/GE Healthcare Biosciences). Control (Ctrl), non-targeting siRNA (siScr), Metformin (MET).

statistical significance. Statistical significance is a metric by which researchers can rank or sort the importance of connections within biological cascades (equivalent to subpathways in our study). Along that line, PATHOME considers all subpathways from KEGG pathways, and tests each subpathway that is associated with differential biological phenotypes (cancer vs normal tissue). Other known tools consider whole pathways; whereas we have designed PATHOME to inspect a greater part of the biological cascade, given a set of pathways. As in Figure 1a, popular tools perform a single test for the toy pathway, removing interconnectivities among biological components. In contrast, PATHOME considers five tests while assessing the interconnectivities for the pathway,

which indicates a more thorough search of the biological pathways.

By applying PATHOME to GC gene expression data sets, we demonstrated its advantages. First, based on a reference pathway set, PATHOME detected more known cancer pathways than the alternative tools we assessed (at a lower FDR). Second, focusing on the WNT pathway that was uniquely detected by PATHOME, we provided strong evidence for the involvement of this pathway in gastric tumorigenesis through perturbation experiments in both cell line and animal studies.

Combining the evidence from perturbation experiments in cell lines and animal models through metformin, our results revealed a



**Figure 5.** Antiproliferative activity in the gastric cancer (GC) cell lines by knockdown of WNT5A. (a) Growth inhibition was observed in NCI-N87 and AGS by knockdown of WNT5A using siRNA-WNT5A (square = siRNA control (CT), triangle = siWNT5A). Cells were counted from the day of siRNA treatment (day 0). (b) Two GC cell lines (NCI-N87 and AGS) knocked down using siWNT5A, and cells were fixed and analyzed for DNA content using propidium iodide and fluorescence-activated cell sorting analysis. siWNT5A cells were compared with control cells. The percentage of cells in each cell cycle was calculated. Apoptotic was observed in knockdown of WNT5A in both cell lines. (c) Inhibition of cyclin D1, showing antiproliferative activity in both NCI-N87 and AGS transfected with siWNT5A.

cross-link between the AMPK metabolic pathway and the WNT signaling pathway, with HNF4α-WNT5A regulation having a key interaction in this link. A study performed by Kato *et al.*<sup>26</sup> suggested an antitumor effect of metformin on cell-cycle regulation in MKN-1, -45 and -74 GC cell lines. We also report a potential metformin inhibitory effect in GC cells and xenograft models. WNT5A is expressed in a variety of human tumors, including those of the esophagus, stomach, pancreas, colon and rectum, breast, lung, prostate gland, endometrial uterus and embryo, as well as in melanoma, osteosarcoma, Ewing sarcoma, neuroblastoma, skin basal cell carcinoma, skin squamous cell carcinoma and leukemia.<sup>27</sup> Kurayoshi *et al.*<sup>28</sup> reported that the WNT5A protein is highly expressed in advanced stages of GC and that its expression correlates with a poor prognosis. In addition, HNF4α and WNT signaling pathways are active members of the same machinery that controls the transcription of differentially zoned HNF4α-dependent genes in the liver, pancreas and biliary tract.<sup>29</sup> Even though WNT5A and HNF4α have been independently reported to be associated with cancer,<sup>27</sup> the link between HNF4α and WNT5A, to our knowledge, was first suggested in GC. Combining the evidence from the computational and perturbation experiments in cell lines and animal models, we demonstrated the oncogenic activity of WNT5A in GC and suggested WNT5A as a potential therapeutic target. Further study is warranted regarding the development of metformin therapeutic options and WNT5A targeted therapies to address GC progression

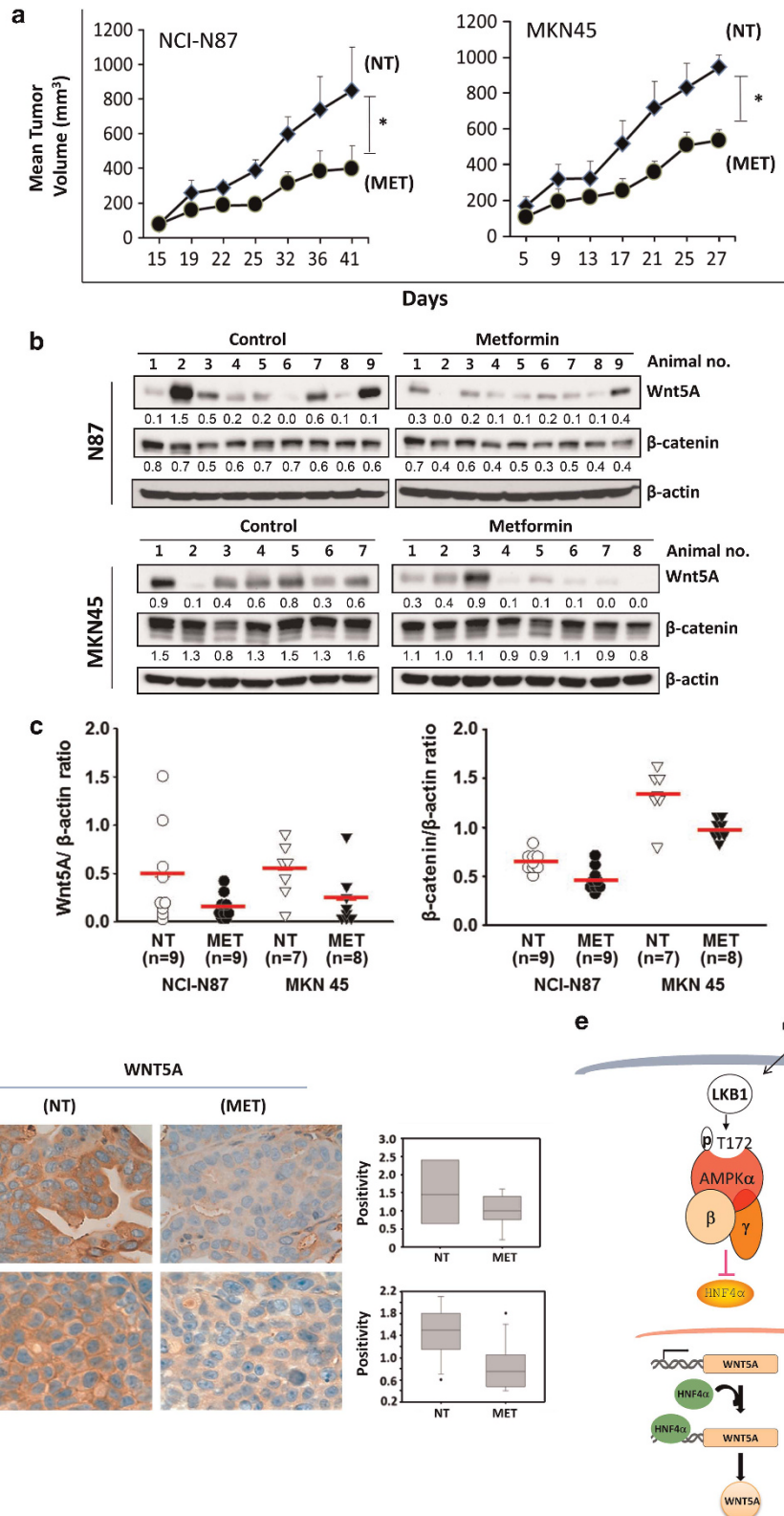
within Asian patient populations. In Figure 6e, we propose cross-talk between HNF4α/WNT pathway.

Despite successful application of PATHOME to GC, PATHOME solely analyzes the linear decomposition of pathways into paths. In the future, loops and loop-backs in the network need to be considered in statistical network analysis. We also note that as the HNF4α-WNT5A link is not annotated in the KEGG pathway database, PATHOME consequently does not pick up the HNF4α-WNT5A link.

## MATERIALS AND METHODS

### PATHOME algorithm

**Overview.** The goal of this algorithm is to identify a set of subpathways that differentiate two experimental groups (for example, cancer vs non-cancer) by considering both prior knowledge about mutual regulations and experimental gene expression data. Here, we used the KEGG pathways as prior knowledge, and assumed the KEGG pathways as a tree structure for the PATHOME application. For computational simplicity, as shown in Figure 1a, we broke down the KEGG pathway maps into each possible path (subpathway) from the top node to the leaf node.<sup>30</sup> A depth-first search algorithm was used to decompose the pathway maps into all possible paths. As our previous study indicated that the number of possible linear paths (subpathways) was huge (~0.13 billion possible paths),<sup>31</sup> we used a selection step before the statistical significance test step to reduce the number of tests (Figure 1).



**Figure 6.** Antitumor activity associated with WNT5A inhibition in gastric cancer (GC) using animal models. **(a)** Antitumor activity after metformin (MET) treatment on NCI-N87 and MKN-45 mouse xenograft models (in scid mice,  $n = 19$  per cell type). Cells were subcutaneously injected and grown for 15 days for the NCI-N87 model and 5 days for the MKN-45 model before treatment with MET. Tumor sizes were measured and compared between mice treated with or without MET, shown by the growth curve (diamond = NT (non-treated), circle = MET), \* $P$ -value  $< 0.05$ . **(b)** Immunoblotting of WNT5A and  $\beta$ -catenin tumors collected from NT and MET-treated mice. **(c)** Quantification of western blot of figure **(b)**. **(d)** Immunohistochemistry results showing WNT5A staining on xenograft models. The stain was scored as 0 to 3, weakly or strongly positive cells. **(e)** Proposed mechanism of action of AMPK/HNF4 $\alpha$ /WNT pathway, which can be utilized to target GC. Immunoblotting was quantified using ImageQuant software normalized to  $\beta$ -actin (Molecular Dynamics/GE Healthcare Bioscience).



**The selection step (selection of candidate subpathways).** As shown in Figure 1c, the rule is defined to associate the regulation information of adjacent entries with their expression correlation in terms of the sign information of the Pearson product-moment correlation coefficients. If two adjacent entries are connected by an edge that denotes activation (arrow-headed edge), the expression correlation between the two entries is assumed to be positive; if the two entries are connected by an edge that denotes inhibition (blunt-ended edge), the expression correlation between the two entries is assumed to be negative. This rule is applied separately to each experimental group. In each group, we identify the consecutive segment starting from the leaf node of each subpathway so that all the edges of the segment should satisfy the association rule. That leads to the determination of the segment (in the subpathway) that is to be statistically evaluated in the test step. Assuming a subpathway with the number of nodes (genes)  $p$ , the leaf node is assigned to index 1 ( $G_1$ ) and the root node to index  $p$  ( $G_p$ ). The two experimental groups are denoted as  $k$  ( $k=1$  or  $2$ ). If the edge ( $e_{i,i+1}$ ) between adjacent genes  $G_i$  and  $G_{i+1}$  is denoted as an activation edge in KEGG,  $e_{i,i+1}$  is encoded as +1. If the edge is represented as an inhibition edge in KEGG,  $e_{i,i+1}$  is set to -1. The Pearson correlation coefficient between the adjacent genes,  $G_i$  and  $G_{i+1}$ , is denoted as  $r_{i,i+1}^k$  in group  $k$  (for example,  $k=1$  for one experimental group;  $k=2$  for the other group). Thus, the association rule assumed in Figure 1c simply satisfies that  $e_{i,i+1} \times r_{i,i+1}^k$  is positive. The length of the segment ( $l^k$ ) of the  $k$ -th group in the subpathway under the rule is mathematically represented as follows:

$$l^k = \operatorname{argmin}_m \left\{ -\sum_{i=1}^m l(\operatorname{sgn}(r_{i,i+1}^k \times e_{i,i+1}) = 1) + \sum_{i=1}^m R(\operatorname{sgn}(r_{i,i+1}^k \times e_{i,i+1})) \right\} + 1, m \in \{1, \dots, p-1\}; R(x) = \begin{cases} 0, & \text{if } x \in \{1\} \\ \infty, & \text{otherwise} \end{cases}$$

where  $\operatorname{sgn}(\cdot)$  is the sign function, and  $l(\cdot)$  is the indicator function. The first term  $-l(\cdot)$  inspects the satisfaction of the association rule, keeping  $l^k$  progressing toward upstream. Inspecting the association rule from the leaf node to the root node, the penalizing term  $R(\cdot)$  stops progressing when the association rule is not satisfied. Given the subpathway, we obtain the two consecutive segments from the two groups (see details about  $l^k$  in Supplementary Figure 2). If both lengths of the two segments are greater than three, we move to the test step for the subpathway.

PATHOME analyzes the interconnectivity between two adjacent nodes. The interconnectivity measure, the Pearson product-moment correlation coefficient, is obtained even in three samples in a group. PATHOME can be applied to a small number of samples, such as three samples in a group. Summarizing the first step, a candidate subpathway for the next step should satisfy the following two conditions: (i) the two experimental groups agree with the association rule between the expression correlation and the edge information for the adjacent entries along the path; and (ii) both consecutive segments for the two groups have at least four elements (three consecutive edges) in order to filter a subpathway with short segments.

**The test step (statistical significance test).** In this step, we identify the subpathway for which the correlation difference between the two consecutive segments for the two experimental groups is statistically significant. As all the consecutive correlation coefficients,  $r_{i,i+1}^k$ 's, in the two  $l$  consecutive segments meet their corresponding regulation information (Figure 1), we do not need to further consider the sign information of the Pearson correlation coefficient in this step. To improve the normality approximation, we transform the absolute value ( $|r_{i,i+1}^k|$ ) of the Pearson correlation coefficient into  $(0, \infty)$  by taking the Fisher transformation, as follows:

$$c_{i,i+1}^k = \frac{1}{2} \ln \frac{1 + |r_{i,i+1}^k|}{1 - |r_{i,i+1}^k|} \in [0, \infty).$$

If the lengths of the two segments are different, we set the minimum (say,  $l^{\min}$ ) of  $l^k$ 's as the length of the two segments to be compared. As a result, we obtain  $\{c_{i,i+1}^k\}$   $k=1$ , and  $i=1, \dots, l^{\min}-1$  and  $\{c_{i,i+1}^k\}$   $k=2$ , and  $i=1, \dots, l^{\min}-1$  in the given subpathway. Let  $\mu^k$  represent the mean of  $c_{i,i+1}^k$ 's in group  $k$ . We test the significance under the null hypothesis:  $H_0: \mu^1 = \mu^2$ . That is, the alternative hypothesis suggests that the global mean of the correlations of the expressions between the two groups are different. We used a z-test statistic to measure significance. We also considered multiple comparisons, and set the FDR at 0.05.<sup>32</sup>

## Computational analysis of GC gene expression data sets

We obtained four GC data sets<sup>14-17</sup> from Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>). We compared our method with two popular GSA methods, DAVID<sup>3</sup> and GSEA.<sup>4</sup> For performance comparison, we used the cancer-related pathways proposed by Vogelstein and Kinzler<sup>19</sup> as a reference standard. We used the three methods (PATHOME, DAVID and GSEA) to examine the agreement between the reference standard and the reported pathways. To examine the TFBSs in the genes of interest, we extracted a 2-kb upstream region of each gene from the UCSC Genome Browser, and obtained the HNF4A-binding site information from the TFBS Conserved track and the Txn Factor ChIP track in the UCSC Genome Browser.

## Biological experiments

**Cells and reagents.** Human NCI-N87 and AGS derived from primary tumors GC cells were obtained from the American Type Culture Collection (ATCC; <http://www.atcc.org/>); and the MKN-1 and MKN-45 cell lines derived from primary tumor and tumor site of liver metastasis, respectively, were made available from Yonsei Cancer Center. SNU-1, -484 and -719 cell lines derived from primary tumor, SNU-5, -16, -620, -601, -638, -668 cell lines derived from ascites and SNU-216 cell line derived from tumor site of lymph node metastasis were made available from the Korean Cell Line Bank ([http://cellbank\\_snu.ac.kr/](http://cellbank_snu.ac.kr/)). Monoclonal antibodies to human HIF-1 $\alpha$  and  $\beta$ -catenin were purchased from BD Transduction Laboratories (BD Biosciences, San Jose, CA, USA); HNF4 $\alpha$ , WNT1, WNT3, GSK, TCF, LEF, pCAMKII (Thr 286), cyclin D1, cyclin D2 and  $\beta$ -actin were purchased from Cell Signaling Tech. Inc. (Boston, MA, USA). WNT5A was purchased from Abgent (San Diego, CA, USA) and metformin was purchased from Sigma-Aldrich (St Louis, MO, USA).

**Cell culture.** Human NCI-N87, AGS, HS 746T, MKN-1, 45, SNU-1 -5, -16, -620, -216, -484, -601, -638, -668, -719 GC cell line studies were analyzed within 6 months of tissue resuscitation; the tissues were cultured in RPMI-1640 (CellGro, Manassas, VA, USA) and 10% fetal calf serum (FCS; Hyclone, ThermoScientific, Waltham, MA, USA) at 37°C in 5% CO<sub>2</sub>. ATCC used short tandem repeat profiling. Cells ( $2.5 \times 10^5$ ) were seeded and incubated under normoxic conditions to 70–80% confluence and then incubated in the presence or absence of metformin at 10 mM concentration for up to 5 days according to the required time in the study.

**Western blotting.** Cells were grown under hypoxic conditions in the presence or absence of 10 mM metformin. The cells were washed twice in a phosphate-buffered saline solution and western blotting was conducted, as previously described.<sup>33</sup> The tumors collected from xenograft mouse models were divided into two pieces for immunoblotting and immunohistochemistry, as previously described.<sup>33</sup> The blots were quantified using ImageQuant software (Molecular Dynamics/GE Healthcare Biosciences, Sunnyvale, CA, USA).

**Real-time RT-PCR analysis.** Total RNA was isolated from cell lysates using the PARIS kit (Ambion/Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocol. Next, TaqMan quantitative RT-PCR was performed on the ABI 7300 system using the TaqMan one-step RT-PCR Master Mix kit and predesigned primer/probe pairs for HNF4 $\alpha$ , WNT5A, TCF4 and  $\beta_2$ -microglobulin (Applied Biosystems). Normalization procedures and analyses were carried out with  $\beta_2$ -microglobulin using the 2(-delta-delta C(T)) method as the internal reference,<sup>34</sup> and using Applied Biosystems GeneAmp 5700 SDS software. All measurements were performed in triplicate.

**siRNA transfection.** siRNA SMARTpool sequences were obtained from Dharmacon/Thermo Fisher Scientific (Waltham, MA, USA); and the cells were transfected with 25 nM siRNA-HNF4 $\alpha$ , siRNA-WNT5A and a siRNA nontargeting control using Dharma-FECT 1 lipid transfection reagent. The transfection medium was removed after 24 h and replaced with fresh medium, and the cells were grown in 5% CO<sub>2</sub> at 37°C for an additional 48–72 h. RT-PCR and/or western blot analyses were performed to confirm target knockdown by siRNA. The transfected cells were treated with metformin and cultured under hypoxic conditions for an additional 18 h.

**Immunohistochemical staining.** Hematoxylin and eosin-stained slides were reviewed and representative areas were selected for tissue microarray. Normal mucosa and cancer tissues were selected, respectively.

In the experiment, 2-mm diameter cores were taken from archival paraffin-embedded blocks using a trephine apparatus (Superbiochips Laboratories, Seoul, Republic of Korea). Tissue microarray blocks were sectioned at a thickness of 3  $\mu\text{m}$ , and these sections were then dried for 1 h at 56 °C. Immunohistochemical staining was performed with the automated staining instrument BenchMark XT (Ventana Medical Systems, Inc., Tucson AZ, USA) as follows: the sections were deparaffinized and rehydrated with EZ Prep (Ventana Medical Systems, Inc.) and washed with Tris-buffered saline. The antigens were retrieved with heat treatment in pH 8.0 Tris-EDTA buffer (CC1, Ventana Medical Systems, Inc.) at 95 °C for 30 min for WNT5A. Endogenous peroxidases were blocked with 3%  $\text{H}_2\text{O}_2$  for 10 min at room temperature. Nonspecific-binding blocking was done with a ready-to-use protein blocker solution (Ventana Medical Systems, Inc.) for 20 min at RT. A primary antibody was applied to the slide section at 42 °C (WNT5A for 30 min at 1:5000 dilutions (clone 6F2, Abgent, AO1264a)). Then the sections were incubated with HRP multimer labeled secondary antibody (ultraView Universal DAB detection kit, Ventana Medical Systems, Inc.) for 20 min at RT and stained using ultraView universal DAB kit (Ventana Medical Systems, Inc.) for 8 min and hematoxylin counterstain.

**Interpretation of immunohistochemistry.** We defined the intensity WNT5A stain as follows: if no signal or only a faint equivocal signal was observed at  $\times 100$  power, it was regarded as negative, 0; if more than 10% of tumor cells showed clear nuclear signals at  $\times 100$  power, similar to those of the foveolar epithelial cells, it was weakly positive, 1; if more than 10% of tumor cells showed obviously stronger signals, at  $\times 40$  power, similar to those of the chief cells for WNT5A, it was strongly positive, 2. The positivity of the WNT5A stain was measured by the percentage of weakly or strongly positive tumor cells. The stain was scored as 0 (negative) if weakly or strongly positive cells were  $\leq 50\%$ ; and as 1 (positive) if they were  $>50\%$ . Representative results of WNT5A immunostaining for GC are shown in Figure 5c.

**In vivo antitumor study.** Approximately  $10^7$  NCI-N87 and MKN-45 GC cells in log cell growth were injected and subcutaneously suspended in 0.2 ml phosphate-buffered saline in the flanks of severe combined immunodeficient (scid) mice. The animals were weighed weekly and tumor diameters were measured twice weekly at right angles ( $d_{\text{short}}$  and  $d_{\text{long}}$ ) with electronic calipers and converted to volume by the formula  $\text{volume} = [(d_{\text{short}})^2 \times (d_{\text{long}})]/2$ . When the tumors reached volumes between 150 and 300  $\text{mm}^3$ , the mice were stratified into groups of 10 animals, total 40 animals having approximately equal mean tumor volumes, and the administration of metformin commenced at 250 mg/kg per oral daily for 25 days. Control animals received vehicle (water) alone. Twenty-four hours after the last metformin administration commenced, tumors were collected for western blot analysis and for immunohistochemistry preparation. When the tumor volume reached  $\geq 1500 \text{mm}^3$  or became necrotic, the animals were euthanized.

**Statistical analysis.** For the experiments on individual genes in this section,  $P < 0.05$  was statistically significant based on a Student's *t*-test or *z*-test to compare the experimental group with the control group.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This study was supported by grants from the National Cancer Center of the Republic of Korea, NCC-1210350-2 (to YHK), NCC-1210460 (to SN); the National Research Foundation of Korea (NRF) grant funded by the Republic of Korea (MSIP) 2012R1A3A2026438, 2008-0062618, 2013M3A9C4078158 (to TP); grants from the US National Institutes of Health, CA95060, CA129616, CA17094 and CA98920 (to GP), CA143883 (to HL); the University of Texas MD Anderson Cancer Center-GS Hogan Gastrointestinal Research Fund and the Lorraine Dell Program in Bioinformatics for Personalization of Cancer Medicine (to HL). The Korea Institute of Science and Technology Information provided SN with a high-performance computing resource. We thank LeeAnn Chastain for editorial assistance.

## AUTHOR CONTRIBUTIONS

HL, TP and YHK conceived and designed the experiments; SN and TP designed the algorithm; SN, DH, HL and YHK analyzed the data; HRC, K-TK, M-CK, CHK, HRJ, HSP, GP, YHK performed the experiments; SN, HL, TP and YHK wrote the paper. All authors have read and approved the manuscript for publication.

## REFERENCES

- Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005; **309**: 1728–1732.
- Soon WW, Hariharan M, Snyder MP. High-throughput sequencing for biology and medicine. *Mol Syst Biol* 2013; **9**: 640.
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**: 44–57.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; **102**: 15545–15550.
- Kanehisa M. A database for post-genome analysis. *Trends Genet* 1997; **13**: 375–376.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**: 25–29.
- Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 2003; **13**: 2129–2141.
- Karp PD, Riley M, Paley SM, Pellegrini-Toole A. The MetaCyc database. *Nucleic Acids Res* 2002; **30**: 59–61.
- Huerta AM, Salgado H, Thieffry D, Collado-Vides J. RegulonDB: a database on transcriptional regulation in *Escherichia coli*. *Nucleic Acids Res* 1998; **26**: 55–59.
- Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011; **12**: 56–68.
- Nam D, Kim SY. Gene-set approach for expression pattern analysis. *Brief Bioinform* 2008; **9**: 189–197.
- Khatiri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* 2012; **8**: e1002375.
- Tarca AL, Draghici S, Khatiri P, Hassan SS, Mittal P, Kim JS *et al.* A novel signaling pathway impact analysis. *Bioinformatics* 2009; **25**: 75–82.
- Cho JY, Lim JY, Cheong JH, Park YY, Yoon SL, Kim SM *et al.* Gene expression signature-based prognostic risk score in gastric cancer. *Clin Cancer Res* 2011; **17**: 1850–1857.
- Takeo A, Takemasa I, Seno S, Yamasaki M, Motoori M, Miyata H *et al.* Gene expression profile prospectively predicts peritoneal relapse after curative surgery of gastric cancer. *Ann Surg Oncol* 2010; **17**: 1033–1042.
- Kim YH, Liang H, Liu X, Lee JS, Cho JY, Cheong JH *et al.* AMPK $\alpha$  modulation in cancer progression: multilayer integrative analysis of the whole transcriptome in Asian gastric cancer. *Cancer Res* 2012; **72**: 2512–2521.
- Cui J, Chen Y, Chou WC, Sun L, Chen L, Suo J *et al.* An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer. *Nucleic Acids Res* 2011; **39**: 1197–1207.
- Parkinson H, Sarkans U, Kolesnikov N, Abeygunawardena N, Burdett T, Dylag M *et al.* ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res* 2011; **39**: D1002–D1004.
- Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nature Med* 2004; **10**: 789–799.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**: 2498–2504.
- Euskirchen GM, Rozowsky JS, Wei CL, Lee WH, Zhang ZD, Hartman S *et al.* Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. *Genome Res* 2007; **17**: 898–909.
- Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* 2011; **39**: D876–D882.
- Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M *et al.* GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* 2003; **4**: R28.
- Khatiri P, Bhavsar P, Bawa G, Draghici S. Onto-Tools: an ensemble of web-accessible, ontology-based tools for the functional design and interpretation of high-throughput gene expression experiments. *Nucleic Acids Res* 2004; **32**: W449–W456.

- 25 Backes C, Keller A, Kuentzer J, Kneissl B, Comtesse N, Elnakady YA *et al*. GeneTrail—advanced gene set enrichment analysis. *Nucleic Acids Res* 2007; **35**: W186–W192.
- 26 Kato K, Gong J, Iwama H, Kitanaka A, Tani J, Miyoshi H *et al*. The antidiabetic drug metformin inhibits gastric cancer cell proliferation *in vitro* and *in vivo*. *Mol Cancer Ther* 2012; **11**: 549–560.
- 27 Katoh M. WNT signaling pathway and stem cell signaling network. *Clin Cancer Res* 2007; **13**: 4042–4045.
- 28 Kurayoshi M, Oue N, Yamamoto H, Kishida M, Inoue A, Asahara T *et al*. Expression of Wnt-5a is correlated with aggressiveness of gastric cancer by stimulating cell migration and invasion. *Cancer Res* 2006; **66**: 10439–10448.
- 29 Colletti M, Cicchini C, Conigliaro A, Amicone L, Tripodi M. Convergence of Wnt signaling on the HNF4alpha-driven transcription in controlling liver zonation. *Gastroenterol* 2009; **137**: 660–672 Erratum in *Gastroenterol* 2010; **138**: 793.
- 30 Chen X, Xu J, Huang B, Li J, Wu X, Ma L *et al*. A sub-pathway-based approach for identifying drug response principal network. *Bioinformatics* 2011; **27**: 649–654.
- 31 Nam S, Park T. Pathway-based evaluation in early onset colorectal cancer suggests focal adhesion and immunosuppression along with epithelial-mesenchymal transition. *PLoS One* 2012; **7**: e31685.
- 32 Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J Roy Stat Soc Series B Stat Methodol* 1995; **57**: 289–300.
- 33 Kim YH, Coon A, Baker AF, Powis G. Antitumor agent PX-12 inhibits HIF-1alpha protein levels through an Nrf2/PMF-1-mediated increase in spermidine/spermine acetyl transferase. *Cancer Chemother Pharmacol* 2011; **68**: 405–413.
- 34 Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C (T) method. *Nat Protoc* 2008; **3**: 1101–1108.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary Information accompanies this paper on the Oncogene website (<http://www.nature.com/onc>)