Gene expression

Deconvoluting essential gene signatures for cancer growth from genomic expression in compound-treated cells

Jinmyung Jung1,2, Yeeok Kang3, Hyojung Paik4, Mijin Kwon3, Hasun Yu1 and Doheon Lee1,3,*

1Bio-Synergy Research Center, 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea, 2Department of Data Science, College of Information Technology, The University of Suwon, Bongdam-eup, Hwaseong, 18323, Republic of Korea, 3Department of Bio and Brain Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea and 4Korea Institute of Science and Technology Information, Center for Applied Scientific Computing, Division of Supercomputing, Daejeon 34141, Republic of Korea

*To whom correspondence should be addressed.

Associate Editor: Janet Kelso
Received on January 3, 2018; revised on August 10, 2018; editorial decision on August 29, 2018; accepted on August 31, 2018

Abstract

Motivation: Essential gene signatures for cancer growth have been typically identified via RNAi or CRISPR–Cas9. Here, we propose an alternative method that reveals the essential gene signatures by analysing genomic expression profiles in compound-treated cells. With a large amount of the existing compound-induced data, essential gene signatures at genomic scale are efficiently characterized without technical challenges in the previous techniques.

Results: An essential gene is characterized as a gene presenting positive correlation between its down-regulation and cell growth inhibition induced by diverse compounds, which were collected from LINCS and CGP. Among 12 741 genes, 1092, 1 228 827 962, 1 664 580 and 829 essential genes are characterized for each of A375, A549, BT20, LNCAP, MCF7, MDAMB231 and PC3 cell lines (\(P\)-value < 1.0E–05). Comparisons to the previously identified essential genes yield significant overlaps in A375 and A549 (\(P\)-value < 5.0E–05) and the 103 common essential genes are enriched in crucial processes for cancer growth. In most comparisons in A375, MCF7, BT20 and A549, the characterized essential genes yield more essential characteristics than those of the previous techniques, i.e. high gene expression, high degrees of protein–protein interactions, many homologs and few paralogs. Remarkably, the essential genes commonly characterized by both the previous and proposed techniques show more significant essential characteristics than those solely relied on the previous techniques. We expect that this work provides new aspects in essential gene signatures.

Contact: dhlee@kaist.ac.kr or jmjung@suwon.ac.kr
Supplementary information: Supplementary data are available at Bioinformatics online.
1 Introduction

Essential genes in cancer are usually defined as genes required (or essential) for cancer growth. Identification of the essential genes is crucial for cancer research, particularly in the field of cancer therapeutics (Cheung et al., 2011; Marcotte et al., 2012). According to the definition of essential gene, an essential gene is regarded as the gene whose down-regulation induces substantial cell growth inhibition. Heretofore, gene knock-down with RNAsi (RNA interference) (Hannon, 2002) is one of the predominant tools for the identification. In this technique, a shRNA (short hairpin RNA) cleaves and down-regulates the mRNA complementary to the shRNA. Recent RNAi technologies, such as lentiviral shRNA pooled screenings have enabled genome-scale analysis and yielded novel essential gene signature in cancer (Gao et al., 2014; Luo et al., 2008; Marcotte et al., 2012). Gene knock-out with CRISPR (clustered regularly interspaced short palindromic repeat)–Cas9 is another great tool for loss-of-function analysis to identify the essential signatures for cancer growth (Shalem et al., 2015). A complex of Cas9 and sgRNAs (single-guide RNAs) causes DNA cleavage of the target gene matched to the sgRNAs, creating frame shift mutations that finally result in the knock-out effect of the gene (Shalem et al., 2014). Recently, a pooled screening with the CRISPR-Cas9 system has allowed genome-wide knock-out experiments (Shalem et al., 2014; Wang et al., 2014). Basically, in both techniques, significant cell growth inhibition with shRNA or sgRNA indicates an essential signature of the target gene.

Here, we propose an alternative method that deconvolutes essential gene signatures for cancer growth from genomic expression in compound-treated cells. This method takes advantage of genomic expressions and growth inhibition profiles in compound-treated cells, largely accumulated so far. For instance, enormous compound-induced expression profiles perturbed by ~30,000 compounds have been archived in LINCS (The Library of Integrated Network-Based Cellular Signatures) database (Subramanian et al., 2017). In addition, CGP (Cancer Genome Project) provides cell growth inhibition information with dose response curves of 132 compounds (Garnett et al., 2012). In this study, with a large amount of the existing compound-induced data, essential gene signatures at the genomic scale are efficiently analysed without technically challenging processes involved in the previous techniques, such as genome-scale shRNA and sgRNA (single-guide RNA) design, their packaging in virus and isolating the shRNA and sgRNA sequences from cultured cells (Shalem et al., 2015).

Whereas only a single gene is down-regulated in the previous techniques using shRNA or sgRNA, numerous gene expressions are usually down-regulated in a compound treatment. From many target proteins of a compound, biological signals are transferred to numerous gene expressions via signalling pathways (such as protein–protein interactions and gene regulation), which causes their perturbations such as up- or down- regulation. Here, we propose a method that extracts essential gene signatures among numerous down-regulated genes by means of associating their cell growth inhibition in compound-treated cells.

In this study, we simultaneously examine the genomic expression profiles and cell growth inhibition in compound-treated cells to characterize the essential signature for cancer growth. 1092, 1228, 827, 962, 1664, 580 and 829 essential genes are characterized for each of A375, A549, BT20, LNCAP, MCF7, MDAMB231 and PC3 cell lines (P-value ≤ 1.0E–05). The essential genes are compared to those of previous techniques, and the enriched functional terms of the common essential genes are identified. Furthermore, they are assessed in terms of the four essential characteristics, i.e. high gene expression, high degrees in protein–protein interactions, many homologs and few paralogs. In the Discussion Section, we describe other possible applications of the proposed method, such as characterization of essential genes involving cancer death and inference of compound MOAs (mode of actions).

2 Materials and methods

2.1 Strategy overview

As mentioned in introduction, an essential gene is believed as the gene whose down-regulation induces substantial cell growth inhibition. In this study, an essential gene is defined as a gene showing positive correlation between its down-regulation and cell growth inhibition induced by diverse compounds. For the purpose, we integrated two kinds of data in compound-treated cells, i.e. compound-induced gene expression profiles from LINCS and compound-induced cell growth inhibition from CGP, for compounds common to the two databases. Then, for each gene, we measure how much positive correlation existed between its down-regulations and cell growth inhibitions (Fig. 1).

2.2 Data preparation

2.2.1 LINCS database

LINCS database provides gene expression profiles with ~30,000 compound experiments from 21 cell lines (Subramanian et al., 2017).

![Compound experiment data](image)

**Fig. 1.** Overall strategy. For each cell line, an integrated table is generated that consists of compound-induced expression and cell inhibition obtained from LINCS and CGP, respectively. Then, an essentiality score is assigned to each gene that measures how much positive correlation existed between its down-regulations and cell growth inhibitions. These processes are identically applied to the seven cell lines, i.e. MCF7, BT20, MDAMB231, LNCAP, PC3, A375 and A549.
2.1 Treatment of compound q with dose r, i.e. $\text{FC}(G_k, C_q, D_r)$, as below:

$$\text{FC}(G_k, C_q, D_r) = \log_2 \left( \frac{\sum_{i=1}^{n} \text{expr}(G_k, C_q, D_r)}{\sum_{j=1}^{m} \text{expr}(G_k, V_j)} \right) - \log_2 \left( \frac{\sum_{i=1}^{n} \text{expr}(G_k, V_j)}{\sum_{j=1}^{m} \text{expr}(G_k, C_q, D_r)} \right),$$  

(1)

where $\text{expr}(G_k, C_q, D_r)^i$ is an expression of gene $k$ acquired by $i$th treatment of compound $q$ with dose $r$, and $\text{expr}(G_k, V_j)$ is an expression of gene $k$ acquired by $j$th treatment of vehicle. The calculated FCs were averaged across plates in terms of compound ID and its dose. This process finally resulted in the 21-fold change tables consisting of compound ID, dose and fold changes (Fig. 2a).

2.2.2 CGP database

CGP database contains cell growth inhibition data of 132 compounds for each of the 639 cell lines, which were measured at 72 h (Garnett et al., 2012). From the database, the parameters of the IC50 (i.e. a dose producing 50% growth inhibition) and beta (i.e. slope of the dose response curve) of 132 compounds were obtained for each of the 639 cell lines, which allows for a cell line specific dose response curve (Fig. 2d). This process finally resulted in the 639 cell growth inhibition tables consisting of compound ID, IC50 and beta (Fig. 2b).

2.2.3 Data integration

We integrated the expression fold changes and cell growth inhibition data for the nine cell lines common to the two datasets, i.e. MCF7, BT20, HS578T, MDAMB231, LNCAP, PC3, A375, A549 and HT29 cell lines. The point of this integration is to compute cell growth inhibition at the dose where each compound was treated in LINCS for measuring expression profiles. It is facilitated by the IC50 and beta provided from the CGP data, which determines compound-specific dose response curves for each cell line (Fig. 2d).

Here, PubChem CIDs were used to pair compounds in the two databases. For a given IC50 and beta of a specific compound, a dose returns the corresponding cell growth inhibition (GI) by a formula as below:

$$\text{GI}(\text{dose}) = \frac{100}{1 + e^{(\text{IC50} - \text{dose})/\text{beta}}}$$  

(2)

By this process, for each of the nine cell lines, we generated an integrated table consisting of compound IDs, doses, expression profiles and cell growth inhibitions. A row of the table is a compound treatment with a particular dose that has induced expression fold changes (FCs) and a cell growth inhibition (GI) (Fig. 2c). Here, the HT29 and HS578T cell lines were excluded from further analysis due to the lack of experiments (less than 20 experiments). Finally, there remains the seven integrated tables for the seven cell lines, i.e. three breast cancer (MCF7, BT20, MDAMB231), two prostate cancer (LNCAP, PC3), one skin cancer (A375), one lung cancer (A549) cell lines (Fig. 2c). The number of compound experiments in the remaining seven cell lines ranged from 78 (PC3 cell line) to 174 (MCF7 cell line) (Supplementary Fig. S1). FCs across the seven cell lines in the integrated tables were normally distributed; however, their GIs were highly skewed towards low GI (Supplementary Fig. S2a and S2b).

2.3 Essentiality score

To determine essential genes required (or essential) for cancer growth, we characterize genes whose down-regulation yielded cell growth inhibition in compound-treated cells. To this end, an essentiality score of a $k$-th gene, i.e. $\text{ES}(G_k)$, is defined as below:

Take $k$ – th gene as $G_k$

Define

$$\text{FC}(G_k) = \text{FC} \text{ of } G_k \text{ in } i \text{ - th compound experiment}$$

$$\text{GI}_i = \text{GI} \text{ in } i \text{ - th compound experiment}$$

$$\text{exp}(G_k) = \{\text{FC}(G_k), \text{GI} \}$$

$$\text{negEXP}(G_k) = \{\text{exp}(G_k) : \text{FC}(G_k) < 0 \}$$

$$\text{negGI}(G_k) = \{\text{GI} : \text{FC}(G_k) < 0 \}$$

$$\text{ES}(G_k) = \text{Spearman(} \text{negEXP}(G_k) \}) \times \text{mean(} \text{negGI}(G_k) \})$$

An ES of a gene is computed with a correlation between its FCs and GIs (Fig. 3). We here employ Spearman (not Pearson) correlation, which is a non-parametric measure of rank correlation,
An ES is a multiplication of the correlation by the mean of their GIs in order to give more weight when more substantial cell growth has been inhibited. In calculation of an ES, only compound experiments (exp) are considered according to the definition of an essential gene [in Equation (3)], which indicates that a random distribution is negEXP-size specific as well. It turns out that the smaller size of negEXP produces a random distribution composed of higher negative ESs. The empirical P-value for a given ES is computed as a rank of the ES in the corresponding distribution. For example, AARSD1 gene in A375 cell line was down-regulated in 43 among 90 compound experiments. Its essentiality score is computed as –14.292, and Its empirical P-value is 0.000087 [Supplementary Table S1]. It means that –14.292 is ranked at 87th in a random distribution generated with the 43 size of negEXP in A375 cell line.

3 Results and evaluations
3.1 Significant essential genes and their enriched functional terms
The main results containing ESs and empirical P-values of the 12 716 genes of the considered seven cell lines are described in Supplementary Table S1, and their averaged P-values are presented in Supplementary Table S2. The distribution of ESs across the seven cell lines were slightly skewed-normally distributed towards low scores [Supplementary Fig. S2c]. For better understanding, we depicted the FC-GI plot of the two top essential genes, i.e. EIF4EBP1 in A375 and NACA2G2 in A549, in Figure 4. As observed in the figure, down-regulation of each gene apparently induced a large amount of cell growth inhibition.

Essential genes (EGs) in each cell line are determined as genes whose empirical p-value is ≤ 0.00001 (1.0E–05). Among 12 741 genes, 1092, 1228, 827, 962, 1664, 580 and 829 essential genes are characterized for each of A375, A549, BT20, LNCAP, MCF7, MDAMB231 and PC3 cell lines. Common essential genes (CEGs) are determined as genes whose averaged empirical P-value across the seven cell line is ≤ 0.00001, and 103 CEGs are characterized. Their statistics and a list of CEGs are displayed in Supplementary Table S3. To examine the enriched biological functions of the CEGs, functional enrichment tests were applied to the CEGs with the gene sets of biological processes (Gene Ontology) and KEGG pathways [Ashburner et al., 2000; Kanehisa et al., 2012]. As a result, they were enriched in DNA replication, DNA repair and many other cell cycle involved processes, yielding that they are closely related to the crucial processes for cancer growth and survival [Supplementary Fig. S3 and Supplementary Table S4].

3.2 Comparison to the essential genes in previous work
To examine how many essential genes characterized by the previous techniques are commonly covered in this study, we collected the essential genes identified by shRNA and CRISPR–Cas9 in available four cell lines. Genes of 18 904 in the A375 cell line were examined by the shRNA and they were ranked by the ∗shRNA depletion scores [Cowley et al., 2014; Luo et al., 2008]. For MCF7, BT20 and A549 cell lines, the top 10% genes with more negative shRNA depletion scores are considered as essential genes. In the proposed method, 12 716 genes were covered in the all four cell lines, and the randomly sampled GIs and negative FCs from the integrated table of the corresponding cell line, which indicates that a random distribution is cell line specific. Moreover, a random distribution was generated for each number of compound experiments that induced down-regulation of a gene, i.e. the size of negEXP [in Equation (3)], which indicates that a random distribution is negEXP-size specific as well.
EGs of each cell line (determined in Section 3.1) are considered as essential genes (Table 1).

For each of the four cell lines, hypergeometric tests were applied to the essential genes identified by the previous and proposed techniques among the genes commonly considered in the both techniques. As a result, the two of four comparisons in A375, A549 produced significant results (P value < 0.05). Especially, extremely high significant results (P value < 5.0E-35) were observed in the A375 cell line characterized by CRISPR, which is a more recent and advanced technique than shRNA. (Fig. 5 and Supplementary Table S5 for more detail information).

For an explanation of less significant results than expected in some comparisons, hypergeometric tests were also applied to the two sets of essential genes identified by the shRNA. The comparison in A549 produced a significant result (P value < 0.005); however, that of MCF7 did not show a significant result (P value > 0.1) (Supplementary Fig. S4). There might be imperfections in the previous techniques in identifying essential genes, which can be an explanation for the less significant results from a few comparisons.

### 3.3 Assessment of essential gene characteristics

Similar to Wang’s work evaluating essential genes in terms of several characteristics of essentiality (Wang et al., 2015), we assessed the four characteristics of essential genes, i.e. (i) higher gene expression, (ii) higher degrees of protein–protein interactions, (iii) more homologs, indicating higher levels of conservation among species and (iv) less paralogs, indicating less functional redundancy. Expression, protein–protein interaction (PPI), homologue and paralog data were acquired from Cancer Cell Line Encyclopedia (CCLE) (Barrettina et al., 2012), BioGIRD (Chatr-Aryamontri et al., 2015), Homologene in NCBI (Ncbi, 2017) and Duplicated Genes Database (Ouedraogo et al., 2012), respectively. Here, non-cell-specific data were used in assessing all characteristics, except for expression characteristics that utilized cell specific expression in CCLE.

For each characteristic, the corresponding values were averaged and compared for each of the four gene sets, i.e. all considered genes in the corresponding characteristic (ALL), essential genes by the previous technique (PREV), essential genes by the proposed method (PROP) and common essential genes between PREV and PROP (BOTH). Here, PREV are specified in Table 1, and PROP are the EGs determined in Section 3.1. The comparisons were carried out for each cell line. In addition to the comparisons of their averages, we applied Welch’s t-test between PREV and PROP and between PREV and BOTH to statistically examine the difference of the characteristic values. The assessment between PREV and BOTH is to determine how much improvement was achieved from the essential genes simultaneously characterized by both techniques compared to the essential genes solely relying on the previous technique.

In assessments between ALL and PROP, PROP represent more essential characteristics than ALL in all the 28 average comparisons, indicating that the essential genes characterized by the proposed technique indeed have more essential characteristics than the rest of the genes (Fig. 6). In assessments between PREV and PROP, PROP show more essential characteristics than PREV in 15 among 16 assessments (except for the homolog evaluation in A375) compared to their averages and produce significant results (P value < 0.01) in the 7 among 16 assessments at Welch’s t-tests (Fig. 6). More noticeably, in assessments between PREV and BOTH, BOTH represented more essential characteristics than PREV in all 16 assessments compared to their averages and show significant results (P value < 0.01) in the 11 among 16 assessments at Welch’s t-tests (Fig. 6). This demonstrated that essential genes that are associated with more essential characteristics can be determined by both previous and proposed techniques than the previous technique only.

### Table 1. Essential genes by previous techniques (shRNA and CRISPR) and the proposed method

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Essential genes by shRNA or CRISPR techniques</th>
<th>Essential genes by the proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>A375</td>
<td>CRISPR</td>
<td>12716</td>
</tr>
<tr>
<td>MCF7</td>
<td>shRNA Cowley 2014</td>
<td>12716</td>
</tr>
<tr>
<td>BT20</td>
<td>shRNA Cowley 2014</td>
<td>12716</td>
</tr>
<tr>
<td>A549</td>
<td>shRNA Luo 2008</td>
<td>12716</td>
</tr>
</tbody>
</table>

**Note:** Genes of 988 are explicitly determined as essential signatures in the A375 cell line by the CRISPR. For MCF7, BT20 and A549 cell lines, the shRNA techniques were employed for essential gene identification, and the top 10% genes with more negative *shRNA depletion scores are considered as essential genes. The EGs of each of the four cell lines are considered as essential genes by the proposed method. * shRNA depletion score: the amount of the underrepresented shRNAs in surviving cells in pooled shRNA experiments. The more negative shRNA depletion score indicates more essentiality for cancer growth.
and PC3 cell lines. Their essential scores and 0.5 for each of A375, A549, BT20, LNCAP, MCF7, MDAMB231 224, 74 and 108 essential genes among 978 genes (Cheng and Li, 2016) showed that the quality of imputed expression genes were considered; however, a few studies (Chen et al., 2014). Out of the top five essential genes in Shalém’s work (RPS19, TCOF1, NIP7, CAR52 and RPL7), two genes (TCOF1 and NIP7) are characterized as essential genes by the proposed method and the other three genes are missing. From the FC-GI plots of the five genes (Supplementary Fig. S5), we noticed common characteristics of the three missing genes, which is that strong down-regulations were not induced by the compound experiments. The most negative fold changes for the two characterized genes (TCOF1 and NIP7) are -1.16 and -1.59; however, those for the three missing genes (RPS19, CAR52 and RPL7) are -0.29, -0.58 and -0.19, which are rarely down-regulated. Here, we noticed a limitation of this study, which is that genes unwilling to be down-regulated by compounds would not be characterized as essential genes. Some genes in a cell might be hardly perturbed by any compounds, and they cannot be identified by the proposed approach even though they are involved in essential functions.

We also observed another limitation due to the utilization of two resources from different research groups, i.e. LINCS and GCP. Experimental environment and protocol can be different between two groups, which would produce inaccurate essential genes. We believe that the proposed method allows more accurate identification of essential genes if the resources are more reliable.

There are a few other applications for the proposed technique in this study. First, not only essential genes for cancer growth but also cancer death can be easily determined without any further complicated data processing. Essential genes for cancer death can be defined as the genes whose up-regulation induces cell growth inhibition. Essential genes for cancer growth and death can be utilized for the identification of oncogenes (causing cancer with high activation) and tumour suppressor genes (causing cancer with loss of their function), respectively, which is a well-known classification of cancer genes in a genomic perspective. Even though only gene expression was examined without genomic mutation data, the identification of essential genes for cancer growth and death could be helpful for functional assessment of cancer genes. For example, two essential genes for cancer death, DNAJB2 in A375 and CD55 in MDAMB231 cell lines, were characterized and are depicted in Supplementary Figure S8. Here, a slightly different equation with Equation (3) has been applied (Supplementary Fig. S8). We plan to examine those genes in more detail in further work. For the same purpose of this application, complex genetic engineering (such as gene editing, gene insertion, or promoter design) would be required in traditional techniques.

Second, the results of this study allow to infer MOAs of the compounds incorporated in the experiments. For example, the treatment with 0.04 μM PD-0325901 induced a great amount of growth inhibition (85.8%), along with the substantial down-regulation of EIF4EBP1 (Fig. 4). Besides the EIF4EBP1 gene that is characterized as an essential gene, other 71 genes were considerably down-regulated (< -2.0-fold change) by the treatment (0.04 μM PD-0325901). If the other 71 genes are not characterized as essential genes, their down-regulations might be merely results (not causes) of cell inhibition or irrelevant to cell inhibition. Thus, in this case, we considered that the paths from the target of

![Fig. 5. Comparison to the essential genes from previous work. Hypergeometric tests were applied to the essential genes identified by the previous and proposed techniques to determine how many significant overlaps there are. It yielded highly significant results (P-value < 5.E-06) in A375 and A549 cell lines. P: P-value, S: the number of the shared genes. (More detail information are specified in Supplementary Table S5)](https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/bty774/5089233)

![Fig. 6. Assessments of essential characteristics. The four characteristics of essential genes (i.e. higher gene expression, higher degrees of protein–protein interactions, more homologs, less paralogs) were examined for all considered genes in the corresponding characteristic (ALL), essential genes by the previous techniques in Table 1 (PREV), essential genes by the proposed method (PROP, i.e. EGs in Section 3.1) and common essential genes between PREV and PROP (BOTH). The P-values written on the yellow and magenta bar indicate the significance of Welch's t-tests between PREV and PROP, and between PREV and BOTH, for each. More detailed statistics are depicted in Supplementary Table S6. MDA: MDAMB231, LNC: LNCAP)](https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/bty774/5089233)

4 Discussion and conclusion

In the LINCS, the expressions of 978 landmark genes (LMGs) were experimentally measured, and the expressions of remaining genes were imputed with a computational model based on the Gene Expression Omnibus (Duan et al., 2014). To characterize essential genes in this study, all genes including the landmark and imputed genes were considered; however, a few studies (Chen et al., 2017; Cheng and Li, 2016) showed that the quality of imputed expression in the LINCS is not perfect. Thus, here, we applied the identical process to the only LMGs, which characterized 147, 144, 102, 148, 224, 74 and 108 essential genes among 978 genes (P-value ≤ 1.0E-05) for each of A375, A549, BT20, LNCAP, MCF7, MDAMB231 and PC3 cell lines. Their essential scores and P-values are depicted in Supplementary Tables S7 and S8. When hypergeometric tests were applied between the essential genes of the LMGs and those of previous techniques (as Section 3.2), only A375 cell line produced significant results (P-value = 4.82E-07) (Supplementary Fig. S5a).

![Not enough commonly considered genes between the two gene sets could be a reason for the non-significant results (Supplementary Fig. S5b). As expected, the assessments of their essentiality characteristics (as Section 3.3) show very high significance in most comparisons (Supplementary Fig. S6).](https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/bty774/5089233)
Deconvoluting essential gene signatures for cancer growth in compound-treated cells

PD-0325901, i.e. MEK, to EIF4EBP1 can be a plausible mechanism of the drug inducing growth inhibition effect. As an example, we depicted the several shortest paths connecting MEK to EIF4EBP1 in the molecular networks consisting of protein interactions (Chatr-Aryamontri et al., 2015) and gene regulations (Matys, 2006) (Supplementary Fig. S9). We plan to examine those paths in more detail in further work.

In this study, the genomic expression profiles and cell growth inhibition in compound-treated cells were exploited to characterize essential signatures for cancer growth. The essential genes indeed have more essential characteristics, such as high gene expression, high degrees of protein–protein interactions, many homologs and few paralogs, than the rest of genes. Noticeably, the essential genes involving more essential characteristics were determined by combining the proposed and previous techniques than those solely that relied on the previous techniques. With an increase of diverse compound experiments on cells, the proposed technique could acquire more reliable results. We believe that this study will provide new aspects in the essential signature for cancer growth and be helpful for cancer research.

Funding
This work has been supported by the Bio-Synergy Research Project (NRF-2012M3A9C4048758) of the Ministry of Science and ICT through the National Research Foundation.

Conflict of Interest: none declared.

References
Subramanian, A. et al. (2017) A next generation connectivity map: 1,000 platform and the first 1,000,000 profiles. Cell, 171, 1437–1452. e1417.