Construction of a functional network for common DNA damage responses in Escherichia coli

Jungeui Hong, Joo-Myung Ahn, Byoung Chan Kim, Man Bock Gu

Abstract

In this study, we aim to identify a common, general mode of toxic action in Escherichia coli when experiencing DNA damage, irrespective of the agents used. We conducted or collected 69 microarray data from seven different DNA damaging agents. In a quantitative manner, we constructed a probable DNA damage stress network, entitled the Functional Linked Network (FLN), which consists of 389 significantly perturbed genes and the 1283 interactions among them. The SOS response related genes (LexA modules) were found to be dominantly activated by DNA damage, irrespective of the agents. Several minor, plausible modules were also implicated in this network, and appear to be related with the metabolic inhibition response to DNA damage or mediate the induction of SOS response. This systems and comparison approach across a variety of genotoxic agents may serve as a starting point to specify some of the unknown and common features of DNA damage responses in bacteria.

Introduction

DNA damage is a general phenomenon that spontaneously alters coding properties or the normal functions of DNA during replication or transcription [1]. This DNA damage can be introduced by both endogenous cellular processes, including oxidation, alkylation, and hydrolysis, and exogenous agents such as UV light, ionizing radiation, and other bulky chemical adducts [1,2]. Since the toxic mechanisms vary greatly, many elaborate and systematic repair systems must be embedded within the genetic make-up of cells in order to compensate for this damage. For example, bacterial cells provoke the so-called ‘SOS response’ to DNA damage, which is mediated and auto-regulated by the LexA and RecA protein pair and includes approximately 40 target genes [3]. As such, most studies concerned with DNA damage and the subsequent responses have focused on the SOS response as the primary defense system in bacteria. Aside from this response, some other unknown responses may occur on the transcriptional or translational level but have been overlooked due to the lack of global insight and molecular experimental data.

Recently, several studies have addressed DNA damage responses in E. coli using DNA microarray techniques and bioinformatics tools [4–10], which enabled us to observe transcriptional profiles at a systems level. Most of these studies aimed to characterize the toxicity of a certain agent using homogeneous experimental conditions and platforms. For example, one study suggested a predicted tran-
Table 1
List of chemicals and microarray data

<table>
<thead>
<tr>
<th>Type</th>
<th>Known mode of toxic action</th>
<th># of arrays</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMC</td>
<td>A family of aziridine-containing natural products, and has antibiotic activity as a potent DNA cross-linker via alkylation</td>
<td>12</td>
<td>This study</td>
</tr>
<tr>
<td>MNNG</td>
<td>A sort of carcinogen and mutagen, which alkylates DNA at multiple sites on nucleotide bases, sugars and produces severe DNA lesions</td>
<td>12</td>
<td>This study</td>
</tr>
<tr>
<td>NDA</td>
<td>A quinolone antibiotic, which targets the GyrA subunit of bacterial DNA gyrase or the topoisomerase IV enzyme, resulting in inhibition of DNA replication and transcription</td>
<td>12</td>
<td>This study</td>
</tr>
<tr>
<td>Norfloxacin</td>
<td>A type of fluoroquinolone, which results in inhibition of DNA replication and transcription as a gyrase inhibitor, like NDA</td>
<td>17</td>
<td>Sangurdekar et al. [5]</td>
</tr>
<tr>
<td>Novobiocin</td>
<td>An aminocoumarin antibiotic, which targets the GyrB subunit of bacterial DNA gyrase, resulting in inhibition of the ATPase reaction</td>
<td>4</td>
<td>Faith et al. [6]</td>
</tr>
<tr>
<td>UV radiation</td>
<td>UV light causes cross-linking between adjacent cytosine and thymine bases, creating pyrimidine dimers and resulting in a distorted DNA structures</td>
<td>7</td>
<td>Courcelle et al. [7]</td>
</tr>
<tr>
<td>Gamma radiation</td>
<td>The most dangerous irradiation, induces DNA alteration and DNA double-strand breaks even in higher eukaryotes</td>
<td>5</td>
<td>Sangurdekar et al. [5]</td>
</tr>
</tbody>
</table>

Total number of microarray used in this study 69

In total, 7 different DNA damaging agents and 69 microarrays were used. More details are described in the Additional Data File 4.

Results and discussion

Statistical index system for the selection of genes that are significantly responsive to DNA damage

We introduced a filtering step to discriminate genes that showed significant changes in their expression from experimental noise, and to quantify the degree of commonness of each gene with DNA damage. Among the approximately 4500 genes in E. coli, the genes for which the fold change compared to control was not more than 2 or less than 0.5 under any experimental condition were excluded, resulting in 3310 genes remaining. Then, we performed a statistical t-test for each gene expression set, and gained a $p$-value for each gene, where the reference set was calculated by averaging the expression values of all the genes for each experimental condition.

Next, the degree of perturbation in the expression level and its consistency toward the up- and down-regulation of each gene were determined. In a biological context, a higher degree of perturbation for a gene, referred to as ‘Norm’ value in this study, guarantees that the gene is more responsive to DNA damaging agents, regardless of whether the expression is up- or down-regulated. To rank genes by their consistency in their expression pattern, a novel index, the Constant Expression Ratio (CER), was introduced; an absolute CER value of close to 1 indicates that the gene is consistently expressed under as many of the experimental conditions as possible.

It was found that CER and $p$-values had a strong correlation with each other (Fig. 1A). Therefore, the CER value can be used as an effective indicator to ascertain the significance of each gene’s expression level under a given stimulus. The Norm and CER are only weakly correlated, which suggests that only few genes were consistently up- or down-regulated with a high degree of gene expression (Fig. 1B).

We tested how this novel index would fit within a real biological context with polycistronic genes (operon genes). Among the 1042 primarily selected genes, 396 genes were found to belong within 152 distinct operon relationships. In most cases, the genes within an operon were expressed in a similar pattern (see the consistency in the size and the color intensity of polycistronic genes in the Supplementary Fig. 1). This is expected since polycistronic genes usually have a similar magnitude of expression. Therefore, this novel index can be used to identify genes responsive to DNA damage, as was planned in this study.

![Fig. 1. Validation of CER and Norm index. In all, 3310 genes were plotted. (A) Scatter plot between the $-\log_{10}$ ($p$-value) and CER value. Interestingly, the $p$-value and the CER value showed an extremely high correlation with each other, which means that the CER can be also used as an effective indicator for significance in gene expression. (B) The Norm and CER value were much less related, but still showed a weak correlation. Scattered spots in the upper right hand sections of both (A) and (B) indicate genes that were significantly and constantly up-regulated during DNA damage, while those in the upper left hand sections were those that were constantly down-regulated. These three threshold values can be used in conjunction with each other to identify significantly perturbed genes.](image-url)
Functional Linked Network construction

The gene filtering steps were initially conducted twice to minimize the exclusion of false negatives. For the primary filtering, genes with a p-value of less than 0.01 were selected, resulting in 1042 genes from among a total of 3310 genes. From these, all the transcription factors (TFs) were investigated via comparison with DBD [18] and RegulonDB; in all, 41 known and 29 putative TFs were selected for further study.

Given the totally 69 microarray data, 906 genes were found to be partially correlated with at least one other gene based on Graphical Gaussian Model (GGM) method [19], and the total number of correlation linkages was 9085 (e.g. co-expression networks). Based on the operon information from RegulonDB [15], 396 genes are located within 152 operon relationships, as already mentioned. We also integrated any protein–protein interaction (PPI) network information available, which was derived from large-scale comprehensive pull-down assays, such as the His-tagged E. coli ORF clone library method [16] and the tandem affinity purification (TAP) method [17]. Among the primarily selected genes, 612 genes were shown to have some PPI interaction. With the transcriptional regulatory network (TRN) information from RegulonDB [15], a total of 34 TFs and 219 target genes, with 289 regulatory relationships between them, were integrated in this network. We also added sigma factor network information from RegulonDB [15], which consists of 4 sigma factors – \( \sigma^{32} \) (RpoD), \( \sigma^{A} \) (RpoE), \( \sigma^{aff} \) (FecI), \( \sigma^{N} \) (RpoE) – and their 374 target genes. From this, the primary ‘Functional Linked Network (FLN)’ consists of 1013 nodes (gene or protein) and 12,486 linkages.

For the secondary refinement step, we empirically determined which genes have Norm \( \geq 4.5 \) and an absolute CER \( \geq 0.35 \), since it approximates to p-value \( < 0.001 \) (Fig. 1). In sum, 400 genes were selected for the final FLN, where 32 were the regulatory factors (16 known TFs, 14 putative TFs and 2 sigma factors) and the other 368 were target genes that showed significant changes in their expression due to DNA damage. Sigma factor \( \sigma^{N} \) (RpoE) was, however, excluded during further analyses since, as a global regulator, it introduced lots of noise. Therefore, the final FLN consisted of 399 nodes and 1268 linkages. For network visualization, the CER value was transformed into a color/intensity map, where red and blue indicate genes that are constantly up- and down-regulated, respectively. The Norm value was set so that it was proportional to the size of each node. Fig. 2 shows the whole network construction scheme, and the primary and the final FLN results.

Clustering and Gene Ontology results for the final selected genes

Before delving deeper into the network structure, the hierarchical clustering analysis was applied for the finally selected 399 genes (31 regulators and their 368 target genes). Although the gene expression patterns when E. coli was exposed to UV, norfloxacin, and novobiocin were more perturbed relative to the others, most genes were classified into two large distinct clusters: activated and repressed (Fig. 3). This suggests that these 399 genes constitute the common DNA damage response despite the differences in the experimental conditions and even the DNA microarray platforms. For example, some SOS response genes regulated by LexA (see the dashed rectangular box in Fig. 3A) were consistently up-regulated, which surely reaffirms that the SOS response commonly occurs regardless of the DNA damage experienced. Fig. 3B shows that the 31 TFs were also separated, i.e., up- and down-regulated, and their CER values and expression patterns were very similar. Since the mRNA expression level and regulatory activity of TFs are not always consistent, we speculate that other TFs, besides LexA, also likely work as regulators of the DNA damage responses in E. coli.

Furthermore, the selected 399 genes were analyzed using Gene Ontology to determine which functional terms were enriched in the DNA damage response. All genes were classified into either an over- and under-represented grouping in terms of their positive and negative CER values, respectively. Using the GOminer tool [20], we extracted any up- and down-regulated GO terms with high level of significance, as shown in Table 2. As already expected, the most remarkably up-regulated terms were primarily DNA damage repair related processes, such as the SOS response. At the same time, several significantly down-regulated terms (flagellar motility, histidine family amino acid biosynthesis process, NADH dehydrogenase, metal ion transport, etc.) were found; with most of these appearing to be related with the inhibition of basal metabolic pathways to minimize DNA damage in the bacterial cells.

Highly enriched functional modules

In the FLN, the number of linkages that each TF has was further considered to distinguish the common, functional modules when E. coli experiences DNA damage. Here, we defined a ‘module’ as a probable functional unit that is enriched by a large number of operons, regulons, protein–protein interactions, and even co-expression relationships in the FLN. Highly linked TFs and their targets were focused on in order to identify the core functional ‘modules’ for the DNA damage responses. The list of TFs ordered by their number of linkages (or edges) is shown in Table 3. In this study, assuming that highly linked TFs may have more significant roles in a biological context, eight TFs, all of which have more than 10 linkages, were selected: Hns, Fur, LexA, CspC, FliD, DnaA, HcaR, and YeiE.

First of all, as expected, the LexA module was the most distinct module involved in DNA damage responses, irrespective of the genotoxic agents. The other modules were relatively skewed to parts of seven different DNA damage agents (see the CER value for each module), and are not well known for possible involvement in DNA damage responses. In particular, two TFs – cspC and yeiE – are putative, and have no prior regulatory information with their latent target genes. Therefore, we tried to infer unknown, conserved regulatory sequences within the latent target genes for each known or putative TF using a conserved motif search tool. However, only the LexA module was shown to have significant one. In this study, we identified eight modules as the most probable DNA damage networks in E. coli (Fig. 4).

However, high-throughput ‘omics’ information integrated in this study might contain potential false positive relationships mainly because of noisy data embedded in public database. In addition, only nodes (genes) by CER and Norm values were scored, not their linkages in this study. In fact, most functional linkages except LexA case don’t appear to be directly related to DNA damage responses in E. coli. Therefore, it could be suggested that a data integration strategy such as Bayesian network [21,22], which tries to score the linkages themselves, could be very useful to increase the significance of the network linkages in further study.

Fig. 2. Network construction scheme and the results. (A) Whole network construction scheme. (B) The primary selected network consists of 1013 nodes and 12,486 edges in total. (C) The final Functional Linked Network (FLN) consists of 399 nodes and 1268 edges in all. After the secondary filtering step, many of partial correlation linkages, which were likely false positively inferred, were removed from the network as shown in the pie charts of (B) and (C). Each node color was represented in red (up-regulated) and blue (down-regulated); the color intensity is proportional to the absolute value of the CER. Each node size is proportional to its Norm value, and the node shape is presented as round (known TFs), hexagonal (putative TFs), diamond (sigma factors), ellipsoid (target genes). Each edge and color was mapped differently, with regard to its network type, such as a partial correlation (pc), operon (op), protein–protein interaction (ppi), sigma network (sigg), positive TF regulation (tf+) and negative TF regulation (tf−).
A 7 Genotoxicants (MMC, NDA, MNNG, UV, 
Gamma, Novobiocin, Norfloxacin)

69 Expression Profiles
(~4,500 ORFs)

Static genomic features

1st FLN
(1,013 Nodes / 12,486 Linkages)

Integration

Assign Norm & CER
(Size & Color)

Correlation Inference

Final FLN
(399 Nodes / 1,286 Linkages)

Primary Filtered genes
C_i, C_j, ..., C_n, C_m

a_i = log_2 ratio of the i_th Gene in the j_th Condition
(n = 69, k = 1,042)

For each i_th gene

Norm(i) = \sqrt{\sum_{j=1}^{n} a_{ij}^2}

CER(i) = \frac{\sum_{j=1}^{n} a_{ij}}{\sqrt{n \times \sum_{j=1}^{n} a_{ij}^2}}

Co-expression Network
(906 Nodes / 9,085 Linkages)

B

C

sig 3%

3%

tf 2%

2%

op 11%

11%

pc 80%

80%

sig 0%

0%

tf 7%

7%

ppi 24%

24%

op 14%

14%

pc 55%

55%
LexA module

The SOS response, as regulated by the LexA suppressor, is the best known response to DNA damage [23] in many bacteria. Under normal conditions, LexA binds to the promoter region of the target genes and represses their expression, while at the onset of DNA damage stress it is cleaved by the RecA protein, which is activated by broken single-strand DNA sequences. This cleavage results in the activation of SOS response genes and the repair of broken DNA [24]. The LexA protein also represses its own gene, *lexA*, and more than 30 genes are known to be precisely
Table 2
Significantly represented GO terms within DNA damage

<table>
<thead>
<tr>
<th>GO ID</th>
<th>Term</th>
<th>Total</th>
<th>Under</th>
<th>Over</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0009432</td>
<td>SOS response</td>
<td>17</td>
<td>0</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>GO:0005694</td>
<td>Chromosome</td>
<td>20</td>
<td>1</td>
<td>5</td>
<td>0.0004</td>
</tr>
<tr>
<td>GO:0051716</td>
<td>Cellular response to stimulus</td>
<td>22</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>GO:0009991</td>
<td>Response to extracellular stimulus</td>
<td>22</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>GO:0033554</td>
<td>Cellular response to stress</td>
<td>22</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>GO:0031668</td>
<td>Cellular response to extracellular stimulus</td>
<td>22</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>GO:0009605</td>
<td>Response to external stimulus</td>
<td>45</td>
<td>5</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>GO:006260</td>
<td>DNA replication</td>
<td>63</td>
<td>6</td>
<td>9</td>
<td>0.0002</td>
</tr>
<tr>
<td>GO:0069794</td>
<td>Response to DNA damage stimulus</td>
<td>68</td>
<td>2</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>GO:0062821</td>
<td>DNA repair</td>
<td>68</td>
<td>2</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>GO:0009719</td>
<td>Response to endogenous stimulus</td>
<td>69</td>
<td>2</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>GO:0006950</td>
<td>Response to stress</td>
<td>151</td>
<td>4</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>GO:0007154</td>
<td>Cell communication</td>
<td>172</td>
<td>11</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>GO:0062529</td>
<td>DNA metabolic process</td>
<td>180</td>
<td>9</td>
<td>28</td>
<td>0</td>
</tr>
</tbody>
</table>

DNA damage repair related processes, such as the SOS response, were over-represented, while flagellar motility, the histidine family amino acid biosynthesis process, NADH dehydrogenase, metal ion transport, etc. were under-represented. All terms in the fl." regulated by the interaction of the LexA-RecA protein pair in E. coli [3].

We found 29 regulatory interactions between LexA and its target genes from the RegulonDB database (release: 5.7, Date: 08-JUN-07) [15]. Among them, 16 (55%) genes were found to be perturbed by DNA damage responses in our FLN scheme, as shown in Fig. 5 (see only the dashed and orange-color arrows). Most importantly, the CER and Norm values for all the genes in the LexA module were noticeably higher than the others, which reaffirms that these are commonly responsive, irrespective of types of agents.

Many of target genes inferred by our FLN were already characterized in a previous work [25]. For instance, dinI and yebG are reported to be involved in the SOS regulon, but their precise TF-binding regions are still putative. dinI is known to have an effect on the activity and function of the RecA protein when DNA damage is introduced [26]. yebG is, however, merely known to be induced by MMC [27], but details about its latent roles in the SOS response have not been reported well. rmuC has also been reported to be induced by MMC and NDA [26], and produce a predicted recombination limiting protein in E. coli [26], and were found to be repressed. Consequently, we speculate that iron uptake is hindered or inhibited after DNA damage occurs in E. coli, and suggests that the iron concentration inside the cells should be controlled by certain defense mechanisms when the bacterium experiences DNA damage.

A recent study has confirmed that iron concentration inside the cells tends to diminish via repression of the iron uptake-related genes so that secondary oxidative damage is minimized after an exposure to gyrase inhibitors, such as norfloxacain [11]. In our FLN, the Fur module genes were perturbed mainly by novobiocin, UV, NDA, and norfloxacain, but not gamma radiation MMC or MNNG. Therefore, inhibition of iron uptake appears to be a specific response to certain types of agents.

**Fur module**

Fur (ferric uptake regulator), usually as a negative regulator, is associated with the regulation of a large number of operons that encode iron transport-related enzymes [32] and several outer-membrane proteins [33]. Most interactions (18 out of 22) in the Fur module were derived from transcriptional regulatory network (Fig. 4). The CER value for the fur gene was relatively high (0.60) even if its expression was not so pronounced in the gamma radiation, MMC, and MNNG cases (Fig. 3B).

Most of its linked genes except only two – flhA and umuD – were repressed in our FLN. Among the 22 genes in the Fur module, several genes, i.e., fecA, fecB, fecC, fecE, entC, entE, and flh, are engaged in ‘iron ion transport’ in E. coli, and were found to be repressed. Consequently, we speculate that iron uptake is hindered or inhibited after DNA damage occurs in E. coli, and suggests that the iron concentration inside the cells should be controlled by certain defense mechanisms when the bacterium experiences DNA damage.

**CspC module**

CspC has sequence homology to CspA, the major cold shock protein in E. coli, but is not cold-shock inducible and constitutively expressed even at 37 °C [34]. It is also known to bind to RNA or single-stranded DNA sequences, specifically AU/AT-rich regions [35], and is predicted to have DNA binding affinity as a putative TF based on information available in DDB and RegulonDB. The CER (−0.55) and the Norm (7.3)
values for the cspC gene in our FLN were not as significant as seen for lexA, but were some of the highest seen. Based on the functional categories in the TIGR database, most genes linked to CspC in the FLN are functionally hypothetical or unclassified, and seem to share no common roles. Most of their linkages to CspC were derived merely from co-expression inference or PPI data (Fig. 4). Furthermore, there is no direct evidence for their regulatory relationship as a part of a DNA damage network in E. coli and no consensus regulatory motif was found via the MEME analysis in this study. Nonetheless, the CspC module might have some specific roles in survival when the cellular DNA was damaged since this protein is a transcriptional regulator. Actually, cspC is known to affect the expression of several stress proteins by regulating the activity of RpoS, a global stress-response regulator [36]. Due to its transcription anti-terminator activity [37], the reduced expression of the genes within the CspC module when the DNA is damaged suggests that the cells try to facilitate the termination of transcription. As such, it would be worthwhile to study this module further to elucidate its functional relevance to DNA damage responses.

Other modules: FlhD, DnaA, HcaR, YeIE

Combined with FlhC, FlhD acts as a compound sigma factor that activates class II flagellar genes [38]. The FlhD module seems to be significantly under the control of Hns and Fur (Fig. 4). The repressed Hns and activated Fur activities mutually suppressed the flhD genes, which may reflect that when the cells experience DNA damage they slow down the production of the flagellar proteins, resulting in a retardation of their motility. It also suggests that the cells reduce and minimize their energy consumption, since the flagellar engine is powered by the proton motive force (e.g., it requires energy).

DnaA is a transcriptional regulator involved in the initiation and activation of chromosome replication [39], and is also known to be a critical mediator of the RecA-independent (non-SOS) response to DNA damage [40]. Especially, the recf gene is induced by DnaA and this protein seems to maintain replication fork arrest during DNA damage by activating the RecA protein [41] and was shown to be relatively up-regulated in this study. Furthermore, expression of dnaA was shown to be significantly perturbed by MMC and UV in another study with B. subtilis [12]. Based on these results, it could be speculated that the DnaA module might be an early inducer of the SOS response or the SOS-independent pathway, but its significance is much lower than LexA, in terms of gene expression.

Another module is regulated by HcaR, which is related with carbohydrate metabolism and oxidative stress responses [42], while YeIE is a putative regulator probably involved in lysine metabolism [43]. These two regulators might constitute minor responses to DNA damage and stress, but have never been studied in detail.

Materials and methods

Bacterial strain and lethality test

In this study, MMC, MNNG and NDA were purchased from the Sigma-Aldrich Co., USA. For each chemical, sub-lethal, LC20 and LC50 concentrations were determined using the cell growth rate compared to control samples. For the test strain, E. coli RMM443 (strR, galK2, lacΔ74) was used as a wild-type (WT) [44–46]. To maintain homogeneous culture conditions, we followed certain steps in all experiments: (i) seed one colony, which was grown on agar plates overnight, into a 15 ml tube (Corning, MA, USA) with 4 ml Luria-Bertani (LB) media, (ii) incubate at 37 °C until the O.D. reached 0.8, with shaking at 200 rpm, (iii) aliquot 1 ml of the culture into a new flask containing 25 ml fresh LB media, (iv) incubate again at 37 °C until the O.D. reached 0.3, i.e., early exponential phase of E. coli growth, shaking at 200 rpm, (v) add each prepared concentration of MMC, MNNG and NDA. The O.D. values for each growth test were measured every 10 min for about 2 h after exposure, at which time the culture was entering the stationary phase.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene description</th>
<th>Norm</th>
<th>CER</th>
<th>p-value</th>
<th># of links</th>
<th>TF</th>
</tr>
</thead>
<tbody>
<tr>
<td>hns</td>
<td>DNA-binding protein HUP-II</td>
<td>7.43</td>
<td>-0.47</td>
<td>7.31E-05</td>
<td>34</td>
<td>Known</td>
</tr>
<tr>
<td>fur</td>
<td>Negative regulator</td>
<td>6.87</td>
<td>0.6</td>
<td>1.49E-08</td>
<td>21</td>
<td>Known</td>
</tr>
<tr>
<td>lexA</td>
<td>Regulator for SOS(lexA) region</td>
<td>10.93</td>
<td>0.74</td>
<td>5.11E-14</td>
<td>20</td>
<td>Known</td>
</tr>
<tr>
<td>cspC</td>
<td>Cold shock protein</td>
<td>7.32</td>
<td>-0.55</td>
<td>2.17E-06</td>
<td>19</td>
<td>Putative</td>
</tr>
<tr>
<td>fhbD</td>
<td>Regulator of flagellar biosynthesis</td>
<td>5.98</td>
<td>-0.47</td>
<td>1.09E-04</td>
<td>17</td>
<td>Known</td>
</tr>
<tr>
<td>dnaA</td>
<td>DNA biosynthesis</td>
<td>5.74</td>
<td>0.55</td>
<td>7.13E-08</td>
<td>1</td>
<td>Known</td>
</tr>
<tr>
<td>hcaR</td>
<td>Transcriptional activator of hca cluster</td>
<td>4.88</td>
<td>-0.45</td>
<td>4.06E-04</td>
<td>1</td>
<td>Known</td>
</tr>
<tr>
<td>yeIE</td>
<td>Putative transcriptional regulator</td>
<td>4.7</td>
<td>-0.58</td>
<td>9.34E-07</td>
<td>10</td>
<td>Putative</td>
</tr>
<tr>
<td>pboB</td>
<td>Positive response regulator for pbo regulon</td>
<td>6.92</td>
<td>0.54</td>
<td>9.53E-07</td>
<td>7</td>
<td>Known</td>
</tr>
<tr>
<td>ycgE</td>
<td>Putative transcriptional regulator</td>
<td>5.72</td>
<td>-0.58</td>
<td>5.62E-07</td>
<td>8</td>
<td>Putative</td>
</tr>
<tr>
<td>matR</td>
<td>Putative repressor of mal regulon</td>
<td>5.72</td>
<td>0.5</td>
<td>2.04E-06</td>
<td>6</td>
<td>Known</td>
</tr>
<tr>
<td>yhiL</td>
<td>Putative transcriptional regulator</td>
<td>5.25</td>
<td>-0.44</td>
<td>3.28E-04</td>
<td>5</td>
<td>Putative</td>
</tr>
<tr>
<td>ycfQ</td>
<td>orf, hypothetical protein</td>
<td>5.96</td>
<td>0.52</td>
<td>2.24E-06</td>
<td>4</td>
<td>Putative</td>
</tr>
<tr>
<td>ydrD</td>
<td>Putative ARAC-type regulatory protein</td>
<td>7.5</td>
<td>0.49</td>
<td>3.10E-04</td>
<td>4</td>
<td>Known</td>
</tr>
<tr>
<td>iclR</td>
<td>Repressor of aceB operon</td>
<td>4.73</td>
<td>-0.45</td>
<td>3.26E-04</td>
<td>4</td>
<td>Known</td>
</tr>
<tr>
<td>fer</td>
<td>Probable RNA polymerase sigma factor</td>
<td>9.31</td>
<td>-0.5</td>
<td>1.85E-05</td>
<td>4</td>
<td>Sigma</td>
</tr>
<tr>
<td>appY</td>
<td>Regulatory protein affecting appA and ot</td>
<td>8.05</td>
<td>0.46</td>
<td>3.77E-05</td>
<td>3</td>
<td>Known</td>
</tr>
<tr>
<td>attE</td>
<td>Activator of attl gene</td>
<td>6.92</td>
<td>-0.45</td>
<td>1.96E-04</td>
<td>3</td>
<td>Putative</td>
</tr>
<tr>
<td>livY</td>
<td>Positive regulator for livC</td>
<td>6.1</td>
<td>-0.48</td>
<td>6.62E-05</td>
<td>3</td>
<td>Known</td>
</tr>
<tr>
<td>cytR</td>
<td>Regulator for deo operon, udp, cdd, tss</td>
<td>7.96</td>
<td>-0.5</td>
<td>2.79E-05</td>
<td>3</td>
<td>Known</td>
</tr>
<tr>
<td>yeIE</td>
<td>orf, hypothetical protein</td>
<td>5.79</td>
<td>-0.53</td>
<td>7.29E-06</td>
<td>3</td>
<td>Known</td>
</tr>
<tr>
<td>glcN</td>
<td>Response regulator for gln (sensor glnL)</td>
<td>5.92</td>
<td>-0.53</td>
<td>6.98E-06</td>
<td>3</td>
<td>Known</td>
</tr>
<tr>
<td>yqel</td>
<td>Putative sensory transducer</td>
<td>8.49</td>
<td>0.38</td>
<td>8.73E-04</td>
<td>2</td>
<td>Putative</td>
</tr>
<tr>
<td>pdhR</td>
<td>Transcriptional regulator for pyruvate d</td>
<td>8.42</td>
<td>-0.5</td>
<td>1.96E-05</td>
<td>2</td>
<td>Known</td>
</tr>
<tr>
<td>ybeF</td>
<td>Putative transcriptional regulator</td>
<td>5.09</td>
<td>0.46</td>
<td>3.02E-05</td>
<td>1</td>
<td>Putative</td>
</tr>
<tr>
<td>envY</td>
<td>Envelope protein; thermoregulation of po</td>
<td>5.48</td>
<td>0.41</td>
<td>2.74E-04</td>
<td>1</td>
<td>Known</td>
</tr>
<tr>
<td>cmn</td>
<td>Putative carbonic anhydrase (EC 4.2.1.1)</td>
<td>9.86</td>
<td>-0.41</td>
<td>9.78E-04</td>
<td>1</td>
<td>Putative</td>
</tr>
<tr>
<td>yidZ</td>
<td>Transcriptional regulator</td>
<td>6.26</td>
<td>-0.43</td>
<td>5.22E-04</td>
<td>1</td>
<td>Putative</td>
</tr>
<tr>
<td>yioU</td>
<td>Transcriptional regulator</td>
<td>5.44</td>
<td>0.57</td>
<td>1.18E-07</td>
<td>0</td>
<td>Putative</td>
</tr>
<tr>
<td>agkK</td>
<td>Prophage P24 protein</td>
<td>8.58</td>
<td>0.39</td>
<td>4.60E-04</td>
<td>0</td>
<td>Putative</td>
</tr>
<tr>
<td>yerK</td>
<td>orf, hypothetical protein</td>
<td>7.84</td>
<td>-0.46</td>
<td>1.15E-04</td>
<td>0</td>
<td>Putative</td>
</tr>
</tbody>
</table>

Assuming a TF that has more linked genes indicates a more probable regulatory network, the TFs were ordered by number of their linkages. TFs which have more than 10 linkages are shown shaded gray, and were selected as probable ‘hub’ regulators in DNA damage networks. The term ‘module’ consists of a TF and its latent target genes and was used since confirmation of the regulons still needs to be performed.
Total RNA isolation and microarray experiment

For each chemical and concentration, 2 ml of the cell cultures was harvested at 5, 25, 45 and 65 min for control and test samples in parallel. The harvested cells were mixed with 4 ml of bacterial RNA protectant (QIAGEN, USA) and placed on ice to inhibit further RNA synthesis and degradation. Total RNA isolation was done using the RNeasy RNA extraction kit (QIAGEN, USA) according to the manufacturer’s protocol. The quality and concentration of each RNA sample was confirmed using ND-1000 UV/Vis spectrophotometer (Nanodrop, Wilmington, USA). Then, 15.9 μl of each purified and concentrated RNA sample, for the control and test, was mixed with 2 μl of random hexamer and heated at 65 °C for 10 min. After cooling on ice and spinning down the samples, 9.6 μl of a pre-reaction mixture (6 μl of 5× First Strand Buffer, 0.6 μl unlabeled dNTPs and 3.0 μl 0.1 M DTT) was added to each sample, respectively. Then 2 μl Cy3-dUTP and Cy5-dUTP were added for sample labeling. Next, 1 μl of Powerscript™ Reverse Transcriptase (Clontech, CA, USA) was added to each sample mixtures, and the resulting 30 μl mixtures were incubated at 40 °C for 2 h for the reverse transcription process. Next, 15 μl NaOH (0.1 N) was added and mixed by vortexing, and the samples were incubated at 65 °C for 10 min to remove any RNA remaining and to stop the PCR reaction. An additional 15 μl of HCl (0.1 N) was added to each sample in order to neutralize the solution. The labeled cDNA solutions for the control and test samples were then mixed into one sample and purified using a PCR purification kit (QIAGEN, USA) according to the manufacturer’s protocol. To gain more concentrated solutions, the elution step was done using microcon filters (Millipore, MA, USA) and with EB buffer and brought up to a final volume of 27 μl. Then 20 μl of 20× SSC, 8 μl of 1.0% SDS, 24 μl of formamide and 1 μl of salmon sperm DNA were mixed with the purified labeled cDNA solutions. These final mixtures were denatured at 100 °C for 3 min, and then centrifuged at 13,000 rpm for 2 min.

Before applying the above final solutions to the arrays, the E. coli Oligo 6K microarray chips (Genomictree, Korea) were pre-hybridized in a blocking solution (5× SSC, 0.1% SDS, 1% BSA with water) at 42 °C for 30 min, and washed with sufficient distilled water for 2 min and isopropanol for 1 min, successively, and then dried by spinning at 600 rpm. Then, the above solutions were applied onto the pre-hybridized chips, covered and sealed within a dark chamber and hybridized at 42 °C for 16 h. Finally, the following solutions were prepared and used for the final washing step successively: (1) 2× SSC and 0.1% SDS buffer twice times for 1 and 5 min, repeatedly, (2) 0.1× SSC and 0.1% SDS buffer for 10 min, (3) 0.1× SSC buffer for 2 min. Afterwards, the chips were dried by centrifugation at 1200 rpm and room temperature for 3 min.

Scanning and data pre-processing of microarray

The completely hybridized microarrays were scanned with a GenePix 4000B laser scanner (Axon Instrument, Inc., CA, USA). The intensity ratio between the two wavelengths for each dye was adjusted to give a value of near 1; and the scanned images were saved as a multi-image TIFF file format. Spot validation was performed with GenePix Pro 3.0 software (Axon Instrument, Inc., CA, USA). For normalization, R language software (version 2.5.1) and a ‘limma’ package (version 2.9.1), as a part of Bioconductor project [47–49], were used. Background correction, scale normalization between the arrays and loess normalization, which are embedded functions in ‘limma’ package [50], were applied. For norfloxacin, novobiocin, UV
and gamma radiation tests, the published data from the GEO database was gathered and used without additional normalization (Supplementary Table 1).

The final result was represented as the log2 ratio value between test and control sample. ID conversion between the Blattner ID and UniProt ID was conducted by referring to the Ecogene[51] and EcoCyc databases [14]. And genes which do not have any absolute log2 expression value over 1 in the 69 samples were removed, resulting in the original set of 3310 genes.

Bioinformatics analysis

For clustering analysis, we used Cluster software [52], where a hierarchical clustering method was chosen: all genes were clustered only by ‘Genes’, not by ‘Arrays’ option, and Metric (distance) and Linkage were set as ‘correlation (centered)’ and ‘Average’, respectively. For the Gene Ontology analysis, we used the GOMiner tool [20], where all options were ‘default’, and the gene ID was transformed into a UniProt ID system. To acquire the biological and statistical significance, only GO terms which have a total gene number of more than 10 and less than 200 were selected, and the p-value threshold was set as less than 0.001.

The degree of correlation between every combination of all genes was calculated by a Graphical Gaussian Model (GGM) [19], embedded in the ‘GeneNet’ R/Bioconductor package. The local FDR cut-off value and the p-value threshold for the partial correlation of each edge were chosen as more than 0.02 and less than 0.0001, respectively, resulting in 906 genes and 9085 edges (gene–gene interactions) as significant. Higher partial correlation values between two given genes mean that they are similarly expressed, and functionally linked to each other. For the network visualization and analysis, Cytoscape software (Version 2.5.1) was used [53].

To confirm the putative target genes and promoter regions of unknown target genes for each TF, the MEME algorithm [30] was applied. In the options, the distribution of occurrences for a single motif among the sequences was set as “Zero or one per sequence”, and the minimum and maximum widths of each motif were set at 6 and 50, respectively. Likewise, the maximum number of motifs was set at 5 and the “Search given strands only” option was selected. The motif with the highest E-value was chosen as the optimal one. For visualization of the chosen consensus motif, the Weblogo program [54] was used, where all the options were ‘default’.

Significance test for the genes

Here, the Expression Matrix (EM) is represented as:

\[ C_j \text{ (condition set)} \]

\[ G_i \text{ (gene set)} \]

where, \( k \) is 3310 (total number of genes), \( n \) is 69 (total number of conditions), and \( a_{ij} \) is a log2 ratio expression value of \( l_{ji} \) gene for \( j \) condition.

p-value calculation for each gene

From each condition set \( C_j = (a_{1j}, a_{2j}, \ldots, a_{nj}) \), a reference set was generated by averaging all the values in each \( C_j \) where \( j \) is from 1 to 69. The reference set serves as a null distribution in the p-value calculation. For all \( G_i = (a_{1i}, a_{2i}, \ldots, a_{ni}) \), the student t-test was conducted, assuming that all distributions of \( a_{ij} \) values are converging to normal distribution. Then, the p-value was calculated, where a two-sided test was selected since we intended to confirm whether or not each \( G_i \) might come from the reference set or just by chance. Lower p-values guarantee that such a gene was significantly perturbed by DNA damage, not merely by chance, when compared to the reference set.

Novel index system for constantly expressed genes

i) Define the degree of perturbation of each gene in gene expression as Euclidean Norm value: Given \( G_i = (a_{1i}, a_{2i}, \ldots, a_{ni}) \) for gene \( i \),

\[
\text{Degree of perturbation} := \text{Norm} = \sqrt{a_{1i}^2 + a_{2i}^2 + \ldots + a_{ni}^2}
\]

\[ (1.1) \]

ii) Assume that the expression consistency of a gene is proportional to the average value of all expression values, and inversе- proportional to the degree of perturbation (Norm):

\[
\text{Expression Consistency} \propto \frac{\text{average}}{\text{Norm}} = \frac{\sum_{i=1}^{n} a_i}{\sqrt{\sum_{i=1}^{n} a_i^2}} = \frac{\sum_{i=1}^{n} a_i}{\sqrt{n \times \sum_{i=1}^{n} a_i^2}}
\]

\[ (2.1) \]
Finally,

\[
\begin{align*}
\max \left( \frac{\text{average}}{\text{Norm}} \right) &= \frac{1}{\sqrt{n}} \quad \text{if } a_1 = a_2 = \ldots = a_n > 0 \\
\min \left( \frac{\text{average}}{\text{Norm}} \right) &= \frac{1}{\sqrt{n}} \quad \text{if } a_1 = a_2 = \ldots = a_n < 0 , \text{ and}
\end{align*}
\]

(2.2)

(Proof: see the Supplementary information).

iii) Based on Eq. (2.2)

For each gene, we can define a normalized index such that

\[
\text{Constant Expression ratio (CER)} := \sqrt{n} \times \frac{\text{average}}{\text{Norm}}
\]

\[
= \sqrt{n} \times \frac{\sum_{j=1}^{n} a_{ij}}{n} = \sqrt{n} \times \frac{\sum_{j=1}^{n} a_{ij}^2}{n} = \sqrt{\frac{\sum_{j=1}^{n} a_{ij}^2}{n}}
\]

Then,

\[
-1 \leq \text{CER} \leq 1
\]

If |CER|≈1, a higher probability exists that the gene will show a constant expression pattern, irrespective of origin of the stress. The sign for the CER value represents up(+) or down(−) regulation of the given gene.

Acknowledgments

This research was supported by Ministry of Science and Technology (MOST) of Korea through Advanced Environmental Monitoring and Evaluation Research Center (ADEMRC) Satellite lab (Germany) project. Both Joohyung Ahn and Jeungeun Hwang were trained in Dr. An-Ping Zeng’s lab at GFB in Braunschweig, Germany for 3 and 2 months, respectively, under the ADEMRC Satellite program. The authors are grateful for the support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jgenet.2009.01.010.

References


