

# Systematic Inspection of the Clinical Relevance of *TP53* Missense Mutations in Gastric Cancer

SeongRyeol Moon<sup>1</sup>, Curt Balch, Sungjin Park, Jinhyuk Lee, Jiyong Sung, and Seungyoon Nam<sup>1</sup>

**Abstract**—The “guardian of the genome,” *TP53*, is one of the most frequently mutated genes of all cancers. Despite the important biological roles of *TP53*, the clinical relevance of *TP53* mutations, in gastric cancer (GC), remains largely unknown. Here, we systematically assessed clinical relevance, in terms of *TP53* mutation positions, finding substantial variability. Thus, we hypothesized that the position of the *TP53* mutation might affect clinical outcomes in GC. We systematically inspected missense mutations in *TP53*, from a TCGA (The Cancer Genome Atlas) GC dataset in UCSC Xena repository. Specifically, we examined five aspects of each mutational position: (1) the whole gene body; (2) known hot-spots; (3) the DNA-binding domain; (4) the secondary structure of the domain; and (5) individual mutation positions. We then analyzed the clinical outcomes for each aspect. These results showed that, in terms of secondary structure, patients with mutations in turn regions showed poor prognosis, compared to those with mutations in beta strand regions (log rank  $p = 0.043$ ). Also, in terms of individual mutation positions, patients having mutations at R248 showed poorer survival than other patients having mutations at different *TP53* positions (log rank  $p = 0.035$ ).

**Index Terms**—Bioinformatics, computational biology, biomedical informatics

## 1 INTRODUCTION

THE gene encoding the tumor suppressor *TP53*, dubbed the “guardian of the genome,” is the most frequently (> 50 percent) mutated gene across all cancers [1]. The extent of *TP53* mutation, however, in specific tumors, varies widely. A previous study of the “mutational landscape,” of 12 distinct tumor types, found 20 biologic processes to be significantly dysregulated. Moreover, in one such process, “genome integrity,” *TP53* mutation rates ranged from 2.2 percent to 94.9 percent, in renal cell and serous ovarian cancers, respectively [2].

Although *TP53* is widely known as the prevailing arbitrator of DNA repair and apoptosis, one study found that its role in DNA repair, in mice, was unrelated to tumor suppression. Rather, activation by the small protein, ARF, in response to oncogenic signaling, resulted in *TP53*'s

antineoplastic properties [3]. However, while those authors accurately noted well-known, frequent deletions, or epigenetic silencing of ARF, they did not examine the transcriptomic context of *TP53* activity, nor its specific mutation [1].

Previously, using patient stratification based on *TP53* mutation status, we identified clinically relevant biomarker candidates, which significantly associated with *TP53* status/mutation in clinical outcomes [4]. Previously, all mutational positions in *TP53* were considered the same. In other words, individual mutation positions were not inspected in terms of clinical relevance. Here we extended that previous work to assess the role of specific *TP53* mutations, on clinical outcomes, finding substantial variability.

## 2 METHODS

### 2.1 Overview of Our Study

Based on the previous *TP53* studies described above, we hypothesized that *TP53* mutational location might affect clinical outcomes, in gastric cancer (GC), via unique mechanisms. To test that posit, we first studied relationships between mutation locations and survival rate, in addition to differentially expressed pathways, using *TP53* missense mutations in the dataset (henceforth, The Cancer Genome Atlas (TCGA) GC dataset, or equivalently TCGA STAD dataset) from Stomach Adenocarcinoma (STAD) in TCGA [5]. Our overall approach is shown in Fig. 1.

The number of samples collected was 289 in total, which included a mutation dataset, and a matched expression dataset, from the TCGA. Among these, 138 samples had mutations in *TP53*, while 151 were *TP53* wild type. Of the 138 *TP53*-mutant patient samples, we specifically examined 80 (58 percent) samples, having only a single missense *TP53* mutation.

- S. Moon is with the Department of Health Sciences and Technology, Gachon Advanced Institute for Health Sciences and Technology (GAIHST), Gachon University, Incheon 21565, Korea. E-mail: moonsr1982@gmail.com.
- C. Balch is with the Complex Biological Systems Alliance, North Andover, MA 01845. E-mail: curt.balch@gmail.com.
- S. Park is with the College of Medicine, Gachon University, Incheon 21565, Korea. E-mail: oscar.park@gmail.com.
- J. Lee is with the Genome Editing Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Korea. E-mail: jinhyuk@kribb.re.kr.
- J. Sung is with the Department of Genome Medicine and Science, Gachon University Gil Medical Center, Incheon 21565, Korea. E-mail: jysung329@gmail.com.
- S. Nam is with the College of Medicine, Gachon University, Incheon 21565, Korea. E-mail: nams@gachon.ac.kr.

Manuscript received 5 Dec. 2017; accepted 3 Mar. 2018. Date of publication 8 Mar. 2018; date of current version 7 Oct. 2019.

(Corresponding author: Seungyoon Nam.)

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.

Digital Object Identifier no. 10.1109/TCBB.2018.2814049

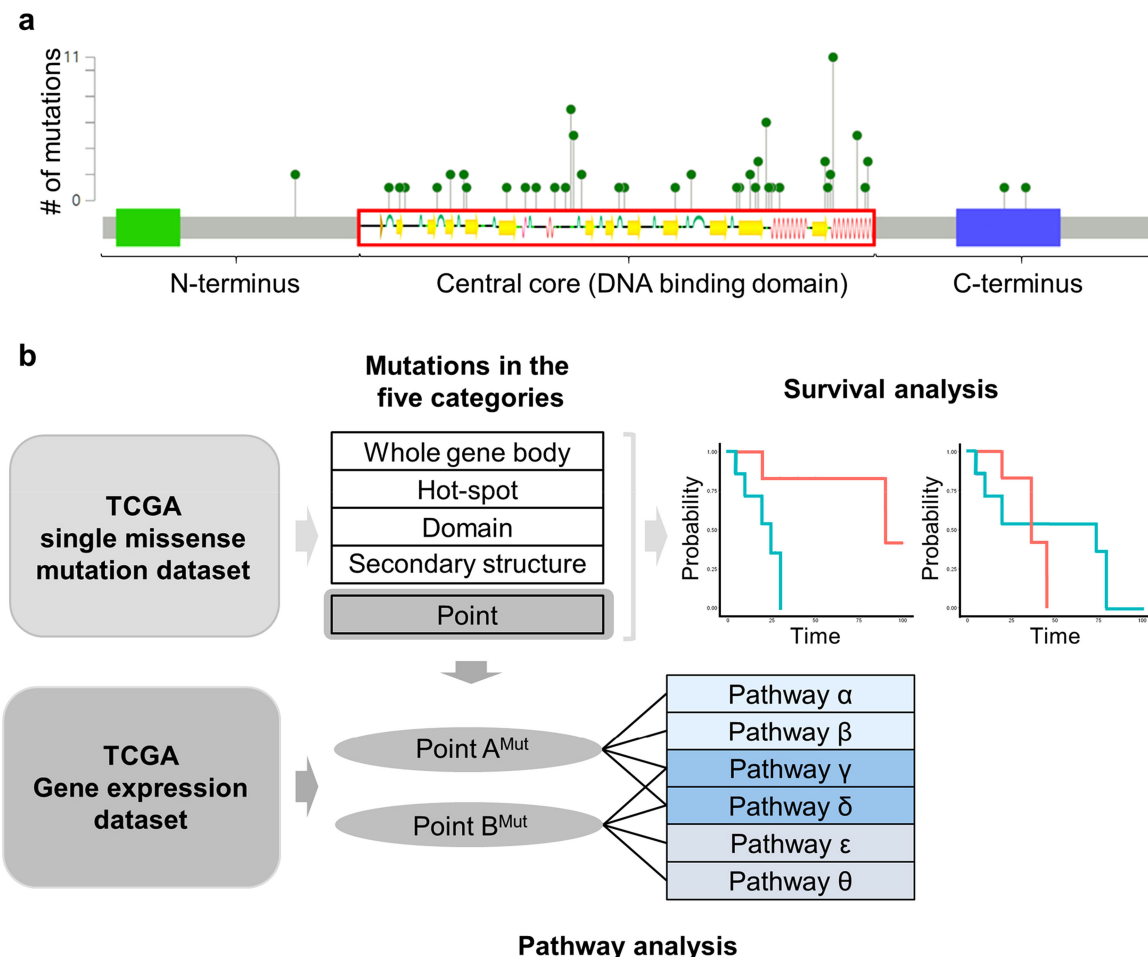


Fig. 1. Scheme of our study. (a) The distribution of TP53 missense mutations in TCGA gastric cancer patients according to domains. The 80 missense mutations in 80 patients were considered in the study. (b) The mutations in a were inspected in terms of five categories (gene body, TP53 hot spots, domains, secondary structures, and points) to identify clinical relevance (e.g., survival). In addition, we inspected pathways associated with clinically relevant mutations, revealing specific signaling, as well as common signaling according to mutation positions.

The distribution of the 80 missense mutations in the 80 patient samples in the TCGA is shown in Fig. 1a. The missense mutations were distributed into 38 distinct positions of the TP53 gene, and up to 11 samples were mutated at the same position. Of the 80 samples with only one missense mutation, 76 samples (95 percent) were mutated in the DNA-binding domain (DBD) (Fig. 1a). Of the four remaining samples, two (2.5 percent) were in the N-terminus and two (2.5 percent) were in the C-terminus.

As shown in Fig. 1b, we next divided the group of 80 missense mutations into five categories: (1) whole gene body; (2) hot spots; (3) domain; (4) secondary structure; and (5) positions (within DBD). We then performed survival analysis for each category. By combining the mutation information with the expression dataset (matched to the same patients), we also assessed differentially expressed pathways, associated with individual missense mutational positions.

## 2.2 Data Collection

We used data from TCGA STAD somatic mutation (curated by the Broad Institute, Cambridge, MA) database (version 2016-04-25), of the UCSC XENA group. According to the data description of UCSC XENA, sequencing data were generated on an Illumina (San Diego, CA, USA) Genome Analyzer system. Calls were generated at the Broad Institute Genome

Sequencing Center, using the MuTect method [6]. This specific mutation dataset holds information on mutations for 289 gastric cancer (GC) patients. Gene expression data was obtained using the TCGA STAD gene expression dataset, as determined by RNA-Seq of the UCSC Cancer Genomic Browser (CGB) group. Patients' gene expression data was matched to that of the mutation data. According to the data manipulation of UCSC CGB, gene expression levels were equal to log2 transformation of RPKM values, with level 3 data downloaded from the TCGA data coordinating center. Clinical information was downloaded from the same web page as the gene expression dataset. Secondary structure data for the TP53 protein was obtained from DSSP for TP53 model 3Q01, deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB-PDB) [7].

## 2.3 TP53 Missense Mutation Visualization

Extraction data with one TP53 missense mutation was used in the TCGA STAD mutation dataset, and the mutation mapper of cBioPortal was used for visualization [8]. The secondary structure of the TP53 DBD was obtained from the PDB.

## 2.4 Patient Grouping for Survival Analysis

The TP53 STAD mutation dataset contained information from 289 patients, with 138 having TP53 mutations, and 151

patients with wild-type *TP53*. Of the 138, 80 patients had one missense mutation in *TP53*. To identify possible relationships between the location and the survival rate of the *TP53* missense mutations, we divided the samples into five categories: (1) gene overall status; (2) known hot spots; (3) domain location; (4) secondary structure; and (5) codon position (within the DBD region). First, these were divided into gene units- 80 patients with only one missense mutation in *TP53*, and 151 patients with wild-type *TP53*. The second category included well-known hot-spots (R175, G245, R248, R273, and R282) [9]. 32 patients had missense mutations in hot spots, while 48 had missense mutations in non-hot-spots. The third category was domain location. In the TCGA GC mutation dataset, we observed that most *TP53* missense mutations were in the DBD. Correspondingly, we mapped *TP53* mutations to the DBD or other regions. The fourth category was mutation-induced secondary structure of the DBD. Based on the DSSP secondary structure alignment extracted from the PDB, we divided the DBD into three protein substructures: (1) a helix region (3/10-helix, alpha-helix); (2) a strand region (beta strand, beta bridge); and (3) a turn region (turn, bend). We also referenced the clinical molecular profile in cBioPortal [8]. The fifth category, codon position, was grouped by the position of the mutation. As in the fourth category, we used samples from the DBD region to create a group. We chose the group of positions (R175, C176, R248, R273, R282) as those that had at least 5 mutation samples in the same position. For each mutation position, survival difference was compared to patients having another position mutations within DBD.

## 2.5 Kaplan-Meier Survival Analysis and Log-Rank Test

We performed a Kaplan-Meier survival analysis and a log-rank test, among the five categorized groups, to confirm the association of the location of the missense mutation with survival rate. In the first category, we carried out survival analysis and log-rank test between one missense mutation in *TP53* and *TP53* wild-type. In the second category, we performed Kaplan-Meier survival analysis and log-rank test between 32 samples with missense mutations at hot-spot locations (R175, G245, R248, R273, and R282), and 48 samples with missense mutations in other non-hot-spots. In the third category, we performed survival analysis and log-rank test between 76 samples with missense mutations in the DNA-binding domain, and 4 samples with missense mutations in other regions. In addition, we performed Kaplan-Meier survival analysis and log-rank test on wild-type samples. Fourthly, we compared various secondary structure groups (strand region-mutated patients vs. turn region-mutated patients; helix region-mutated patients vs. strand region-mutated patients; helix region-mutated patients vs. turn region-mutated patients; turn region-mutated patients vs. WT; helix region-mutated patients vs. WT; and strand region-mutated patients vs. WT), performing Kaplan-Meier survival analysis and log-rank test. In the fifth category (codon), we performed Kaplan-Meier survival analysis and log-rank test between each position (from R175, C176, R248, R273 and R282) and the other positions within DBD.

## 2.6 Pathway Analysis

We used a fifth category group to identify differentially expressed pathways, depending on the position of missense mutations within DBD. We used the RPKM value of the expression data (matched to the mutation data), and samples without expression data were excluded. We used the PATHOME algorithm to find differently expressed pathway contexts. The mutation position-associated pathways were visualized using Cytoscape software [10].

## 3 RESULTS

### 3.1 Survival Analyses of the *TP53* Gene Body and Mutation Hot-Spots

We compared survival between the 80 patients having missense mutations, in the *TP53* whole gene body, and 151 patients in the *TP53* wild-type (WT) group. Without regard to positions of the mutations, the 80 patients were categorized as the *TP53* mutant group. Kaplan-Meier survival analysis and log-rank test, between the two groups, revealed no significant survival difference between *TP53* missense mutation and *TP53* WT patients (log rank  $p = 0.81$ ) (Fig. 2a).

Next, we inspected survival associated with mutations in *TP53* hot spots. Since we had assigned the 80 mutations into either *TP53* known hot spot, or non-hot spot, groups (Fig. 2b), we examined possible clinical differences between the two groups. 32 of the 80 GC tumors were assigned to known hot spot (R175, G245, R248, R273, and R282) [9], and 48 to non-hot spot, groups. There was no significant correlation between the two groups in terms of survival (log rank  $p = 0.71$ ). This is contradictory to a previous result [9] that indicated clinical significance of *TP53* hot-spot mutations, possibly resulting from different GC clinico-pathological characteristics.

### 3.2 Survival Analyses of Missense Mutations within the *TP53* DNA-Binding Domain vs. Other Regions

Because 95 percent of the 80 missense mutations were in the *TP53* DNA-binding domain (DBD), we next assessed survival differences between DBD-mutated GC patients, vs. GC WT patients or those with other (i.e., non-DBD) mutations. In the first comparison, DBD-mutated patients ( $n = 76$ ) were considered as one group, and the 151 *TP53* WT patients as another group. Comparing survival between the two groups showed no significant survival differences (log rank  $p = 0.77$ ), using Kaplan-Meier survival analysis and log-rank test (Fig. 2c). In the second analysis, we compared DBD-mutated GC patients ( $n = 76$ ) to patients ( $n = 4$ ) with mutations elsewhere. That study showed no significant correlation between survival rate, and missense mutations, of 76 samples with missense mutations in the DBD, compared to the 4 with non-DBD mutations (Fig. 2d), although we concede this low-numbered sample size.

### 3.3 Survival Differences of GC Patients with Mutations, According to Secondary Structures, in the DBD

*TP53*'s secondary structure possesses three major types of known secondary structures: helix, strand, and turn. A helix is a spiral formation, a strand is part of a sheet-like

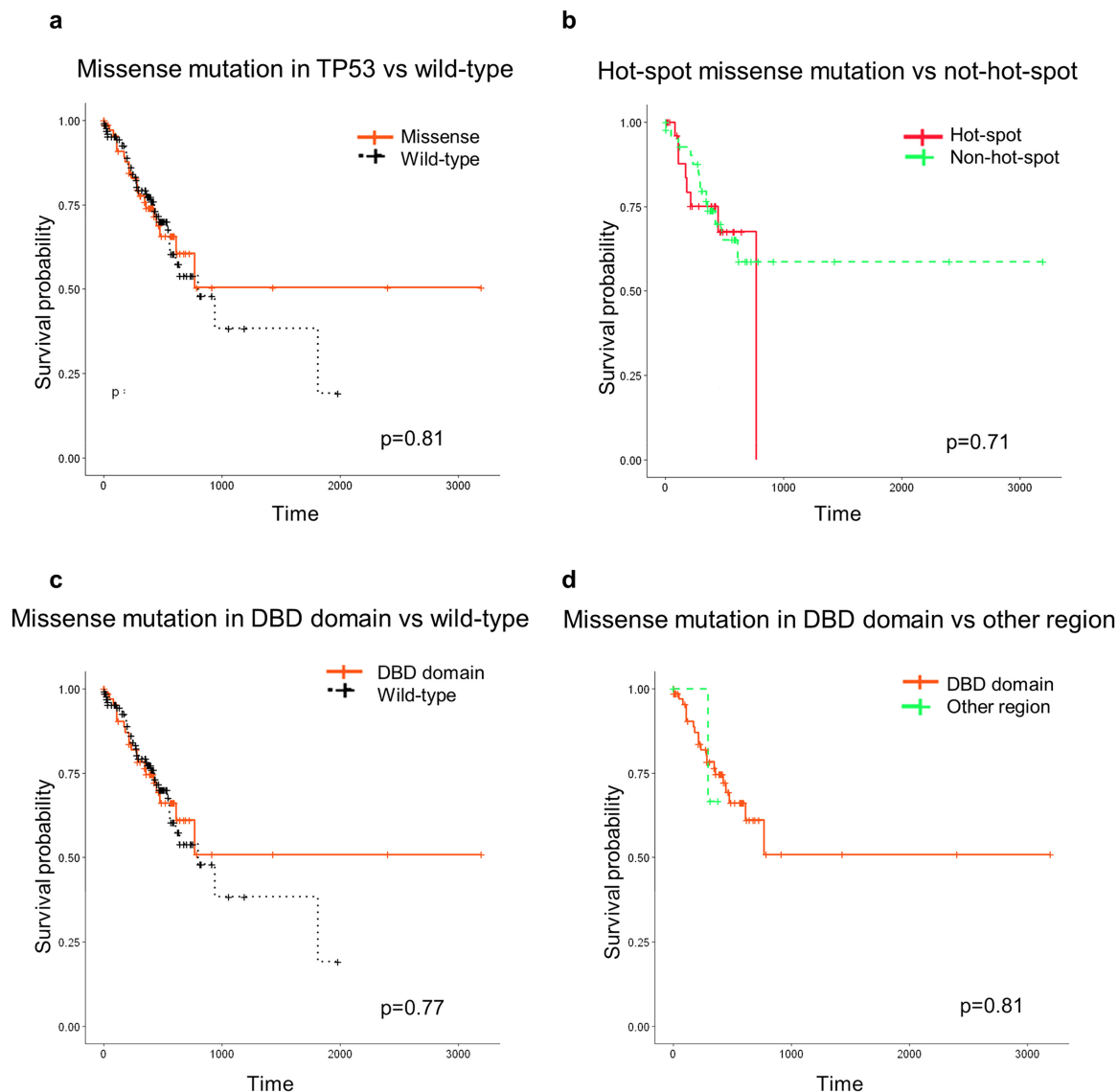


Fig. 2. Survival analyses according to TP53 whole gene body, mutation hot-spots, and domains. (a) Survival between the 80 patients having missense mutations in TP53 (whole gene body) with 151 TP53 wild-type (WT) patients. (b) Survival between patients of TP53 known hot spot mutations and those of non-hot spot TP53 mutations. (c) Survival between patients having mutations in DBD domain and TP53 WT patients. (d) Survival between DBD domain mutated patients and the other mutated regions. The y-axis indicates survival probability and the x-axis time (days).

structure, while a turn is a region where the polypeptide turns in the opposite direction [11].

Consequently, we inspected how 76 GC patients' mis-sense mutations, within the TP53 DBD, would affect TP53 secondary structure. 11 of the 76 were within the helix, 26 samples were in strands, 15 samples were in turns, and for 24 patients, the secondary structure could not be determined (Fig. 3a). We then compared survival differences between patient groups, according to these different secondary structures.

Survival analysis comparisons between the 15 turn region-mutated patients and the 26 strand region-mutated patients showed a significant difference (log rank  $p = 0.043$ ) (Fig. 3b). However, survival analysis between the helix region- vs. strand region-mutated patients showed no significant difference (log rank  $p = 0.074$ ) (Fig. 3c). Likewise, survival analysis comparing TP53-mutated helix region GC patients and TP53-mutated turn region patients showed no significant difference in survival (log rank  $p = 0.70$ ) (Fig. 3d). Similarly, survival analysis between the turn

region-mutated vs. WT-TP53 ( $n = 151$ ) patients showed no significant difference (log rank  $p = 0.23$ ) (Fig. 3e), nor did strand region-mutated patients vs. WT TP53 patients (log rank  $p = 0.17$ ) (Fig. 3g), nor when comparing helix region-mutants vs. WT TP53 (log rank  $p = 0.49$ ) (Fig. 3f).

We also inspected clinico-molecular profiles according to secondary structure (Supplementary Fig. S1, which can be found on the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TCBB.2018.2814049>), showing that two groups (turn region-mutated TP53 vs. strand region-mutated TP53) differed in tumor/node/metastasis (TNM)-stage to some extent (Supplementary Fig. S1 available online). Clinico-molecular comparison between strand region-mutated- and helix region-mutated-TP53 showed a visually slight difference between the two groups (Supplementary Fig. S1 available online). However, we note that the number of samples was small, and no further statistical tests could be performed (Supplementary Fig. S1 available online).



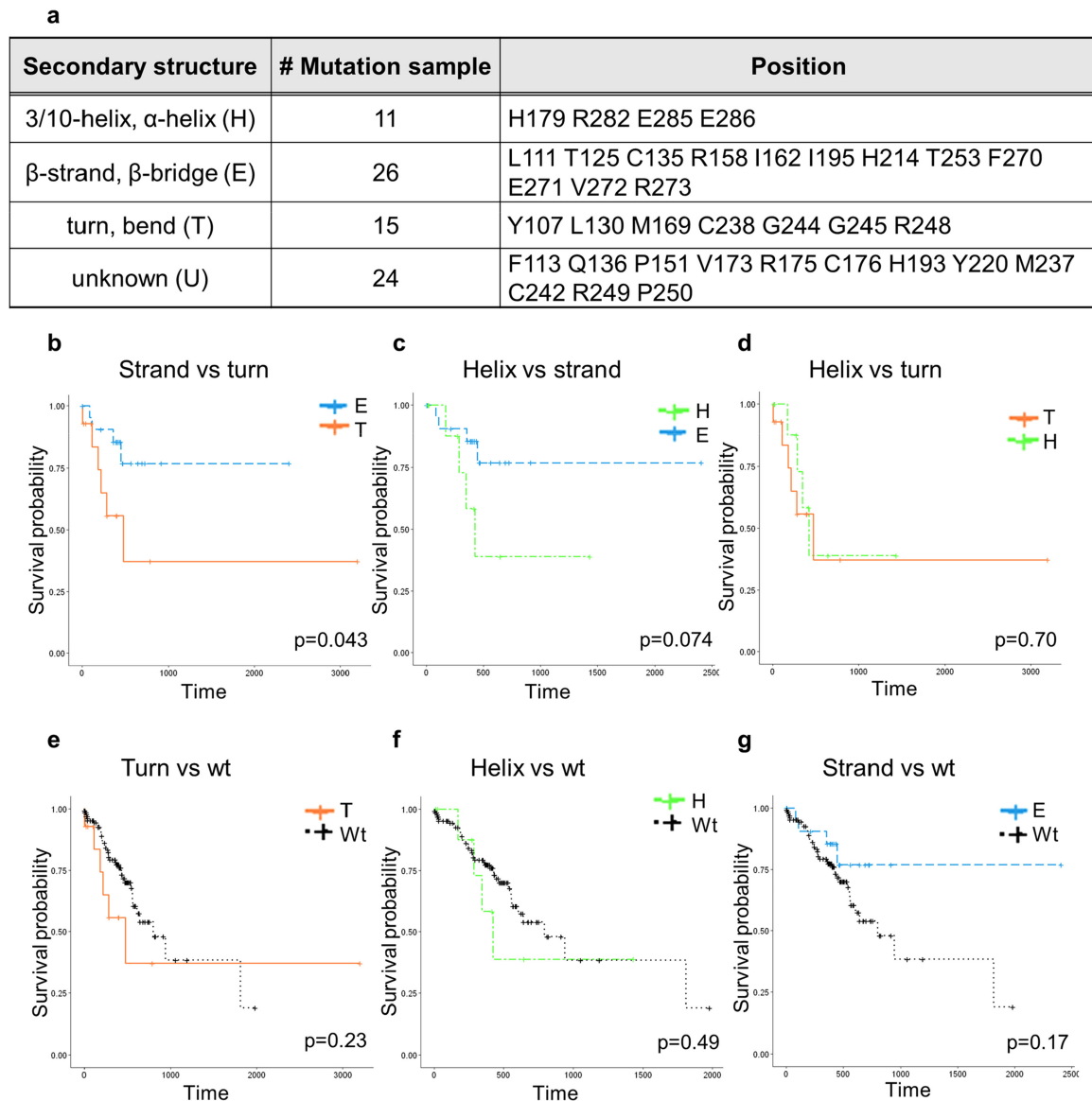


Fig. 3. Survival differences of mutations according to secondary structures, in the DNA-binding domain (DBD). (a) Classification according to secondary structure of mutation patients in the DBD. The single characteristics in parenthesis of the first column are used in the group notation of the following survival plots. (b) Survival between the turn region-mutated patients and the strand region-mutated patients. (c) Survival between the helix region-mutated patients and the strand region-mutated patients. (d) Survival between the helix region-mutated patients and the turn region-mutated patients. (e) Survival between the turn region-mutated patients and the WT patients. (f) Survival between the helix region-mutated patients and WT patients. (g) Survival between the strand region-mutated and WT patients. The y-axis indicates survival probability and the x-axis time (days).

We next compared *TP53* gene expression, according to secondary structure groups (e.g., turn-mutated-, helix-mutated-, and strand-mutated patient groups), vs. wild-type *TP53* GC patients. That comparison showed no statistically significant expression level difference between *TP53* WT patients and secondary structure-mutated patient groups (Supplementary Fig. S2 available online).

### 3.4 Survival Difference of Mutations, According to Mutational Positions within the DBD and Mutation Position-Associated Pathways

In the 76 *TP53*-DBD region-mutated GC patients, we performed survival analysis between patients having mutations in a specific position, vs. those having mutations elsewhere within the DBD. To secure a sufficient number of samples for statistics, we considered a mutation position that had at least five samples, allowing us to assess five

well-known mutation positions (R175, C176, R248, R273, and R282) within the DBD. Survival differences between R248-mutated samples ( $n = 6$ ) and patients' *TP53*-mutated in other codons within the DBD ( $n = 70$ ) showed statistically significant differences (log rank  $p = 0.035$ ) (Fig. 4a), while *TP53*-R175-mutated patients ( $n = 7$ ) vs. those mutated in other DBD positions ( $n = 69$ ) showed no statistically significant differences (log rank  $p = 0.25$ ). Moreover, analysis of *TP53*-C176 ( $n = 5$ ) vs. other *TP53* DBD-mutated regions ( $n = 71$ ) were not different (log rank  $p = 0.79$ ), nor were *TP53*-R273 mutants ( $n = 11$ ) vs. other DBD-mutated regions ( $n = 65$ ) (log rank  $p = 0.72$ ). Finally, survival analysis between *TP53*-R282-mutated GC patients ( $n = 5$ ), and *TP53*-other DBD-mutants ( $n = 71$ ) showed no significance (log rank  $p = 0.93$ ).

We further investigated signal pathways likely affected by the five mutational positions (R175, C176, R248, R273,

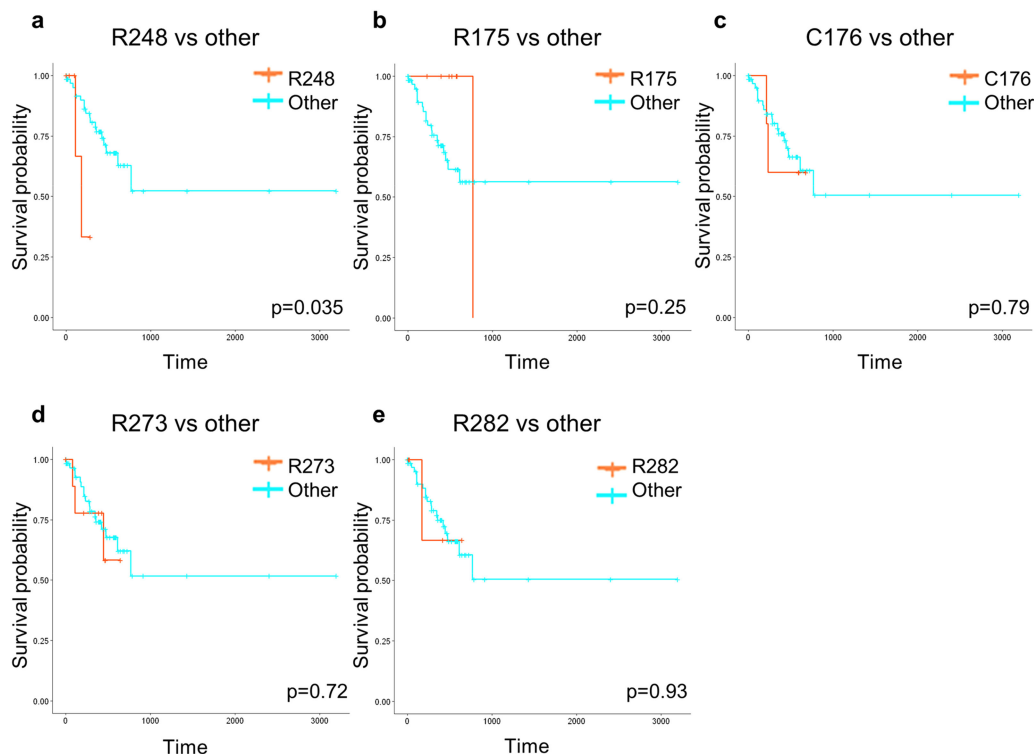


Fig. 4. Survival differences according to mutation positions within the DBD. (a) Survival between R248-mutated patients and other mutated positions. (b) Survival between R175-mutated patients and other mutated positions. (c) Survival between C176-mutated patients and other mutated positions. (d) Survival between R273-mutated patients and other mutated positions. (e) Survival between R282-mutated patients and other mutated positions. The y-axis indicates survival probability and the x-axis time (days).

R282) in *TP53* DBD [12]. For each position, we performed two pathway analyses, using our previously developed PATHOME method [13], and the matched TCGA expression dataset [5]. In the first pathway analysis, for each position (R175, C176, R248, R273, and R282), two patient groups were defined: one having *TP53* mutations in the position of interest, and the other group having *TP53* mutations in other positions within the DBD. Matched expression data for the two groups was used to identify pathway contexts by our PATHOME algorithm. In the second pathway analysis, for the position of interest, one group had the same mutational position as in the first analysis, compared to *TP53* WT GC patients. Pathway contexts were identified by PATHOME, based on expression data for the two groups. For the position of interest, we intersected pathway contexts from the two pathway analyses, to select both significantly changed pathway transcriptomal contexts, as compared to other *TP53*-mutated positions, as well to WT patients. As shown in Fig. 5, pathways associated with R248 (melanoma, non-small cell lung cancer) and those associated with R282 (thyroid cancer, colorectal cancer, and chronic myeloid leukemia) were “hubs” having pathway connections to a variety of cancer contexts. By contrast, R175 (apoptosis, cell cycle) and R273 (axon guidance, neurotrophin) were the least associated with malignant processes.

## 4 DISCUSSIONS

The tumor suppressor p53 (the “guardian of the genome,” and “policeman of oncogenes”) [1] is the most frequently mutated gene across all cancers [12]. Previously, it was shown that the “mutational landscape,” of 3,281 tumors, of 12 cancer

types, possessed *TP53*-driven clusters in breast, head and neck, and ovarian cancers [2]. In this study, we set forth to assess the effects of specific, well-known missense *TP53* mutations in gastric cancer (GC). Missense mutations, which include “point” mutations of a single nucleotide, alter a single codon (i.e., a single amino acid), in contrast to “indels” (insertions/deletions) that are functionally catastrophic.

In our previous work, we found biomarker candidates that combined with *TP53* mutation, in GC tumor samples, increasing the sensitivity and specificity over either biomarker alone [4]. However, the current study did not consider the transcriptomal context of *TP53* mutation, even while this was found to be important (Fig. 5), similar to our previous study [4], and the work of others, as mentioned above [2]. When comparing subsets of *TP53* mutant GC patients, we found that “hot-spot,” *TP53*-mutated patients showed no survival difference. This finding strongly agrees with the long-held “dominant-negative” model of normal DNA-binding, by nonfunctional *TP53* mutants, competing with wild-type *TP53* for binding to its target sequences [12].

As mentioned above, frameshift indels and nonsense mutations (premature stop codons) result in entire protein loss or production of scrambled proteins, generating totally different (and largely unknown) 3-dimensional structures. Thus, functional domains cannot be identified. Missense mutations, by contrast, keep protein structures intact, enabling us to infer their effect on 3-dimensional *TP53* structures. Consequently, our purpose in this study was to inspect how individual amino acid changes, in the intact structures, differentially affect survival.

With these considerations, we explored possible disruptions in the *TP53* protein structure. Using secondary

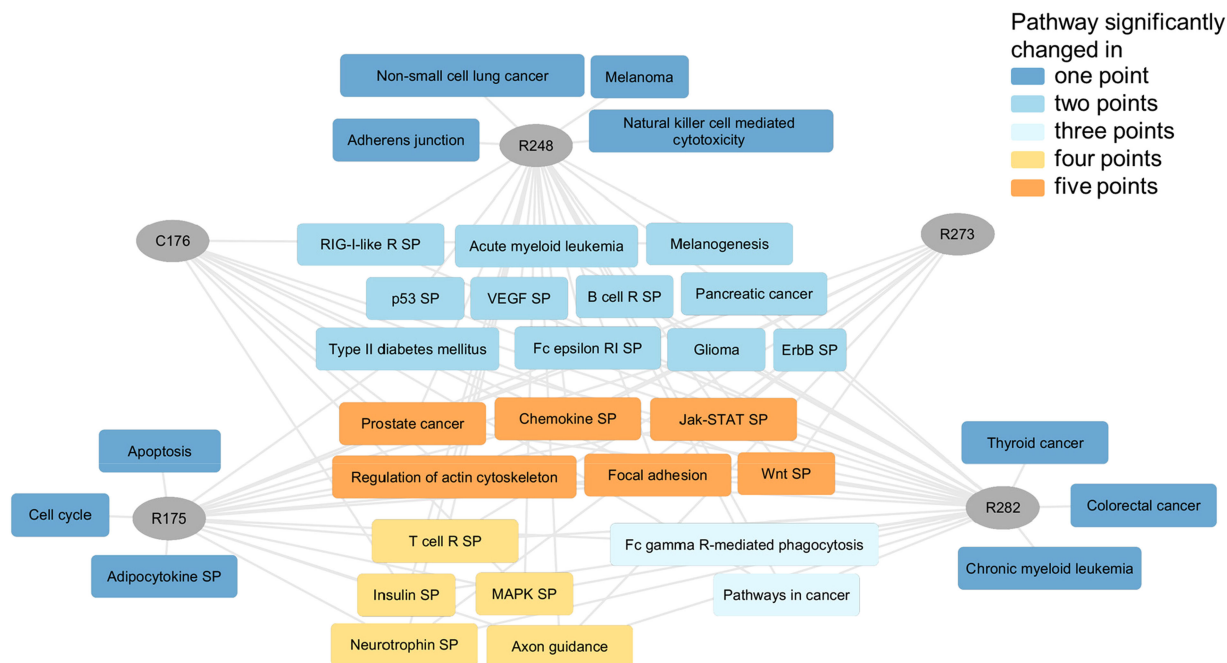


Fig. 5. Visualized mutation position-associated pathways. Here, we performed the intersection of pathway contexts from two pathway analyses, for selecting both significantly changed pathway contexts, compared to other positions mutated patients, in addition to WT patients (see details in main text). We also visualized the intersected pathway contexts in both pathway analysis. The “one point” (in upper right) indicates unique pathways in one location. The “two points” indicates two mutated locations shares common pathways. So, the figure shows a location-specific pathways and common pathways. (R:receptor; SP: signalling pathway).

structure, we found significantly increased survival in patients with beta-strand mutations, compared to other structural mutations. This might likewise be explained by normal (dominant-negative) DNA binding of the mutant protein, as the beta-strand is located outside the DNA-binding domain [14]. Further, we found wild-type *TP53* to correlate with clinical parameters, as compared to *TP53* mutants, including molecular subtype and copy number clusters (Supplementary Fig. S1 available online). Overall, survival of *TP53* missense mutation patients was not different than *TP53* wild type patients. Moreover, the various subsets of *TP53* missense mutations seemed to be irrelevant, except for *TP53* R248 mutations.

In a recent cancer study [15], truncating mutations and clonality were considered as critical to the evolution of blood cancer, and here we examined truncating mutations and clonality of GC, as follows.

50 samples from the TCGA GC dataset had *TP53* truncating mutations, including nonsense mutations, frameshift indels, and splicing site mutations. Comparison of survival differences between the 50 functionally abrogated *TP53* samples, and the *TP53* wild type samples, demonstrated no statistical difference (Supplementary Fig. S3 available online).

We also considered “double-hit” events for *TP53* alleles in GC patients having one allele with a truncating mutation, and the other allele with copy number deletion (Supplementary Fig. S4 available online). However, survival comparison of 36 patients having double-hit events, and *TP53* wild type patients, showed no survival difference (Supplementary Fig. S5 available online).

In another recent molecular classification of a TCGA GC dataset, copy number alteration-dominant GC cases corresponded with chromosomal instability (CIN), and mutation-dominance with microsatellite instability (MSI) [5]. Also,

MSI-type GC patients were reported to be hypermutated [5]. Selecting patients from these two GC types from the TCGA, we again compared survival, showing no significance difference (Supplementary Fig. S6 available online).

For considering confounding factors, including age and gender, in survival analysis, the Cox proportional hazard model was used. Here, we compared *TP53* missense mutation patients with *TP53* wild type patients, adjusting for age and gender. Again, however, *TP53* missense mutational status did not statistically associate with survival (p-value: 0.375, Supplementary Table ST1 available online).

As shown in Fig. 4, we performed five individual tests, to determine the FDRs (false discovery rates) for the five log-rank tests (R175, C176, R248, R273, and R282). That result, in comparing the FDR q-value for *TP53* R248 missense mutation versus different mutational positions (in Fig. 4a) was 0.175. Consequently, the results shown in Fig. 4a should be carefully interpreted.

As *TP53* is a transcription factor that binds DNA to regulate expression of its target genes [16], DNA contact occurs at R248 (Supplementary Fig. S7 available online). In other words, DNA-contact R248 mutant proteins keep the overall structure of the DBD intact, but cannot contact DNA. Correspondingly, such loss of contact may prevent DNA binding [16].

Gene-set enrichment analysis (GSEA), plus the Molecular Signatures Database (MSigDB) analysis [17], was performed for *TP53* hotspot mutation patients versus wild type patients (Supplementary Table ST2 available online). Those assays showed that two hotspot mutations (R248, R273) demonstrated differential pathways, when compared to wild type patients and other mutation patients (Supplementary Table ST2 available online).

To inspect clonality for six patients having *TP53* R248 missense mutations (from the TCGA GC dataset), we used



the ABSOLUTE method [15]. Following the Landau et al [18] criteria for clonality we showed that in three of the six patients, the *TP53* R248 missense mutation was clonal (Supplementary Table ST3 available online).

In the TCGA colorectal cancer dataset [19], we compared *TP53* R248 missense mutation patients with *TP53* wild type patients in terms of survival. No significant survival difference was observed (Supplementary Fig. S8 available online), indicating that the clinical relevance (e.g., survival) of R248 missense mutations depends on cancer context.

This study has some limitations; the most severe being sample size. Also, the significance levels we reported are marginal, and the results should be carefully interpreted. Nonetheless, as more and more genomic and transcriptomic datasets become publically available, statistical power will increase. Moreover, we did not study epigenetics of the *TP53* gene (nor its associated genes). This would also be of strong interest, as several *TP53* inducers (WAF1, ARF) are well known to be epigenetically silenced in gastric and other cancers [20]. Also, the *TP53* protein undergoes extensive post-translational modifications, including acetylation, phosphorylation, and ubiquitination; these should also be considered in the function of the wild-type or mutant proteins [21]. In summary, we believe that such characterizations could reveal prognostic biomarkers, and increase understanding of the etiology of diseases that lack function of this “master” tumor suppressor.

## 5 CONCLUSIONS

This study comprehensively characterized the status of the *TP53* tumor suppressor, in gastric cancer datasets. Significant findings were that: (1) overall *TP53* status (i.e., wild-type vs. mutated), in gastric cancer, had no effect on overall survival; (2) mutations in the DNA-binding (DBD) domain were not statistically significantly different from those outside the DBD, in terms of survival; (3) secondary structure disruption in the *TP53* helix or turn was much more deleterious than perturbation to a beta strand; and (4) patients mutated at R248 showed poorer survival than patients with mutations at different positions. In conclusion, we believe that our approach used here could be valuable for studying the consequences and phenotypes of specific genetic events and increase our understanding of physiologic responses involving master transcription regulators, such as *TP53*.

## ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1B03933145 to S.N.; NRF-2017R1D1A1B 03032077 to S.P.).

## REFERENCES

- B. Leroy, M. Anderson, and T. Soussi, “TP53 mutations in human cancer: Database reassessment and prospects for the next decade,” *Human Mutation*, vol. 35, no. 6, pp. 672–688, Jun. 2014.
- C. Kandath, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, M. D. Leiserson, C. A. Miller, J. S. Welch, M. J. Walter, M. C. Wendt, T. J. Ley, R. K. Wilson, B. J. Raphael, and L. Ding, “Mutational landscape and significance across 12 major cancer types,” *Nature*, vol. 502, no. 7471, pp. 333–339, Oct. 17, 2013.
- A. Efeyan and M. Serrano, “p53: Guardian of the genome and policeman of the oncogenes,” *Cell Cycle*, vol. 6, no. 9, pp. 1006–1010, May 02, 2007.
- S. Park, J. Lee, Y. H. Kim, J. Park, J. W. Shin, and S. Nam, “Clinical relevance and molecular phenotypes in gastric cancer, of TP53 mutations and gene expressions, in combination with other gene mutations,” *Sci. Rep.*, vol. 6, Oct. 06, 2016, pp. 34822.
- Cancer Genome Atlas Research Network, “Comprehensive molecular characterization of gastric adenocarcinoma,” *Nature*, vol. 513, no. 7517, pp. 202–209, Sep. 11, 2014.
- K. Cibulskis, M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E. S. Lander, and G. Getz, “Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples,” *Nat. Biotechnol.*, vol. 31, no. 3, pp. 213–219, Mar. 2013.
- P. W. Rose, A. Prlic, A. Altunkaya, C. Bi, A. R. Bradley, C. H. Christie, L. D. Costanzo, J. M. Duarte, S. Dutta, Z. Feng, R. K. Green, D. S. Goodsell, B. Hudson, T. Kalro, R. Lowe, E. Peisach, C. Randle, A. S. Rose, C. Shao, Y. P. Tao, Y. Valasatava, M. Voigt, J. D. Westbrook, J. Woo, H. Yang, J. Y. Young, C. Zardecki, H. M. Berman, and S. K. Burley, “The RCSB protein data bank: Integrative view of protein, gene and 3D structural information,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D271–D281, Jan. 04, 2017.
- J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, and N. Schultz, “Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal,” *Sci. Signal*, vol. 6, no. 269, Apr. 02, 2013, Art. no. p11.
- T. Tahara, T. Shibata, Y. Okamoto, J. Yamazaki, T. Kawamura, N. Horiguchi, M. Okubo, N. Nakano, T. Ishizuka, M. Nagasaka, Y. Nakagawa, and N. Ohmiya, “Mutation spectrum of TP53 gene predicts clinicopathological features and survival of gastric cancer,” *Oncotarget*, vol. 7, no. 27, pp. 42252–42260, Jul. 05, 2016.
- K. Ono, B. Demchak, and T. Ideker, “Cytoscape tools for the web age: D3.js and Cytoscape.js exporters,” *F1000 Res.*, vol. 3, 2014, Art. no. 143.
- G. E. Schulz and R. H. Schirmer, *Principles of Protein Structure*, pp. 66–94, New York, NY, USA: Springer, 1979.
- P. A. Muller and K. H. Vousden, “p53 mutations in cancer,” *Nat. Cell Biol.*, vol. 15, no. 1, pp. 2–8, Jan. 2013.
- S. Nam, H. R. Chang, K. T. Kim, M. C. Kook, D. Hong, C. H. Kwon, H. R. Jung, H. S. Park, G. Powis, H. Liang, T. Park, and Y. H. Kim, “PATHOME: An algorithm for accurately detecting differentially expressed subpathways,” *Oncogene*, vol. 33, no. 41, pp. 4941–4951, Oct. 09, 2014.
- B. Kamaraj and A. Bogaerts, “Structure and function of p53-DNA complexes with inactivation and rescue mutations: A molecular dynamics simulation study,” *PLoS One*, vol. 10, no. 8, 2015, Art. no. e0134638.
- S. L. Carter, K. Cibulskis, E. Helman, A. McKenna, H. Shen, T. Zack, P. W. Laird, R. C. Onofrio, W. Winckler, B. A. Weir, R. Beroukhi, D. Pellman, D. A. Levine, E. S. Lander, M. Meyerson, and G. Getz, “Absolute quantification of somatic DNA alterations in human cancer,” *Nat. Biotechnol.*, vol. 30, no. 5, pp. 413–421, May 2012.
- M. Olivier, M. Hollstein, and P. Hainaut, “TP53 mutations in human cancers: origins, consequences, and clinical use,” *Cold Spring Harb Perspect Biol.*, vol. 2, no. 1, Jan. 2010, pp. a001008.
- A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, “Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles,” *Proc. Natl. Academy Sci. USA*, vol. 102, no. 43, pp. 15545–15550, Oct. 25, 2005.
- D. A. Landau, S. L. Carter, P. Stojanov, A. McKenna, K. Stevenson, M. S. Lawrence, C. Sougnez, C. Stewart, A. Sivachenko, L. Wang, Y. Wan, W. Zhang, S. A. Shukla, A. Vartanov, S. M. Fernandes, G. Saksena, K. Cibulskis, B. Tesar, S. Gabriel, N. Hacohen, M. Meyerson, E. S. Lander, D. Neuberger, J. R. Brown, G. Getz, and C. J. Wu, “Evolution and impact of subclonal mutations in chronic lymphocytic leukemia,” *Cell*, vol. 152, no. 4, pp. 714–726, Feb. 14, 2013.
- Cancer Genome Atlas Research Network, “Comprehensive molecular characterization of human colon and rectal cancer,” *Nature*, vol. 487, no. 7407, pp. 330–337, Jul. 18, 2012.
- Y. Qu, S. Dang and P. Hou, “Gene methylation in gastric cancer,” *Clin. Chim. Acta*, vol. 424, pp. 53–65, Sep. 23, 2013.



- [21] Y. Xu, "Regulation of p53 responses by post-translational modifications," *Cell Death Differ.*, vol. 10, no. 4, pp. 400–403, Apr. 2003.



**SeongRyeol Moon** received the BSc degree in biology from Kyunghee University, South Korea, in 2009. Currently, he is working toward the MS degree in the Department of Health Sciences and Technology, Gachon Advanced Institute for Health Sciences and Technology (GAIHST), Gachon University, Incheon, South Korea. His current research interests include network analysis, machine-learning application in biological big data, and next-generation sequencing (NGS) analysis in clinical samples.



**Curt Balch** received the PhD degree and postdoctoral training from the University of Cincinnati, prior to eight years of service as a research faculty member with the Indiana University School of Medicine. There, he performed extensive research in cancer epigenomics and subsequently, in mid-2013, founded a biomedical consulting company, Bioscience Advising. Now, he is a member of the Complex Biological Systems Alliance, North Andover, MA. His interests include not only translational research but also systems biology/bioinformatics

approaches to integrate the epigenomic, transcriptomic, and genomic anomalies that govern specific phenotypes (e.g., the "epithelial-to-mesenchymal transition," etc.), allowing the identification of specific "regulatory networks" that control tumor progression.



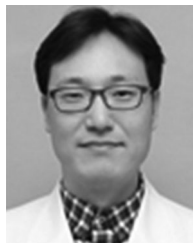
**Sungjin Park** received the PhD degree in biomedical engineering from Inje University, in 2016. Currently, he is a research professor in the College of Medicine Gachon University, South Korea. His research interests include computing effect of mutational interaction, cellular compartments interaction, and epistasis in cancer systems biology.



**Jinhyuk Lee** received the PhD degree in chemistry from Seoul National University, in 2004. He experienced three post-doctoral associates from 2004 to 2010 at the Scripps Research Institute (TRSI) in San Diego, the University of Kansas in the Department of Bioinformatics, and the Korea Institute of Advanced Study (KIAS). After the successful achievements during his post-doctoral associates, he was given a regular position at the Genome Editing Research Center at the Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, South Korea, and also taught graduate students in the University of Science and Technology (UST) as an associate professor. He is studying in silico protein-ligand interaction and developing a new computational methodologies for drug discovery based on Big Data.



**Jiyong Sung** received the BSc degree in physics from Yonsei University, South Korea, in 2008. Currently, she is a researcher with Gachon University Gil Medical Center, Incheon, Korea. Her interests include systems biology application in metabolism.



**Seungyoon Nam** received the PhD degree in bioinformatics from Seoul National University, in 2008. From 2008 to 2010, he was a postdoctoral fellow with the Indiana University School of Medicine, Indiana, performing computational cancer epigenomics studies. From 2010 to 2015, he worked as a senior researcher in a field of cancer bioinformatics at the Korea Institute of Information and Science Technology, and the National Cancer Center of Korea. Now, he is an assistant professor in the College of Medicine, Gachon

University, Korea. His research interests include systems biology, miRNA biology, Next-Generation Sequencing (NGS) clinical tests, and druggable genome in various diseases. He has served as a member of the program committee at the IEEE International Conference on Bioinformatics & Biomedicine, Workshop on Data Mining from Genomic Rare Variants and Its Application to Genome-Wide Analysis since 2014.

▷ For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).