In silico Drug Repositioning of bortezomib to reverse metastatic effect of *GALNT14* in lung cancer

Running title**: A data-driven route to drug discovery for undruggable targets**

Ok-Seon Kwon[1*], Haeseung Lee[2*], Hyeon-Joon Kong[3], Ji Eun Park[1], Wooin Lee[1], Seungmin Kang[2], Mirang Kim[4], Wankyu Kim[2#], Hyuk-Jin Cha[1#]

[1] College of Pharmacy, Seoul National University, Seoul 08826, Republic of Korea [2] Research Center for Systems Biology, Department of Life Sciences, Ewha Womans University, Seoul 03760, Republic of Korea [3] College of Natural Sciences, Department of Life Sciences, Sogang University, Seoul 04107, Republic of Korea SD [4]Personalized Genomic Medicine Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Republic of Korea

[*] These authors contributed equally to this work.

[#]Corresponding authors: hjcha93@snu.ac.kr and wkim@ewha.ac.kr

Hyuk-Jin Cha, PhD

College of Pharmacy, Department of Pharmacy, Seoul National University

1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

Tel.: [+]82-2-880-7825; Fax: [+]82-2-880-9122; E-mail: hjcha93@snu.ac.kr

Wan-Kyu Kim, Ph.D

Ewha Research Center for Systems Biology, Department of Life Sciences, Ewha Womans University

52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760 Republic of Korea

Tel: +82-2-3277-4132, Fax: +82-2-3277-6809, Email: wkim@ewha.ac.kr

Conflicts of interest: The author declares that he has no conflict of interest.

Abstract

Although many molecular targets for cancer therapy have been discovered, they often show poor druggability, which is a major obstacle to develop targeted drugs. As an alternative route to drug discovery, we adopted an *in silico* drug repositioning (*in silico* DR) approach based on large-scale gene expression signatures, with the goal of identifying inhibitors of lung cancer metastasis. Our analysis of clinoco-genomic data identified GALNT14, an enzyme involved in O-linked N-acetyl galactosamine glycosylation, as a putative driver of lung cancer metastasis leading to poor survival. To overcome the poor druggability of GALNT14, we leveraged Connectivity Map approach, an *in silico* screening for drugs that are likely to revert the metastatic expression patterns. It leads to identification of bortezomib (BTZ) as a potent metastatic inhibitor, bypassing direct inhibition of poorly druggable target, GALNT14. The anti-metastatic effect of BTZ was verified *in vitro* and *in vivo*. Notably, both BTZ treatment and *GALNT14* knockdown attenuated TGFβ-mediated gene expression and suppressed TGFβ-dependent metastatic genes, suggesting that BTZ acts by modulating TGFβ signaling. Taken together, these results demonstrate that our *in silico* DR approach is a viable strategy to identify a candidate drug for undruggable targets, and to uncover its underlying mechanisms.

Keywords: GALNT14, connectivity map, drug repositioning, bortezomib, TGFβ, metastasis, undruggable targets

Introduction

In the context of personalized anti-cancer therapy based on targeting specific proteins with the goal of lowering cancer-related mortality (1), a great deal of effort has been devoted to identifying both molecular targets and accompanying drugs (2,3). However, the fraction of patients eligible for personalized anti-cancer therapy is very limited (4) due to the poor druggability of newly identified molecular targets, notwithstanding recent advances in strategies in drugging 'undruggable' proteins (5,6).

An alternative approach to matching candidate drugs to poorly druggable cancer targets is *in silico* drug repositioning (*in silico* DR) (7,8). Due to the well-characterized pharmacology and safety of approved drug libraries(9), this approach has the potential to reduce cost and attrition during the clinical phases of drug development. Several approaches to DR have been tested in the context of oncology (10) and a few of the resultant drugs, including celecoxib and thalidomide, have been approved by the FDA for repurposing as anti-cancer therapeutics (11). Along with recent advances in sequencing technologies, chemogenomics databases containing drug-induced gene expression profiles provide clues regarding potential treatments for personalized cancer targets, and can also suggest candidate drugs based on tailored gene signatures of cancers upon identification of molecular targets (12). The Connectivity Map (CMap), a collection of genome-wide expression profiles of cell lines treated with >20,000 chemicals(13), has been used to identify candidate drugs for certain cancer types (14,15), (16).

N-acetyl-galactosaminyltransferases (GalNAc-Ts or GALNTs) are key enzymes that initiate O-linked N-acetyl galactosamine (GalNAc) glycosylation. This process is an important step in the synthesis of Thomsen-nouvelle (Tn) antigens, which

are well-characterized tumor-associated molecules (17). In particular, *GALNT14* has been examined in the context of apoptotic signaling (18,19); invasion and migration of breast (20,21), ovary (22), and lung (23) cancers; and multi-drug resistance of breast cancer cells (24). Moreover, *GALNT14* expression is not only a prognostic marker in neuroblastoma (25) and lung cancer (23), but also a predictive marker for Apo2L/TRAIL-based cancer therapy (18), although a randomized phase II study based on the predictive marker of GALNT14 for dulanermin did not improve patient outcome (26).

In this study, through transcriptome analysis of the TCGA dataset and *in vitro* and *in vivo* studies, we demonstrated that *GALNT14* is strongly associated with lung cancer recurrence due to the high migration and invasion properties of tumor cells. Rather than attempting direct inhibition of the poorly druggable GALNT14 protein or downstream signaling, we leveraged large-scale drug-induced transcriptome data to identify candidate drug(s) likely to reverse *GALNT14*-dependent gene expression, *i.e.*, drugs that led to transcriptomic changes similar to those induced by *GALNT14* depletion. We successfully identified an anti-metastatic candidate drug that mimicked *GALNT14* depletion. The results demonstrate that this approach represents a viable strategy for discovering candidate drugs for many other undruggable targets.

Materials and Methods

Cell line establishment and Cell culture

H460 cell line which was purchased from Korean cell line bank (KCLB) were maintained in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (FBS), gentamicin (50 μg/ml) at 37**°C** in a humidified atmosphere of 5% CO_2 in the air. GALNT14 knockdown cell lines with shRNA were established as previously described (23).

Reagents and antibodies

The primary antibodies against cleavage caspase- 3 (#9664), cleavage caspase-9 (#9505) and pSmad2 (# 310S) were obtained from Cell Signaling Technology. Antibodies against β-Actin (sc-47778), PARP (sc-7150), p53(sc-126), p27 (sc-528), p21 (sc-397), CyclinD1(sc-718) and CyclinB1 (sc-245) were obtained from Santa Cruz Biotechnology Inc. and β-catenin (BD 610153) was purchased from BD Biosciences pharmigen. Bortezomib (S1013) and Carfilzomib (S2853) were purchased from selleckchem.

RNA-sequencing and analysis

Total RNA was isolated using the Trizol according to the manufacture instruction. For library construction, we used the TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA). Briefly, the steps of strand-specific protocol are: first strand cDNA synthesis; second strand synthesized using dUTPs instead of dTTPs; end repair, A-tailing, and adaptor ligation; PCR amplification. Then, each library was diluted to 8 pM for 76 cycles of paired-read sequencing (2 X 75bp) on the Illumina NextSeq 500 per the manufacturer's recommended protocol.

TCGA data processing and analysis

TCGA lung adenocarcinoma (LUAD) cohort (n=576), containing mRNA gene expression and clinical data on 388 primary lung cancers, 128 lung cancers with recurrence, and 59 benign lung tissues were collected from the Broad GDAC Firehose (https://gdac.broadinstitute.org/). Total 494 patients with clinical information tracked for at least one month were used for survival analysis. For 20,531 genes, all patients were divided into high and low groups by the median expression of each gene and relapse-free survival analysis was performed according to the group difference. Hazard ratio and *P* value were calculated by Cox proportional hazards regression model and log-rank test respectively. With 58 patients who have gene expression profiles of normal benign tissues, differential expression in LUAD compared to matched normal samples were measured from the likelihood ratio test. RNA-seq profiles were normalized and processed using R package 'limma' and 'DESeq2' and survival analysis was conducted by R package 'survival'.

Metastasis and Tumorigenesis signatures

From MSigDB manually curated gene sets (C2), we collected the 35 metastasis- and 44 tumor-related gene sets that are annotated by 'metastasis' / 'epithelial-mesenchymal transition' and 'cancer'/ 'tumorigenesis', respectively. We took only Up-regulated genes when both UP- or DOWN-regulated sets were available. Then, we selected 32 and 23 genes as the metastatic and the tumorigenesis signature, respectively, by taking the consensus genes common to at least three or more gene sets.

Anti-metastatic drug prediction using CMap dataset

The updated CMap, or LINCS L1000(13), provides an extensive catalog of >1.2 million transcriptome signatures from 71 human cell lines in response to each of 20,413 small molecule treatments and are different from the original CMap based on the

microarray(27). This dataset contains multiple expression signatures even for the same drug and cell line depending on dose, sample time, and batch. Therefore, we developed our own DR method to handle such redundancy, which is distinct from the original CMap method based on rank-based Kolmogorov-Smirnov (KS) test. Differential gene expression matrix (level 5, replicate-collapsed z-score) and metadata were obtained from the Gene Expression Omnibus (acession id: GSE92742). Each column in the matrix represents an experiment consisting of chemical perturbation in a cell line, and each row represents a gene whose degree of differential expression was quantified in that experiment, respectively. After filtering out the experiments which consist of less than three replicates and low-quality (distil_cc_q75 < 0.2, and pct_self_rank_q25 > 5), the 100 most up- and down-regulated genes were selected as the expression signature of an experiment. Given a set of genes representing a specific phenotype, a series of gene set analysis toward gene signatures of all experiments was performed using *Jaccard index*. After calculating the similarities, multiple similarity scores of a compound are combined into a single score, drug-repositioning (DR) score. DR score is defined as the negative logarithm of the *P*-value of hypergeometric test that quantifies the degree of over-representation of experiments of a compound within the top 10% of similarity scores.

GALNT14-TGFβ signature

TGFβ downstream target genes (39 SMAD4 dependent genes and 65 SMAD4 independent genes) were collected from the literature (28). Among the SMAD4 dependent targets, a set of genes commonly down-regulated by both BTZ and sh*GAL* (*ATF3*, *ARNTL*, *COPA*, *NEDD9*, *LAMC2*, *RAB27B*, and *PCDH7*) was defined as '*GALNT14*–TGFβ signature'. To measure the average activity of the signature, KS

statistic was used to estimate the degree of up- or down-regulation of the seven genes in a sample's gene expression profile. In TCGA LUAD cohort, the patients with activity in the top 10% (n=50) or lower 10% (n=50) were classified as 'high' or 'low' respectively.

Statistical analysis

The graphical data were presented as mean ± S.E.M. Statistical significance among the three groups and between groups was determined using one-way or two-way analysis of variance (ANOVA) following Turkey post-test and Student's t-test respectively. Significance was assumed for $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

Results

GALNT14 as a putative molecular target for lung cancer metastasis

Identification of molecular targets in recurrent cancers is essential not only for predicting prognosis, but also for matching specific drug–target pairs if they are available. To identify potential molecular targets related to cancer recurrence, we assembled transcriptome data and clinical information from 516 lung cancer patients from the TCGA LUAD cohort (Fig. 1A). Concentrating on molecular targets relevant to recurrent lung cancer, we performed a series of relapse-free survival (RFS) analyses and differential expression analyses. Expression of seven genes (*GALNT14, COL7A1, GPR115, C1QTNF6, KRT16, INHA*, and *TNFSF11*) was significantly associated with cancer progression, recurrence (Fig. 1B), and overall survival (Fig. S1A), indicating that these genes are potentially valuable as predictors of poor prognosis. Notably, metastasis-related genes were significantly overrepresented in gene lists selected by both RFS ($P = 2.3 \times 10^{-6}$) and differential expression ($P = 5.2 \times 10^{-19}$) analysis, including two genes (*e.g., TNFSF11* and *INHA*) among the seven aforementioned candidates. To further confirm the relevance of each gene to metastasis or tumorigenesis, we divided lung cancer patients into two groups (low or high) using the median expression of each gene as the cutoff. Both metastasis and tumor signatures were positively enriched in the high-expression groups of all seven genes (Fig. 1C), and the significant enrichment was observed for *GALNT14* (Fig. 1D).

GALNT14, which encodes a glycosyltransferase involved in O-glycosylation, has been implicated in both tumor malignancy (25) and metastasis (20,21,23). As expected, metastatic (Fig. 1E) and tumorigenic potentials (Fig. 1F) were markedly attenuated by loss of *GALNT14*, indicating that the gene is important for metastasis as

well as tumorigenesis. Further, metastatic lung cancer cells were more vulnerable to *GALNT14* depletion than non-metastatic or other types of cancers in the *Project Achilles* dataset(29), a genome-scale RNAi screening data from for 501 cancer cell lines, including 126 cell lines originating from metastatic patients (Fig. 1G). Other candidate genes were less vulnerable in metastatic lung cancer (Fig. S1B). Consistent with this, *GALNT14* expression shows a clearly negative correlation with both locoregional recurrence-free survival (LRFS) and distant metastasis-free survival (DMFS) (Fig. 1H) as well as overall survival (Fig. S1A) in the TCGA LUAD cohorts. In addition, normal lung expresses only a low level of *GALNT14*, and there is a large gap between normal and lung cancer tissue (Fig. S1C)(30). All these results suggested *GALNT14* as a promising molecular target for lung cancer metastasis to improve patient survival.

Computational repositioning of BTZ to reverse the *GALNT14* expression signature

Although *GALNT14* may be a potential therapeutic target for metastatic lung cancer, GALNTs remain poorly druggable despite several attempts to find specific inhibitors (31,32). Notably in this regard, GALNT14-dependent metastatic potential is governed by induction of transcription factors (e.g. *HOXB9* or *SOX4*) rather than by altered glycosylation (20,23). Therefore, rather than inhibiting *GALNT14* directly, we leveraged the CMap dataset to virtually screen for drugs that mimic the effect of *GALNT14* depletion at the transcriptome level. To this end, we generated two gene signatures: (i) 73 genes down-regulated in the *GALNT14*-low vs *GALNT14*-high group in the TCGA LUAD cohort, and (ii) 129 genes down-regulated by *GALNT14* knockdown in the H460 cell line. Next, we collected a comprehensive list of (iii) 3711 metastasis-related genes. Finally, we refined the gene signatures (i, ii) to 20, and 49 genes as *GALNT14*

signatures by taking common genes with the genes of (iii). Accordingly, our subsequent predictions would prioritize drugs that may be relevant to both *GALNT14*-dependence and metastasis, but irrelevant genes tend to be filtered out.

Using the two *GALNT14* signatures, we performed two independent predictions by CMap analyses (Fig. 2A). Candidate drugs were prioritized according to their DR scores (See Materials and Methods in detail). Two drugs, dexamethasone (DEX, an anti-inflammatory corticosteroid) and bortezomib (BTZ, a first-in-class proteasome inhibitor used to treat multiple myeloma), were among the top ten candidates in both predictions. We then validated expression levels of *SOX4, AREG,* and *VCAN,* which are strongly associated with metastasis and regulated by *GALNT14* (20,23) (Figs. S2B and S2C). Among genes differentially expressed in response to either DEX or BTZ in the CMap dataset, *SOX4* (but not *AREG* or *VCAN*) was commonly altered, indicating that *SOX4* could serve as a validation marker (Fig. S2C). As predicted, we observed dose-dependent suppression of *SOX4* (but not *VCAN*) in H460 cells treated with either DEX or BTZ. Although BTZ suppressed *SOX4* less effectively than DEX (EC_{50}: 15 nM *vs.* 5 nM, respectively) (Fig. S2D), BTZ treatment led to a significant reduction in the migration capacity of H460 cells, whereas DEX did not (Fig. 2C). The EC_{50} for proteasome inhibition was around 20 nM (Fig. 2D and S2E), a concentration at which BTZ clearly inhibited migration (Figs. 2E and S2F) and invasion (Fig. 2F) by lung cancer cells, while it did not affect GALNT14 expression (Fig. S2G), cell viability (Figs. 2G and S2H) nor the cell proliferation (Figs. 2H and S2I). Of note, H358 with relatively lower GALNT14 expression than H1299 and H460 did not respond to BTZ treatment in migration (Fig. S2K) and invasion (Fig. S2L) unlike H1299, while proteasome inhibition by BTZ occurred (Fig. S2M), suggesting that anti-invasion/migration effect

of BTZ would be associated with GALNT14 expression.

The effect of BTZ is independent of proteasome inhibition

Given that the anti-migration/invasion effect of BTZ occurred at a concentration that also inhibited proteasome activity (Figs. 2D and E), we sought to determine whether this effect was a result of proteasome inhibition *per se*. To investigate this issue, we first compared the chemical structure of three FDA-approved proteasome inhibitors, BTZ, carfilzomib (CFZ), and ixazomib (IXZ), all of which are approved for treatment of multiple myeloma (33). CFZ (but not IXZ) was structurally similar to BTZ as calculated by Tanimoto coefficient or Jaccard index of molecular fingerprints(34) (Fig. 3A). In contrast to BTZ, the concentration of CFZ that inhibited proteasome activity (Fig. S3A) and stabilized well-characterized proteasome targets β-catenin, Cyclin D1, and p27 (Fig. 3B) failed to suppress migration (Fig. 3C) and invasion (Fig. 3D). Moreover, the boronic acid moiety responsible for the proteasome inhibition (35) was present in both BTZ and IXZ. However, like CFZ, treatment with IXZ could not inhibit migration (Fig. S3B). These data suggest that BTZ has an off-target effect that is independent of proteasome inhibition.

To confirm that the anti-migration/invasion effect of BTZ is not dependent on proteasome inhibition, we compared transcriptome profiles of H460 cells treated with BTZ, CFZ and depleted of *GALNT14* (sh*GAL*). Perturbation by BTZ and sh*GAL* (but not CFZ) induced similar transcriptomic changes relative to the control (Fig. 3E). Moreover, the expression patterns of metastatic signature genes were even more similar between BTZ and sh*GAL*, whereas CFZ had a minimal effect on only a few genes (Fig. 3E, middle panel). By contrast, expression of proteasome-related genes was altered

significantly by both drugs, but only marginally by sh*GAL* (Fig. 3E, right panel). Genes down-regulated by BTZ and sh*GAL* overlapped significantly (101 common genes; $P = 5.3 \times 10^{-54}$), suggesting that BTZ treatment partially mimics depletion of *GALNT14* (Fig. 3F). These results are consistent with the phenotypic outcomes, i.e., BTZ, but not CFZ, suppressed cell migration and invasion similarly to *GALNT14* depletion.

Attenuation of the TGFβ gene response by BTZ treatment or GALNT14 knockdown

We hypothesized that a subset of the 101 genes down-regulated by both BTZ treatment and *GALNT14* depletion could account for anti-migration/invasion effects of BTZ. To investigate the drug's mode of action, we conducted pathway enrichment analysis of the 101 genes and investigated the clinical significance of each pathway (Fig. 4A). Among the most enriched pathways was TGFβ signaling (hazard ratio [HR] = 1.2). BTZ treatment induced changes in expression of individual TGFβ signaling genes that were very similar to those induced by *shGAL* (Fig. S4A for BTZ and Fig. S4B for *shGAL*). Moreover, genes commonly down-regulated (e.g., *INHBA, FST,* and *BMPR*) among targets of TGFβ signaling were indeed suppressed (Fig. 4B). Suppression of the TGFβ-dependent gene signature, a common effect of BTZ treatment and *GALN14* depletion, was validated by reporter assays using the Smad-binding element (SBE), activin-response element (ARE), and BMP-response element (BRE) (Fig. 4C). Similarly, TGFβ reporter activity decreased after treatment with BTZ, but not CFZ (Fig. 4D), while β-catenin and Cyclin D1 were stabilized by proteasome inhibition following treatment with either BTZ or CFZ (Fig. 4E). Concomitant with the reduction in Smad2

phosphorylation with (Fig. 4E) or without TGFβ stimulation (Fig. S4C), SMAD4 nuclear translocation was inhibited significantly by BTZ treatment (Fig. 4F). Consistently, TGFβ dependent gene responses (determined by SBE or BRE) upon TGFβ stimulation, were significantly attenuated by BTZ treatment (Fig. S4D). As depletion of SMAD4 was sufficient to inhibit both migration and invasion (Figs. 4G and H), inhibition of SMAD2 phosphorylation (Figs. 4E and S4C) and subsequent delay of SMAD4 nuclear translocation (Fig. 4F) by BTZ would be a possible mode of action of BTZ. It is important to note that the TGFβ signaling pathway has been studied extensively as a tumor suppressor, a tumor promoter (36), and a promoter of metastasis (37). To determine whether the TGFβ signaling response is associated with *GALNT14*, we selected a set of TGFβ downstream targets (28) and examined their correlations with *GALNT14* expression and patient prognosis (Fig. 4I). Notably, SMAD4-dependent targets *PCDH7* and *LAMC2,* previously shown to induce metastasis (38,39) or tumorigenicity (40,41), were highly correlated with *GALNT14* (Fig. 4I). Moreover, the SMAD4-dependent TGFβ targets was associated to RFS in the *GALNT14*-high group (P = 0.045, Fig. 4J), suggesting that some SMAD4-dependent targets responsible for cancer recurrence are strongly associated with *GALNT14* expression. These results imply that BTZ treatment, like *GALNT14* depletion, exerts its anti-metastatic effect by interfering with nuclear translocation of SMAD4 (Fig. 4F) and with the SMAD4-dependent gene expression response (Fig. 4D). Finally, we defined a set of genes commonly down-regulated by both BTZ and sh*GAL* among the SMAD4 dependent targets as '*GALNT14*–TGFβ signature' (See Materials and Methods). The average activity of the *GALNT14*–TGFβ signature strongly discriminated patient RFS ($P = 4.0 \times$

10^{-4}, Fig. 4K), and *GALNT14* expression was significantly higher in lung cancer patients with higher levels of the signature (P =0.025, Fig. 4L). Overall, these results suggest that suppressing TGFβ signaling and gene expression responses relevant to BTZ (similar to the response observed after *GALNT14* depletion) makes a major contribution to reducing migration and invasion.

In vivo validation of ant-metastatic effect of BTZ

Given the anti-migration/invasion effect of BTZ in a lung cancer cell model (Fig. 2), we next tested the *in vivo* efficacy of BTZ against cancer metastasis *in vivo*. For this purpose, local metastasis was induced in mice by tail vein injection of H460 lung cancer cells, followed by treatment with or without BTZ or CFZ twice weekly for three weeks (Fig. 5A). The proteasome-inhibitory effect of BTZ (0.1mg/kg) with CFZ (0.5 mg/kg) was examined by measuring proteasome activity in blood (Fig. 5B). Under this concentration, the mice tolerated both BTZ and CFZ, exhibiting neither significant loss of body weight nor any other abnormalities (Fig. 5C). Consistent with the *in vitro* assay, the number of metastatic nodules in the lungs of BTZ-treated mice was significantly lower than that in CFZ- or vehicle-treated animals (Fig. 5D and S5A). Close examination of cancer tissue also revealed that inflammatory lesions, which provide favorable microenvironments for tumor formation (42), were present in both CFZ- and vehicle-treated mice (Fig. S5B). Taken together, the *in vivo* and *in vitro* data reveal that BTZ has a significant therapeutic advantage over CFZ in that it inhibits cancer metastasis without significant undesirable side effects.

Discussion

Although many molecular targets for tumorigenesis and metastasis have been identified, most remain undruggable. For example, the GALNTs, expression of which is strongly associated with various properties of cancer (20,21) (22-25), have yet to be drugged, although a few attempts have been made to develop inhibitors of GALNT-dependent O-linked glycosylation (31,32). Thereby, instead of searching direct inhibitors of GALNT14, we adopted an *in silico* DR approach to reverse the *GALNT14*-dependent metastatic expression signature with the goal of finding a candidate drug that could interfere with the *GALNT14*-dependent cancer phenotype. Notably, *HOXB9* and *SOX4,* transcription factor genes regulated by *GALNT14,* are responsible for metastasis (23) and self-renewal (20), respectively, suggesting that downstream transcriptional modulation would be a promising strategy.

Unlike similar studies in the past that used the CMap method to analyze differences in gene expression signatures between normal and cancerous tissue (14,43), we focused exclusively on genes (e.g., *GALNT14*) related to the pertinent phenotype (*i.e.*, the metastatic gene signature) (Fig. 1) and identified BTZ as a drug candidate with novel anti-metastatic effects both *in vitro* (Fig. 2C–2H) and *in vivo* (Fig. 5). Importantly, we also demonstrated that the anti-metastatic effect of BTZ, in contrast to that of CFZ, was independent of proteasome inhibition (Fig. 3). Moreover, the metastatic gene signature of CFZ was also distinct from that of BTZ, whereas the proteasome gene signatures of the two drugs were relatively similar (Fig. 3). Recent studies examined the inhibitory effect of BTZ on TGFβ-dependent responses such as fibrosis (44) and survival of lymphoma (45), and the results support our conclusions. We also identified the *GALNT14*-TGFβ signature that serves as clear indicators of poor prognosis (Fig. 4).

Thus, attenuation of TGFβ signaling by BTZ, depletion of *GALNT14*, or inhibition of TGFβ signaling all decrease invasive properties *in vitro* (Fig. 4) and lung metastasis *in vivo* (Fig. 5), suggesting the drug's mode of action. Accordingly, the *GALNT14*–TGFβ signature represents potentially useful prognostic marker for lung cancer patients, and could be used alongside previously reported marker [*e.g.,* the TGFβ-response signature (TBRS) (46) and the MAPK pathway activity score (MPAS) (47)]. Of note, although BTZ had no undesirable side effects in our *in vivo* experiments, the risk of peripheral neuropathy in patients treated with BTZ (48) merits a further search for other candidate drugs with safer profiles that could also reverse the gene signature associated with *GALNT14* or *GALNT14*-TGFβ activity. The continued search for improved candidates could be performed using recently reported *in silico* (or computational) DR tools (8,49,50). Notably in this regard, a recent *in silico* approach can predict candidate drugs capable of modulating the activities of oncogenic transcription factors, a class of proteins that has yet to be drugged (7). In the future, it would be interesting to apply this type of approach to modulate *GALNT14*-regulated transcription factor genes such as *HOXB9* and *SOX4,* which mediate metastasis (23) and self-renewal (20), respectively.

Besides predicting potential candidate drugs, our *in silico* DR approach enabled identification of several marker genes that turned out to be strongly associated with clinical outcomes such as RFS. This strategy, which integrated multiple independent expression signatures from cancer patients, genetic perturbation (*e.g.,* knockdown or overexpression), and drug treatment (CMap), would be applicable generally to any other types of target. Thus, our results provide a strong "proof-of-concept" that our DR method is a viable strategy for accessing undruggable molecular targets, leading to identification of candidate drugs that target specific cellular processes such as cancer

metastasis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

We appreciate Jeong-Hwan Kim, Seon-Young Kim and Dong-Uk Kim at Korea Research Institute of Bioscience and Biotechnology (KRIBB) for helpful discussions.

FUNDING

This work was supported by a grant from the National Research Foundation of Korea (NRF-2017M3C9A5028691 and NRF-2017M3A9B306184 from HJ.C and NRF-2017 M3A9B3061843 from W.K).

DISCLOSURE DECLARATION

The authors declare that they have no conflict of interest.

Authors' contributions

HJ.C and W.K conceived the overall study design and led the experiments. OS.K and H.L mainly conducted the experiments, data analysis, and critical discussion of the results. HJ.K, JE.P, and W.L conducted the mouse xenograft experiments. S.K, JH.K and M.K generated RNAseq data performed analysis. All authors contributed to manuscript writing and revising, and endorsed the final manuscript.

Figure Legends

Figure 1 *GALNT14* as a putative molecular target for lung cancer metastasis

A. Identification of therapeutic targets for both cancer progression and recurrence based on transcriptomic analysis of lung cancer patients. **B**. Selection of therapeutic target candidates that are differential expressed between tumor and normal tissue as well as significantly associated to RFS in the 516 patient dataset from TCGA LUAD. Red circles indicate metastasis-related genes annotated by MSigDB. **C**. Enrichment analysis of metastatic and tumor signatures between high- and low-expressed patient groups for each of the seven candidate genes. **D**. The normalized enrichment score (NES) was calculated by Gene Set Enrichment Analysis (GSEA) for the metastasis (left) and tumorigenesis signature (right) after all the genes were ranked by their expression fold change. **E-F.** 1 X 10^6 control H460 (shCont) and GALNT14 knockdown (shGal) cells were injected into lateral tail vein (E) or flanks (F) of nude mice. The representative H&E staining images of tumor-bearing lung (E) and tumors (F) were presented. **G**. Comparison of *GALNT14* dependencies among metastatic and primary cell lines from lung and other types of cancer from the *Project Achilles* dataset. Only metastatic lung cells show a significant dependency on *GALNT14*. **H**. Comparison between the high- and low- expression group for *GALNT14* in terms of locoregional recurrence-free survival (LRFS), and distant metastasis-free survival (DMFS) in LUAD patients.

Figure 2 Computational repositioning of BTZ to reverse the *GALNT14* expression signature

A. CMAP analyses to prioritize anti-metastatic candidate drugs using the two independent *GALNT14* signatures. Candidate drugs were prioritized according to the similarities between drug-induced down-DEGs and the two *GALNT14* signatures. **B**.

DR score of top candidate drugs selected by the *GALNT14* signatures from TCGA (x-axis) and H460 (y-axis). The two drugs marked in red (dexamethasone and bortezomib) got high scores in both predictions. **C.** Cell migration rate after treatment of BTZ and DEX respectively was measured. Representative images (left) and quantification graph (right) were exhibited **D.** Proteasome activity with IC50 concentration in H460 cell line was measured after indicated dose of BTZ treatment **E.** IC50 value of BTZ on cell migration capacity was measured in H460 cell line (right). Recovery ratio was measured 43 hours after BTZ treatment and representative image was shown (left). **F.** Representative image of cell, which is invaded through trans-well membrane was shown (left) and quantification of invaded area was presented (right). **G.** Flow cytometry plot (upper) of Annexin V and 7-AAD staining 24 hours after BTZ treatment (indicated concentration) and the quantification graph (lower) of cell death population were presented. **H.** Propidium iodide (PI) was stained to analyze the cell cycle profiles after BTZ treatment. Flow cytometry plot (left) and quantification graph of population in each cell cycle phase (right) were presented.

Figure 3 The effect of BTZ is independent of proteasome inhibition

A. Chemical structure of proteasome inhibitors (bortezomib, carfilzomib, and ixazomib) and their Tanimoto similarity heatmap. Red circle indicates the boronic acid structure. **B.** Immunoblotting for β-catenin, Cyclin D1 and p27 after treatment of BTZ (20 nM) and CFZ (20 nM) and α-tubulin for equal protein loading control **C.** Cell migration ratio were measured after 48 hours with 20nM of BTZ and CFZ treatment **D.** Area of cells, invaded through trans-well membrane was measured at 24 hour after treatment of BTZ and CFZ (20 nM) **E**. Sample clustering using t-SNE based on the expression of whole genes, metastasis-related genes and proteasome-related genes. Each sample was colored

according to its perturbation type. **F**. Venn diagram of differentially down-regulated genes by BTZ, CFZ and sh*GAL* perturbation in the gene space of whole genes.

Figure 4 Attenuation of the TGFβ gene response by BTZ treatment or GALNT14 knockdown

A. Significantly enriched pathways using the 101 genes (hypergeometric test, *P* value < 0.01 and FDR < 0.1) and the distribution of each of their RFS HR in TCGA LUAD cohort. The red circle represents the median of HR, and the bar across the circle represents the range of HR. **B.** mRNA expression of 10 common differentially expressed genes with loss of GALNT14 and BTZ treatment. **C.** The activity of TGFβ/Smad signaling in H460 shCont and H460 shGal was determined with SBE, ARE and BRE luciferase reporter systems. Luciferase activity was measured and its fold induction is shown as a graph. **D.** BRE luciferase activity was measured at 24 hours after BTZ and CFZ treatment. **E.** Protein level of Smad2 (pSmad2) and ERK2 (pERK2) phosphorylation were measured by immunoblotting 24 hours after BTZ and CFZ treatment. β-actin used as an equal protein loading control. **G-H** Wound healing assay (G) and Two-chamber invasion assay (H) were performed after TGFβ signaling suppression with siRNA targeting SMAD4 (siSMAD4). **I-L.** Analysis of TCGA LUAD cohort, **I**. The relationships between correlation score with *GALNT14* expression (z-transform of Spearman correlation, x-axis) and RFS HR (y-axis) of each gene in TGFβ target genes. The dependency of TGFβ target genes on Smad4 was labeled with two different colors (yellow: Smad4-dependent, blue: Smad4-independent). **J**. KM plot of RFS stratified by the combination of *GALNT14* expression (low and high) and TGFβ target activity (low and high). TGFβ target activity was measured using SMAD4 dependent genes (left) and independent genes (right) individually. Significant

differences between the patient groups were marked with the asterisk (*). **K**. KM plot of RFS stratified by the *GALNT14*-TGFβ signature. *P* values and HR were calculated with the log-rank test and Cox regression, respectively. **L**. Differences in *GALNT14* expression by *GALNT14*-TGFβ signature in TCGA LUAD cohort. *P* values were calculated with Student's t-test.

Figure 5 *In vivo* validation of anti-metastatic effect of BTZ

A. Schematic overview of *in vivo* experimental procedure **B.** Proteasome activity with whole blood collected 1 hour after BTZ (0.1 mg/kg) or CFZ (0.5 mg/kg) was shown in a bar graph. **C.** Body weight of each mouse, measured twice in a week was shown at indicative days. **D.** Representative lung image of each group was shown. White arrowheads indicate tumor nodule (top panels). Number of tumor nodules in the lung at each condition was shown in the table (bottom panel) **E.** H&E staining of lung from tumor-bearing mice after treatment of BTZ and CFZ

References

1. Spear, B.B., Heath-Chiozzi, M. and Huff, J. (2001) Clinical application of pharmacogenetics. *Trends Mol Med*, **7**, 201-204.
2. Olsen, D. and Jorgensen, J.T. (2014) Companion diagnostics for targeted cancer drugs - clinical and regulatory aspects. *Front Oncol*, **4**, 105.
3. Kim, E.S., Herbst, R.S., Wistuba, II, Lee, J.J., Blumenschein, G.R., Jr., Tsao, A., Stewart, D.J., Hicks, M.E., Erasmus, J., Jr., Gupta, S. *et al.* (2011) The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov*, **1**, 44-53.
4. Marquart, J., Chen, E.Y. and Prasad, V. (2018) Estimation of The Percentage of US Patients With Cancer Who Benefit From Genome-Driven Oncology. *JAMA Oncol*.
5. Lazo, J.S. and Sharlow, E.R. (2016) Drugging Undruggable Molecular Cancer Targets. *Annu Rev Pharmacol Toxicol*, **56**, 23-40.
6. Dang, C.V., Reddy, E.P., Shokat, K.M. and Soucek, L. (2017) Drugging the 'undruggable' cancer targets. *Nat Rev Cancer*, **17**, 502-508.
7. Gayvert, K.M., Dardenne, E., Cheung, C., Boland, M.R., Lorberbaum, T., Wanjala, J., Chen, Y., Rubin, M.A., Tatonetti, N.P., Rickman, D.S. *et al.* (2016) A Computational Drug Repositioning Approach for Targeting Oncogenic Transcription Factors. *Cell Rep*, **15**, 2348-2356.
8. Nagaraj, A.B., Wang, Q.Q., Joseph, P., Zheng, C., Chen, Y., Kovalenko, O., Singh, S., Armstrong, A., Resnick, K., Zanotti, K. *et al.* (2018) Using a novel computational drug-repositioning approach (DrugPredict) to rapidly identify potent drug candidates for cancer treatment. *Oncogene*, **37**, 403-414.
9. Ashburn, T.T. and Thor, K.B. (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nature reviews. Drug discovery*, **3**, 673-683.
10. Bertolini, F., Sukhatme, V.P. and Bouche, G. (2015) Drug repurposing in oncology--patient and health systems opportunities. *Nat Rev Clin Oncol*, **12**, 732-742.
11. Sleire, L., Forde, H.E., Netland, I.A., Leiss, L., Skeie, B.S. and Enger, P.O. (2017) Drug repurposing in cancer. *Pharmacol Res*, **124**, 74-91.
12. van't Veer, L.J. and Bernards, R. (2008) Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature*, **452**, 564-570.
13. Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K. *et al.* (2017) A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, **171**, 1437-1452 e1417.
14. van Noort, V., Scholch, S., Iskar, M., Zeller, G., Ostertag, K., Schweitzer, C., Werner, K., Weitz, J., Koch, M. and Bork, P. (2014) Novel drug candidates for the treatment of metastatic colorectal cancer through global inverse gene-expression profiling. *Cancer research*, **74**, 5690-5699.
15. Chen, B., Ma, L., Paik, H., Sirota, M., Wei, W., Chua, M.S., So, S. and Butte, A.J. (2017) Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat Commun*, **8**, 16022.
16. Lee, H., Kang, S. and Kim, W. (2016) Drug Repositioning for Cancer Therapy Based on Large-Scale Drug-Induced Transcriptional Signatures. *PLoS One*, **11**,

e0150460.

17. Ten Hagen, K.G., Fritz, T.A. and Tabak, L.A. (2003) All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology*, **13**, 1R-16R.

18. Wagner, K.W., Punnoose, E.A., Januario, T., Lawrence, D.A., Pitti, R.M., Lancaster, K., Lee, D., von Goetz, M., Yee, S.F., Totpal, K. *et al.* (2007) Death-receptor O-glycosylation controls tumor-cell sensitivity to the proapoptotic ligand Apo2L/TRAIL. *Nat Med*, **13**, 1070-1077.

19. Wu, C., Shan, Y., Liu, X., Song, W., Wang, J., Zou, M., Wang, M. and Xu, D. (2009) GalNAc-T14 may be involved in regulating the apoptotic action of IGFBP-3. *J Biosci*, **34**, 389-395.

20. Song, K.H., Park, M.S., Nandu, T.S., Gadad, S., Kim, S.C. and Kim, M.Y. (2016) GALNT14 promotes lung-specific breast cancer metastasis by modulating self-renewal and interaction with the lung microenvironment. *Nat Commun*, **7**, 13796.

21. Huanna, T., Tao, Z., Xiangfei, W., Longfei, A., Yuanyuan, X., Jianhua, W., Cuifang, Z., Manjing, J., Wenjing, C., Shaochuan, Q. *et al.* (2015) GALNT14 mediates tumor invasion and migration in breast cancer cell MCF-7. *Mol Carcinog*, **54**, 1159-1171.

22. Wang, R., Yu, C., Zhao, D., Wu, M. and Yang, Z. (2013) The mucin-type glycosylating enzyme polypeptide N-acetylgalactosaminyltransferase 14 promotes the migration of ovarian cancer by modifying mucin 13. *Oncol Rep*, **30**, 667-676.

23. Kwon, O.S., Oh, E., Park, J.R., Lee, J.S., Bae, G.Y., Koo, J.H., Kim, H., Choi, Y.L., Choi, Y.S., Kim, J. *et al.* (2015) GalNAc-T14 promotes metastasis through Wnt dependent HOXB9 expression in lung adenocarcinoma. *Oncotarget*, **6**, 41916-41928.

24. Shan, J., Liu, Y., Wang, Y., Li, Y., Yu, X. and Wu, C. (2018) GALNT14 Involves the Regulation of Multidrug Resistance in Breast Cancer Cells. *Transl Oncol*, **11**, 786-793.

25. De Mariano, M., Gallesio, R., Chierici, M., Furlanello, C., Conte, M., Garaventa, A., Croce, M., Ferrini, S., Tonini, G.P. and Longo, L. (2015) Identification of GALNT14 as a novel neuroblastoma predisposition gene. *Oncotarget*, **6**, 26335-26346.

26. Soria, J.C., Mark, Z., Zatloukal, P., Szima, B., Albert, I., Juhasz, E., Pujol, J.L., Kozielski, J., Baker, N., Smethurst, D. *et al.* (2011) Randomized phase II study of dulanermin in combination with paclitaxel, carboplatin, and bevacizumab in advanced non-small-cell lung cancer. *J Clin Oncol*, **29**, 4442-4451.

27. Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N. *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, **313**, 1929-1935.

28. Levy, L. and Hill, C.S. (2005) Smad4 dependency defines two classes of transforming growth factor {beta} (TGF-{beta}) target genes and distinguishes TGF-{beta}-induced epithelial-mesenchymal transition from its antiproliferative and migratory responses. *Molecular and cellular biology*, **25**, 8108-8125.

29. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G.,

Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M. *et al.* (2017) Defining a Cancer Dependency Map. *Cell*, **170**, 564-576 e516.

30. Stern, H.M., Padilla, M., Wagner, K., Amler, L. and Ashkenazi, A. (2010) Development of immunohistochemistry assays to assess GALNT14 and FUT3/6 in clinical trials of dulanermin and drozitumab. *Clin Cancer Res*, **16**, 1587-1596.
31. Gross, B.J., Swoboda, J.G. and Walker, S. (2008) A strategy to discover inhibitors of O-linked glycosylation. *J Am Chem Soc*, **130**, 440-441.
32. Hang, H.C., Yu, C., Ten Hagen, K.G., Tian, E., Winans, K.A., Tabak, L.A. and Bertozzi, C.R. (2004) Small molecule inhibitors of mucin-type O-linked glycosylation from a uridine-based library. *Chem Biol*, **11**, 337-345.
33. Gandolfi, S., Laubach, J.P., Hideshima, T., Chauhan, D., Anderson, K.C. and Richardson, P.G. (2017) The proteasome and proteasome inhibitors in multiple myeloma. *Cancer Metastasis Rev*, **36**, 561-584.
34. Maggiora, G., Vogt, M., Stumpfe, D. and Bajorath, J. (2014) Molecular similarity in medicinal chemistry. *J Med Chem*, **57**, 3186-3204.
35. Groll, M., Berkers, C.R., Ploegh, H.L. and Ovaa, H. (2006) Crystal structure of the boronic acid-based proteasome inhibitor bortezomib in complex with the yeast 20S proteasome. *Structure*, **14**, 451-456.
36. Wakefield, L.M. and Roberts, A.B. (2002) TGF-beta signaling: positive and negative effects on tumorigenesis. *Curr Opin Genet Dev*, **12**, 22-29.
37. Padua, D. and Massague, J. (2009) Roles of TGFbeta in metastasis. *Cell Res*, **19**, 89-102.
38. Li, A.M., Tian, A.X., Zhang, R.X., Ge, J., Sun, X. and Cao, X.C. (2013) Protocadherin-7 induces bone metastasis of breast cancer. *Biochem Biophys Res Commun*, **436**, 486-490.
39. Moon, Y.W., Rao, G., Kim, J.J., Shim, H.S., Park, K.S., An, S.S., Kim, B., Steeg, P.S., Sarfaraz, S., Changwoo Lee, L. *et al.* (2015) LAMC2 enhances the metastatic potential of lung adenocarcinoma. *Cell Death Differ*, **22**, 1341-1352.
40. Zhou, X., Updegraff, B.L., Guo, Y., Peyton, M., Girard, L., Larsen, J.E., Xie, X.J., Zhou, Y., Hwang, T.H., Xie, Y. *et al.* (2017) Dissecting the Role of PCDH7, an Oncogenic Cell Surface Receptor, in Non–Small Cell Lung Cancer. *Journal of Thoracic Oncology*, **12**, S1542.
41. Huang, D., Du, C., Ji, D., Xi, J. and Gu, J. (2017) Overexpression of LAMC2 predicts poor prognosis in colorectal cancer patients and promotes cancer cell proliferation, migration, and invasion. *Tumour Biol*, **39**, 1010428317705849.
42. Gupta, G.P. and Massague, J. (2006) Cancer metastasis: building a framework. *Cell*, **127**, 679-695.
43. Jahchan, N.S., Dudley, J.T., Mazur, P.K., Flores, N., Yang, D., Palmerton, A., Zmoos, A.F., Vaka, D., Tran, K.Q., Zhou, M. *et al.* (2013) A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. *Cancer Discov*, **3**, 1364-1377.
44. Zeniya, M., Mori, T., Yui, N., Nomura, N., Mandai, S., Isobe, K., Chiga, M., Sohara, E., Rai, T. and Uchida, S. (2017) The proteasome inhibitor bortezomib attenuates renal fibrosis in mice via the suppression of TGF-beta1. *Sci Rep*, **7**, 13086.
45. Chang, T.P., Poltoratsky, V. and Vancurova, I. (2015) Bortezomib inhibits expression of TGF-beta1, IL-10, and CXCR4, resulting in decreased survival

and migration of cutaneous T cell lymphoma cells. *J Immunol*, **194**, 2942-2953.

46. Padua, D., Zhang, X.H., Wang, Q., Nadal, C., Gerald, W.L., Gomis, R.R. and Massague, J. (2008) TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell*, **133**, 66-77.

47. Wagle, M.C., Kirouac, D., Klijn, C., Liu, B., Mahajan, S., Junttila, M., Moffat, J., Merchant, M., Huw, L., Wongchenko, M. *et al.* (2018) A transcriptional MAPK Pathway Activity Score (MPAS) is a clinically relevant biomarker in multiple cancer types. *Npj Precis Oncol*, **2**.

48. Argyriou, A.A., Iconomou, G. and Kalofonos, H.P. (2008) Bortezomib-induced peripheral neuropathy in multiple myeloma: a comprehensive review of the literature. *Blood*, **112**, 1593-1599.

49. Oprea, T.I. and Overington, J.P. (2015) Computational and Practical Aspects of Drug Repositioning. *Assay Drug Dev Technol*, **13**, 299-306.

50. Iorio, F., Rittman, T., Ge, H., Menden, M. and Saez-Rodriguez, J. (2013) Transcriptional data: a new gateway to drug repositioning? *Drug Discov Today*, **18**, 350-357.

Figure. 1

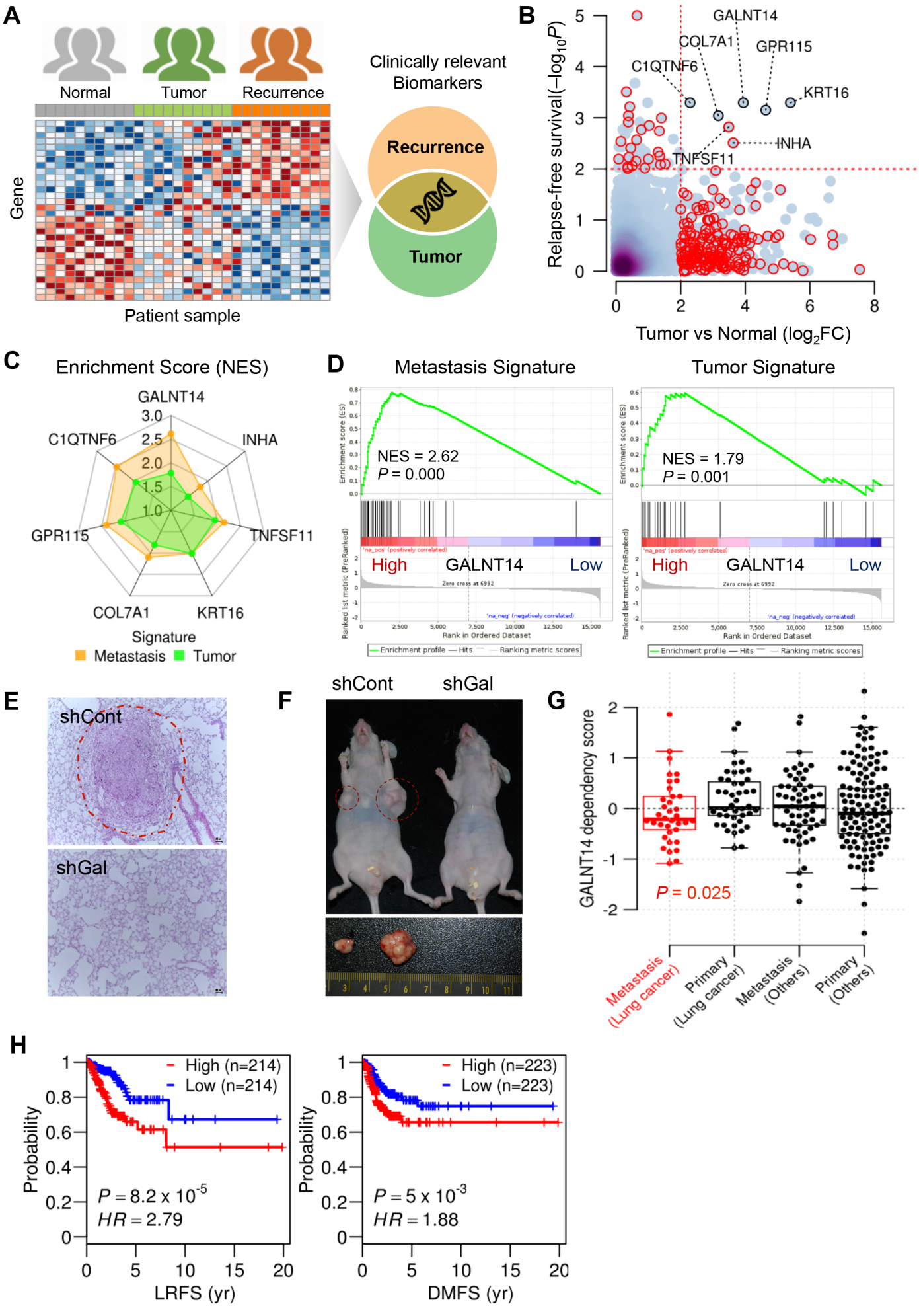


Figure. 2

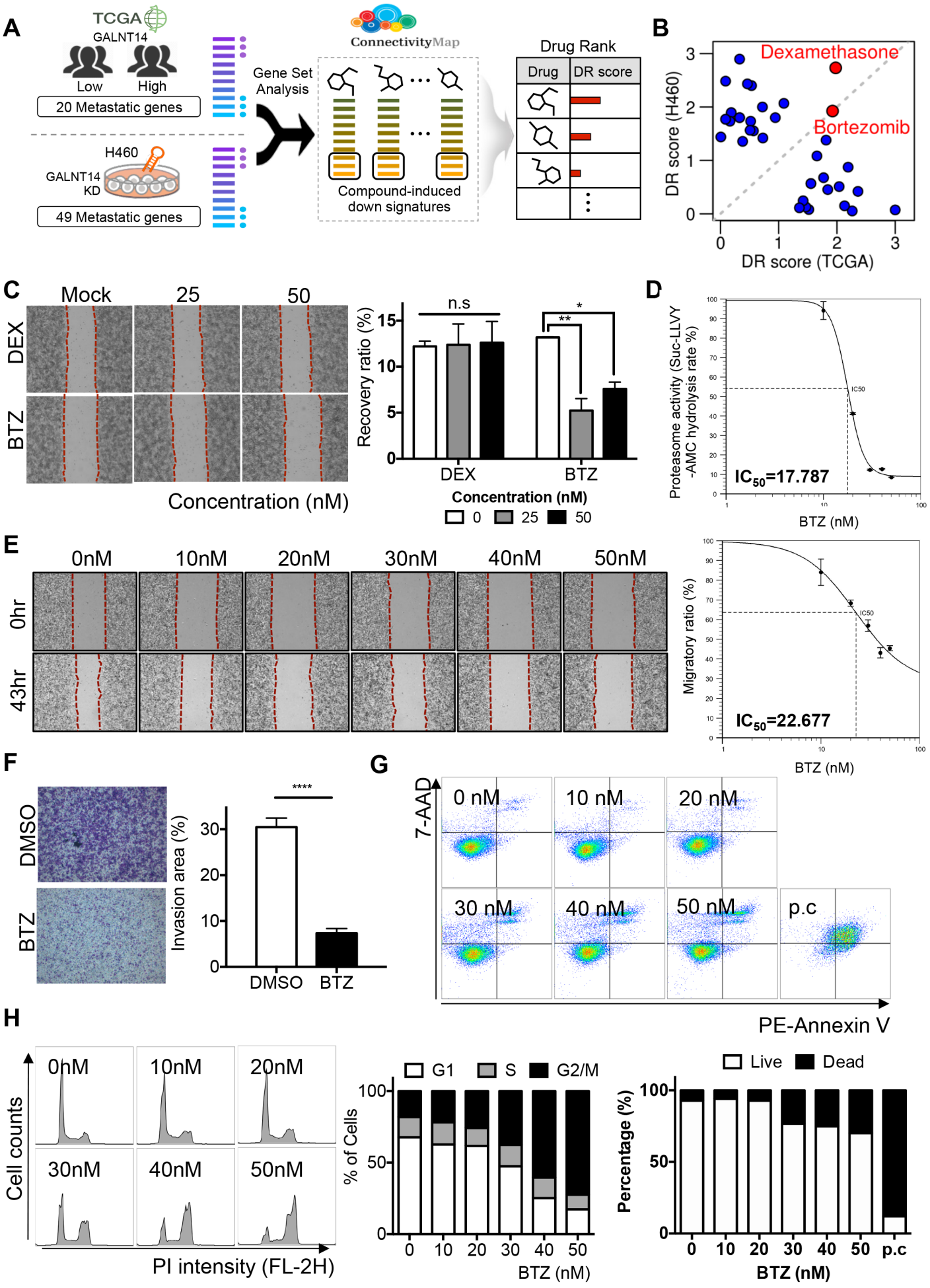
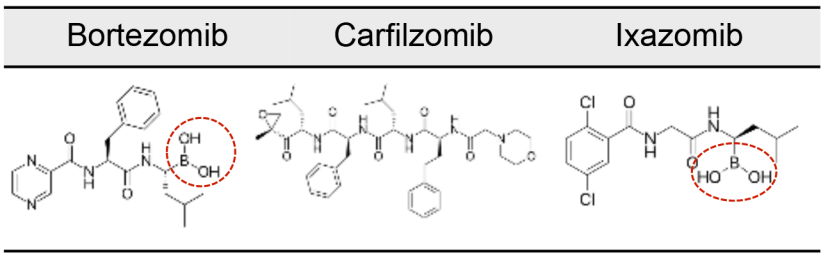
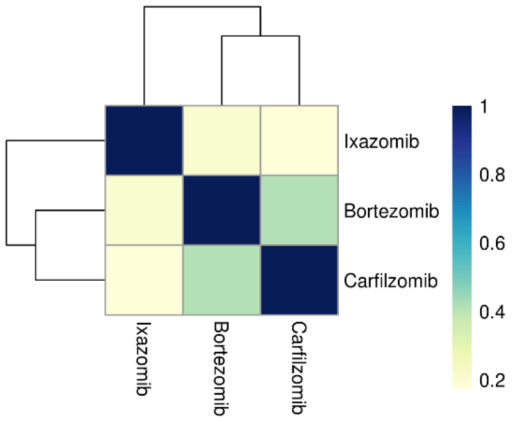


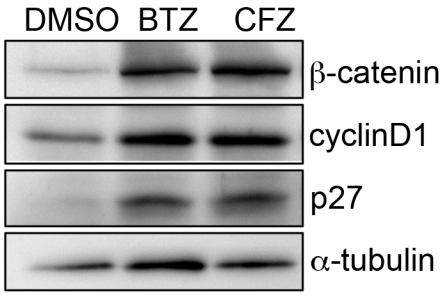
Figure. 3

A

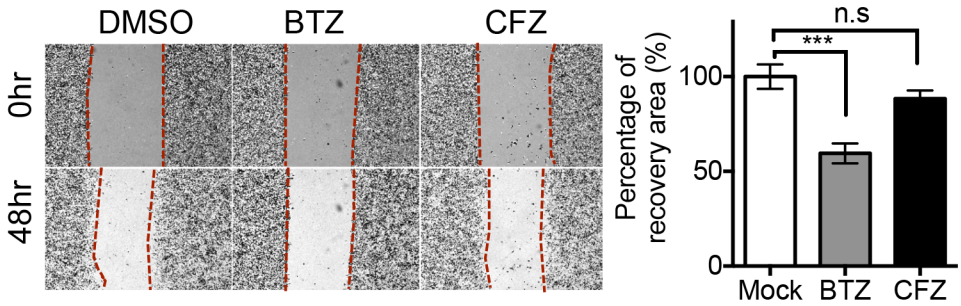




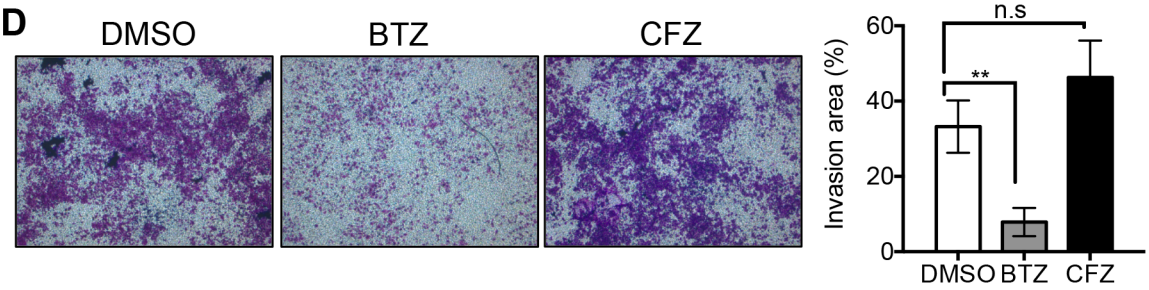
B



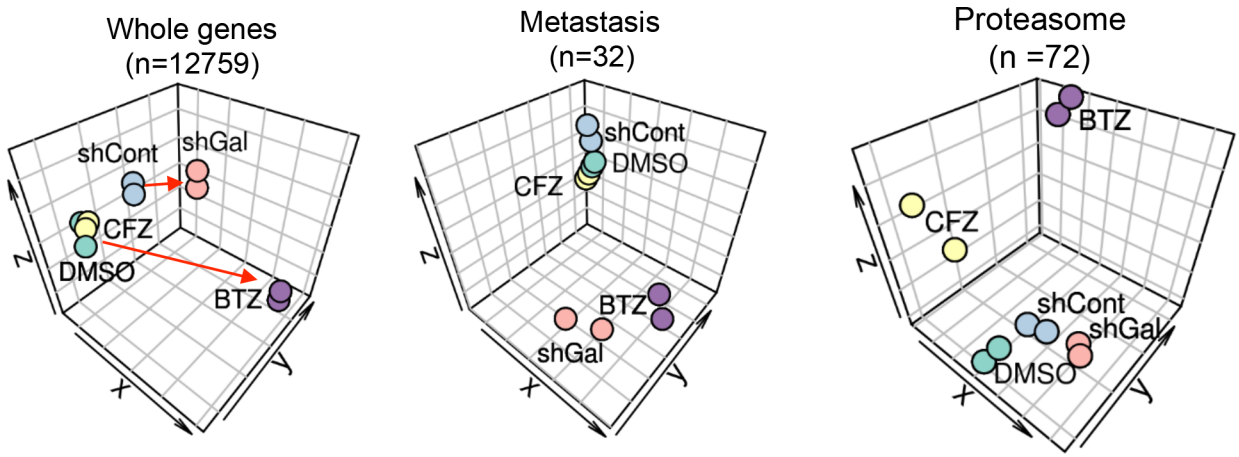
C



D



E



F

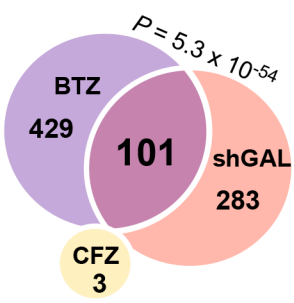
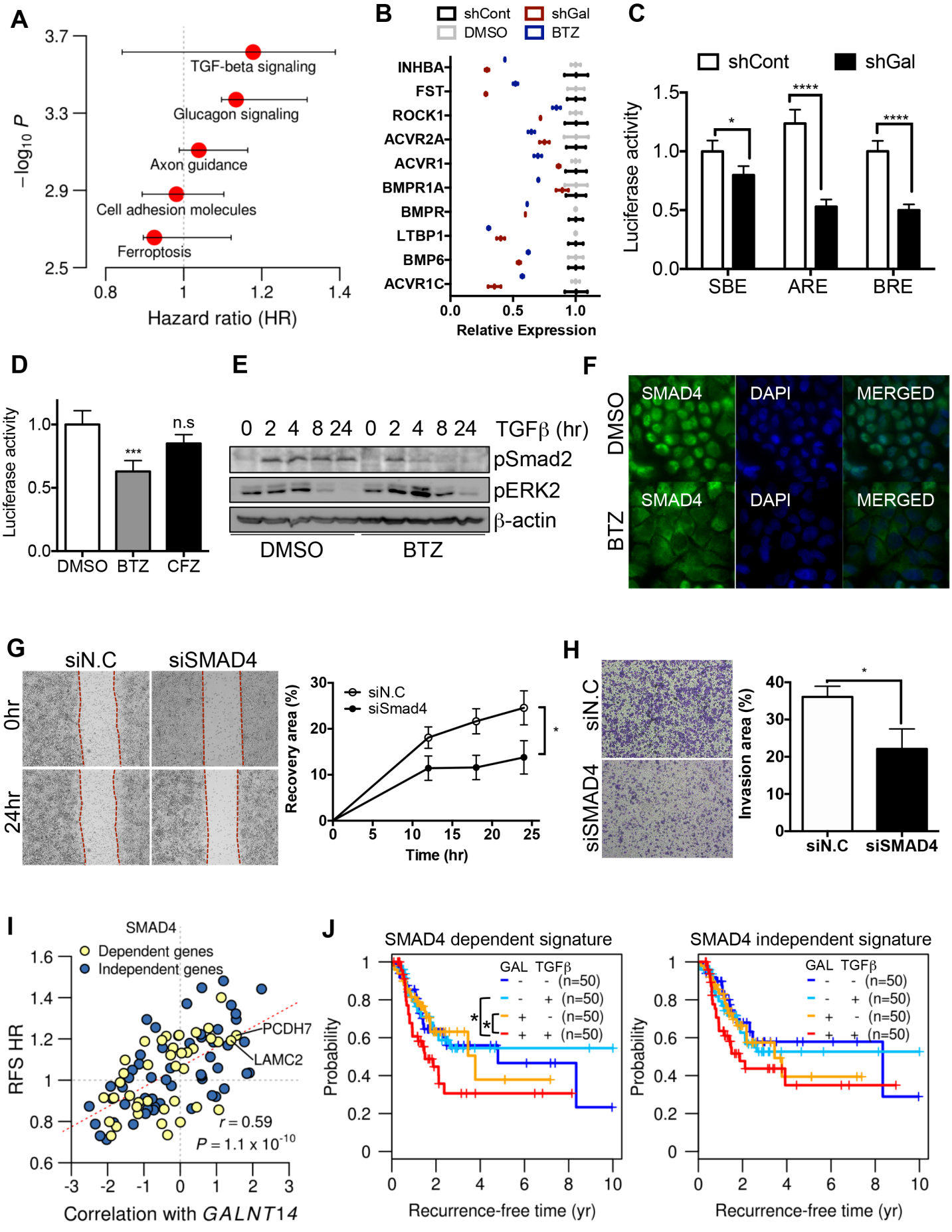


Figure. 4



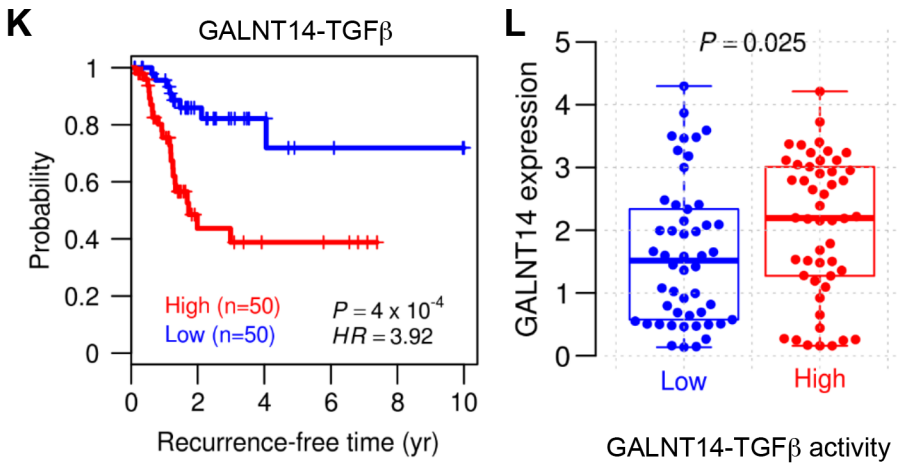


Figure. 5