Systematic Prioritization of Druggable Mutations in ~5000 Genomes Across 16 Cancer Types Using a Structural Genomics-based Approach*‡

Junfei Zhao‡‡‡, Feixiong Cheng‡‡‡, Yuanyuan Wang‡, Carlos L. Arteaga§¶¶, and Zhongming Zhao‡‡‡‡

A massive amount of somatic mutations has been cataloged in large-scale projects such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium projects. The majority of the somatic mutations found in tumor genomes are neutral ‘passenger’ rather than damaging “driver” mutations. Now, understanding their biological consequences and prioritizing them for druggable targets are urgently needed. Thanks to the rapid advances in structural genomics technologies (e.g., X-ray), large-scale protein structural data has now been made available, providing critical information for deciphering functional roles of mutations in cancer and prioritizing those alterations that may mediate drug binding at the atom resolution and, as such, be druggable targets. We hypothesized that mutations at protein–ligand binding-site residues are likely to be druggable targets. Thus, to prioritize druggable mutations, we developed SGDriver, a structural genomics-based method incorporating the somatic missense mutations into protein–ligand binding-site residues using a Bayes inference statistical framework. We applied SGDriver to 746,631 missense mutations observed in 4997 tumor-normal pairs across 16 cancer types from The Cancer Genome Atlas. SGDriver detected 14,471 potential druggable mutations in 2091 proteins (including 1,516 recurrently mutated proteins) across 3558 cancer genomes (71.2%), and further identified 298 proteins harboring mutations that were significantly enriched at protein–ligand binding-site residues (adjusted p value < 0.05). The identified proteins are significantly enriched in both oncoproteins and tumor suppressors. The follow-up drug-target network analysis suggested 98 known and 126 repurposed druggable anticancer targets (e.g., SPOP and NR3C1). Furthermore, our integrative analysis indicated that 13% of patients might benefit from current targeted therapy, and this proportion would increase to 31% when considering drug repositioning. This study provides a testable strategy for prioritizing druggable mutations in precision cancer medicine. Molecular & Cellular Proteomics 15: 10.1074/mcp.M115.053199, 642–656, 2016.

Several large-scale cancer genome sequencing projects, such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), have uncovered a large volume of somatic mutations in human cancers. These data provide an unprecedented opportunity to identify somatic mutations that play a pathogenic role in cancer development and progression (e.g., “driver” mutations) that, as such, can be exploited with a therapeutic intent (e.g., druggable targets) (1–3). However, the majority of these alterations detected in tumor genomes are “passenger” mutations, obscuring those mutations that are actionable in a background of errors, noise, and random alterations (4–7). The recent advances in structural genomic technologies, like x-ray and nuclear magnetic resonance, have helped investigators generate large-scale three-dimensional (3D) protein structure data. As of February 2015, more than 100,000 protein 3D structures (including ~27,000 human 3D protein structures) have been curated in the Protein Data Bank (PDB) database (8). Protein 3D structure and biological function are closely related.
related, especially in the case of protein–ligand binding-site residues, which are local regions that perform critical functions in cells such as binding with small molecules including drugs (9–12). Thus, development of new computational approaches that efficiently prioritize driver mutations according to their functional impact on protein 3D structures may provide an opportunity for the identification of molecules that can be the focus of targeted therapies.

Recently, several computational approaches have been developed for the prediction of putative cancer driver genes (13). These methods are classified into four main classes. The first class utilizes mutation frequency information and other genomic features (e.g., DNA replication time) to detect driver genes under the assumption that these genes have a higher mutation rate in comparison to the expected background mutation model, such as MutSigCV (14) and DrGap (15). The second class utilizes evolutionary conservation of amino acids to predict functional impact of nonsynonymous mutations, such as SIFT (16) and MutationAssessor (17). The third class of approaches examines the distribution of nonsynonymous mutations in specific protein regions or protein 3D structures to estimate their functional impacts, such as Oncodrive-CLUST (18), e-Driver (19), and mutation set enrichment analysis (MSEA) (20). MSEA was implemented by two methods, namely MSEA-clust and MSEA-domain, to predict cancer driver genes using mutation hotspot patterns in coding sequences or protein domain sequences (20). The application of MSEA to somatic mutations in seven cancer types in TCGA found a total of 82 putative cancer genes bearing significant mutation hotspots. The last class is based on the network perturbation hypothesis in cancer, that is, genetic aberrations may cause network architectural changes by affecting or removing a node or its connection within the network, or changing the biochemical properties of a node or protein (21). Representative methods include VarWalker (22) and HotNet2 (23). Although many methods and tools have been developed, so far, how to efficiently integrate somatic mutation data with structural genomics information to identify potentially druggable cancer driver gene products or actionable mutations for precision medicine remains a monumental challenge (24).

In this study, we proposed SGDriver, a novel, structural genomics-based method that incorporates the somatic missense mutations into the protein–ligand binding-site residues to decipher the functional role of somatic mutations and prioritize putative druggable molecular targets using a Bayes inference statistical framework (Fig. 1). We found significantly higher mutation rates at ligand binding-site residues in comparison with the whole protein sequences in all the 16 cancer types that we examined. Additionally, mutations at protein–ligand binding-site residues are more likely to be deleterious when compared with those at nonbinding-site residues. We applied SGDriver to over 740,000 missense mutations in ~5000 tumor-normal pairs in 16 cancer types in TCGA. We found 298 significantly mutated proteins (SMPs) that have enriched missense mutations at their protein–ligand binding-site residues (adjusted p value (q) < 0.05). We further performed a drug-target interaction network analysis by systematically integrating large-scale drug pharmacological data. This analysis found 98 known anticancer targets as well as 126 new mutated proteins that may be repurposed as potential druggable anticancer targets. Our integrative analysis suggested that 13% of cancer patients might benefit from currently approved targeted therapies, with this proportion potentially increasing to 31% by considering drug repositioning. In summary, SGDriver provides a useful tool to identify potential druggable SMPs that can be further studied for precision cancer medicine.

**EXPERIMENTAL PROCEDURES**

Collection and Preparation of Somatic Mutations—We downloaded the somatic mutation data from three sources: (1) Sanger website: (ftp://ftp.sanger.ac.uk/pub/cancer/AlexandrovEtAl), which contained 4,938,362 mutations from 7042 tumor samples (December 16, 2013); (2) Elledge’s Laboratory website at Harvard University (http://elledge-lab.med.harvard.edu/?page_id = 689; accessed in March 2014) (25), which contained 1,195,223 somatic mutations from 8207 tumor samples across 30 major cancer types; (3) COSMIC: Catalogue of Somatic Mutations in Cancer (v69), which is the largest database for somatic mutations. To reduce redundancy and ensure the quality of somatic mutation data, in this study, we only focused on the somatic mutations in TCGA tumor-normal matched samples from aforementioned three datasets. We used ANNOVAR (26) to map these somatic mutations into the protein sequences for identifying the corresponding amino acid changes based on RefSeq ID. The functional impact of nonsynonymous SNVs (single nucleotide variants) was assessed using SIFT (16) and PolyPhen 2 scores (27). For this analysis, we obtained SIFT and PolyPhen 2 scores from ANNOVAR annotation database. Then, we converted RefSeq ID to UniProt ID with the aid of UniProt ID mapping tool (http://www.uniprot.org/uploadlists/).

Gene Expression Data—Normalized gene expression data of RNA-Seq (V2) was extracted from TCGA using R package (TCGA-Assembler) (28). All public data files on TCGA Data Coordinating Center (DCC) server were gathered on January 5, 2015.

Collection and Preparation of Protein Structural Genomics Data—We extracted protein–ligand binding site data from BioLIP, which is a semi-manually curated database for high-quality, biologically relevant protein–ligand binding interactions (version October 10, 2014, the most updated version) (29). These annotated ligand molecules include DNA or RNA ligands, peptide ligands, metal ligands and regular ligands (i.e., the common small-molecule ligands). There were a total of 264,692 ligand–protein binding interactions, including 148,054 interactions for regular ligands, 31,797 interactions for DNA or RNA ligands, 72,088 interactions for peptide ligands, and 12,753 interactions for metal ligands. For each UniProt protein, we combined the protein–ligand binding-site residues of all the corresponding PDB structures. In total, there were 17,595 UniProt proteins having protein–ligand binding site information, including 2735 unique human proteins with the number of ligand binding-site residues ranging from 2 to 545 and median number being 19.

Description of SGDriver—SGDriver was developed under the hypothesis that somatic mutations that alter protein–ligand binding-site residues are more likely to induce the initiation and progression of cancer and mediate anticancer drug responses. SGDriver assumes that the observed number of somatic mutations across samples, y, at the ligand-binding-site residues of a protein, follows a Zero Inflated Poisson distribution (ZIP):
Structural Genomics-based Prioritization of Mutations

\[
f(y|\mu) = \begin{cases} 
  p + (1 - p)\mu^{-y}, & y = 0 \\
  (1 - p)\mu^{-y}y!, & y = 1, 2, 3, \ldots 
\end{cases}
\]

This means that if a SNV is present, \( y \) follows a Poisson distribution with background mutation rate parameter \( \mu > 0 \).

Given one protein sequence having \( n \) ligand-binding-site residues, \( Y = y_1, \ldots, y_n \) is the vector of observed number of mutations across samples per residue. For each ligand-binding-site residue, the number of mutations is counted at the direct protein-ligand binding-site residue and two flanking residues immediately upstream or downstream of the direct binding residue. The likelihood of observation is obtained as

\[
f(y_1, \ldots, y_n|\mu) = \prod_{i=1}^{n}(f(y_i|\mu) \propto e^{-\mu} \mu^{y_i})
\]

According to Bayes’ theorem, given a continuous prior \( f(\mu) \), the posterior distribution of \( \mu \) can be stated as

\[
f(\mu|y_1, \ldots, y_n) \propto \frac{f(y_1, \ldots, y_n|\mu)f(\mu)}{\int_0^\infty f(y_1, \ldots, y_n|\mu)f(\mu)d\mu}
\]

We choose the Jeffreys prior for the Poisson

\[
f(\mu) \propto \frac{1}{\sqrt{\mu}}, \mu > 0.
\]

The resulting posterior is a Gamma(\( \alpha, \beta \)) density with parameters \( \alpha = \frac{1}{2} \) and \( \beta = n \).

The posterior distribution is computed for each UniProt protein. The minimal \( p \) value of all ligand-binding residues is deemed as the significance for each protein. The previous study revealed that gene expression level affects somatic mutation rates in tumor genomes (14). In this study, we used the specific background mutation rates corrected by gene expression level to predict putative SMPs in each type of cancer. Specifically, we first computed average gene expression for each gene, and assigned these values to the corresponding gene products. For each UniProt protein of interest, we identified the top 100 proteins with the most similar expression levels; their mutational profiles were merged and used to compute the posterior distributions and \( p \) values. Finally, \( p \) values were adjusted for multiple testing by the Benjamini-Hochberg method (30). The adjusted \( p \) value \( (q) < 0.05 \) was used as the cutoff to choose the SMPs that harbor significantly enriched somatic mutations at the protein-ligand binding residues.

Collection of Cancer-associated Genes—We collected a large number of cancer genes from several publicly available sources. First, a total of 547 genes were downloaded from Cancer Gene Census (31) (accessed on November 5, 2014, denoted as CGC genes). CGC genes are well-curated and have been widely used as a reference cancer gene set in many cancer-related projects (32). Second, we collected 693 significantly mutated genes (SMGs) in cancer from TCGA publications and several other publications (details are shown in supplemental Table S3). Finally, we collected 478 oncogenes and 937 tumor suppressor genes as described in our previous studies (21, 33).

Construction of Drug-target Interaction Network—We constructed a drug-target interaction (DTI) network by a systematic integration of drug pharmacological data from three databases: DrugBank (v3.0, http://www.DrugBank.ca) (34), ChEMBL (v18) (35), and BindingDB (36). First, we collected the information of ~7700 drugs, including 1600 FDA-approved small molecule drugs, 1600 FDA-approved biotech drugs, and over 6000 experimental drugs. In order to improve the quality of DTI data, we only used the DTI pairs with pharmacological profiles (IC50, EC50 or \( K \)) less than 10 \( \mu \)M based on a previous study (37). All drugs were annotated using the detailed therapeutic information based on the Anatomical Therapeutic Chemical (ATC) classification system codes from the DrugBank database as described in previous studies (39, 39). All genes encoding drug targets were mapped to their Entrez IDs based on the National Center for Biotechnology Information (NCBI) database (40) as well as their official gene symbols based on HUGO Gene Nomenclature Committee (http://www.genenames.org).

Molecular Docking—We performed molecular docking simulation to repurpose potential drug candidates that are more likely to target mutant protein than wild-type (WT) protein using estrogen receptor (ER) as a case study because ER has both WT and driver mutant protein 3D structures (41, 42). We downloaded PDB files of wild-type ER (PDB ID: 1R5K) and p.Tyr537Ser mutant ER (PDB ID: 2QA8) from a previous study (42). We used MGLtools software (v1.5.6) (43) for receptor structure preparation by adding hydrogen and computing Gasteiger charges. The residues within 5 Å of ligands were chosen as docking pocket. We collected 40 known ER small-molecule ligands (e.g. ER antagonists and modulators) in SMILES format from the DrugBank database (34). We transferred SMILES to 3D structures for each ligand using SMILES translator and structure file generator from National Cancer Institute (http://cactus.nci.nih.gov/translate/). We performed ligand preparation and optimization using UCSF Chimera software (v 1.10.2) (44) by adding hydrogen and Amber ff14SB charges for standard residues and AM1-BCC for other residues. Structure energy minimum was calculated using the steepest method, with a minimum of 100 steps. Finally, we performed molecular docking experiments by using Autodock_vina software (v 1.1.2) (45) and retained top three docking poses that had the lowest free energy for further analysis.

Statistical Analysis and Network Visualization—All the statistical analyses were performed by the R package (v3.1.2, http://www.r-project.org/). We illustrated the network graphs using Cytoscape (v2.8.1) (46).

RESULTS

Overview of SGDriver—We examined the somatic mutation profiles in 4,997 tumor-normal pairs in 16 major cancer types using TCGA data (Fig. 1A and supplemental Table S1). These 16 major cancer types consist of acute myeloid leukemia (LAML), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon and rectal adenocarcinoma (COAD/READ), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), prostate adenocarcinoma (PRAD), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC). In this study, we used only missense mutations in our analyses because nonsense mutations typically have different mechanisms, like protein outcome (truncation or termination of translation). The mutational load (the mean number of missense mutations per
Structural Genomics-based Prioritization of Mutations

A: 
- GBM (318)
- HNSC (306)
- KICH (65)
- KIRC (448)
- LAML (185)
- LUAD (392)
- LUSC (176)
- OV (470)
- PRAD (150)
- SKCM (260)
- STAD (253)
- THCA (326)
- COAD/READ (491)

B: 
- Bladder
- Breast
- Colorectal/Rectum
- Glioblastoma
- Head and Neck
- Kidney (Chromophobe)
- Kidney (Clear)
- Myeloid
- Lung (Adeno)
- Lung (Squamous)
- Ovarian
- Prostate
- Skin Melanoma
- Stomach
- Thyroid
- Uterin

C: 
Prioritize putative cancer driver genes from the integration of structural genomics and next generation sequencing data using a Saxeye inference statistics framework

D:

E:

F: 
An integrated pipeline to prescribe drugs targeting significantly mutated proteins

Known anticancer drugs

Drug-target network visualization

Drug repositioning

Known non-anticancer drugs

Direct prescription
Structural Genomics-based Prioritization of Mutations

for each cancer type is shown in Fig. 1B. The missense muta-
tion rate of individual patient ranges from 8 to 497 with an average of 160 across the 16 cancer types. Among the 16 cancer types, LAML had the lowest mean mutation rate (8 missense mutations per sample), whereas UCEC had the highest mutation rate (497 per sample). We then developed SGDriver, a protein structural genomics-based approach to prioritize putative SMPs harboring missense mutations that are significantly enriched at protein–ligand binding-site residues using a Bayes inference statistical framework (see Methods). SGDriver postulates that if a gene product (protein) contains somatic mutations that are significantly enriched in protein–ligand binding-site residues, this protein would be more likely cancer-associated or related to anticancer drug responses (10, 12) (Fig. 1C). By applying SGDriver to 746,631 missense mutations in 4997 cancer genomes in 16 cancer types from TCGA (Fig. 1A and 1B), and using a statistical significance cutoff value \( q \leq 0.05 \), we identified 298 putative SMPs \( q < 0.05 \) in a pan-cancer analysis and a total of 335 putative SMPs in individual cancer analyses (Fig. 1D and 1E).

Finally, we used the drug pharmacological data to explore the potential druggable SMPs or actionable mutations in order to assist anticancer drug development in precision cancer medicine (Fig. 1F).

Pan-cancer Analysis of Protein–Ligand Binding-site Residue Mutations—We mapped missense mutations in 4997 tumor-normal pairs across 16 cancer types from TCGA into 2735 human protein–ligand complexes with unique UniProt ID. The final list was comprised of 7967 unique missense mutations at the protein–ligand binding-site residues covering 1119 unique human proteins (supplemental Table S5). After including the somatic mutations at the two immediate flanking residues of each direct binding residue, we obtained 14,471 mutations on 2091 proteins (including 1516 recurrently mutated proteins).

We first studied the 1119 proteins having mutations directly at the ligand-binding residues. We examined all mutations in the whole protein sequences and the mutations at the direct protein–ligand binding residues, respectively, in 16 cancer types (Fig. 2A). We found a significantly higher mutation rate at the direct ligand binding residues than that of the whole protein sequence; this is consistent among all the 16 cancer types \( p \) values ranged from \( 1.2 \times 10^{-14} \) to \( 7.8 \times 10^{-283} \), Wilcoxon rank-sum test, supplemental Table S2). For the instance of BRCA, there were 358 somatic missense mutations scattered at 60,946 protein–ligand binding residues; this compared with a total of 3256 mutations on 1,079,303 amino acids at the whole protein level. The mutation rate at the binding residues is nearly twice of that at the whole protein level \( p = 8.5 \times 10^{-20} \). These observations indicate that protein–ligand binding residues are significantly altered by somatic missense mutations. We further performed the comparison of mutation patterns between nonsmall molecule ligand binding-site residues and regular small-molecule ligand binding-site residues. As shown in supplemental Fig. S1, we did not observe a significant difference in mutation pattern between regular small-molecule ligands and nonsmall molecule ligands.

Ligand-binding-site Residue Mutations Tend to be Deleterious—We hypothesized that somatic missense mutations at protein–ligand binding-site residues have more deleterious impact than other missense mutations. We tested this hypothesis by measuring the functional impact scores of mutations using two popular and complementary tools: SIFT and PolyPhen2. SIFT (16) score is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences to predict whether an amino acid substitution affects protein function, whereas PolyPhen2 (27) score is based on the integration of eight sequence-based and three structure-based predictive features. We examined the cumulative distribution of SIFT and PolyPhen2 scores for protein–ligand binding-site residue mutations (direct position and two flanking residues) and nonbinding-site residue mutations (amino acid positions excluding the direct binding residues or two flanking residues). The distribution is shown in Fig. 2B and 2C. We defined the mutations with SIFT scores < 0.95 or PolyPhen2 score > 0.909 as deleterious mutations based on previous studies (12). We found that protein–ligand binding-site residue mutations are more likely to be deleterious mutations in comparison to nonbinding-site residue mutations when evaluated using both SIFT \( p < 2.2 \times 10^{-16} \), Fisher’s exact test, Fig. 2B) and PolyPhen2 \( p < 2.2 \times 10^{-16} \), Fig. 2C) scores. Next, we examined the cumulative distributions of deleterious mutations at different residues as a function of the distance from the binding-site residue. As shown in supplemental Fig. S2, we observed a clear trend that the enrichment of deleterious mutations diminishes by distance from the binding-site residue. The distribution of the weight of
deleterious mutations across 16 cancer types is shown in supplemental Fig. S4.

Prioritizing Significantly Mutated Proteins in Cancer Using SGDriver—Based on the observation that somatic mutations are enriched at protein–ligand binding-site residues (Fig. 2A) and that mutations at protein–ligand binding-site residues are more likely to be deleterious than those at other residues (Fig. 2B and 2C), we used SGDriver to prioritize putative SMPs harboring statistically significant excess number of mutations at protein–ligand binding-site residues among 16 cancer types. As shown in Fig. 3, we identified 298 SMPs harboring mutation-enriched protein–ligand binding-site residues with \( q < 0.05 \) in pan-cancer analysis (supplemental Table S4). All top five SMPs are well-known cancer proteins: BRAF (\( q < 10^{-200} \)), KRAS (\( q < 10^{-200} \)), EGFR (\( q = 4.9 \times 10^{-174} \)), IDH1 (\( q = 1.1 \times 10^{-147} \)), and TP53 (\( q = 3.1 \times 10^{-121} \)) (32). We then evaluated our method using the benchmark data. Specifically, we examined the enrichment of these 298 SMPs using two well-curated cancer gene data sets: (1) CGC genes (547 experimentally validated cancer genes), and (2) a curated list of published cancer SMGs (693 genes) from TCGA genome projects and other published literatures (supplemental Table S3). We found that the genes harboring a significant number of protein–ligand binding-site residue mutations were significantly enriched in CGC (\( p = 1.6 \times 10^{-12} \), hypergeometric test) and SMGs (\( p = 1.2 \times 10^{-14} \)), respectively.

Next, we prioritized potential SMPs for each individual cancer type using SGDriver. In total, we identified 335 SMPs with \( q < 0.05 \) in 16 cancer types, including 56 CGC proteins (\( p = 2.5 \times 10^{-13} \), hypergeometric test) and 72 published SMG products (\( p = 1.1 \times 10^{-19} \)). In the case of BRCA, we found 9 SMPs. All top five SMPs are well-known breast cancer proteins: AKT1 (\( q = 1.4 \times 10^{-138} \)), KRAS (\( q = 1.1 \times 10^{-16} \)), TP53 (\( q = 1.5 \times 10^{-8} \)), ERBB2 (\( q = 5.3 \times 10^{-8} \)), and GATA3 (\( q = 7.5 \times 10^{-4} \)) (47, 48). In the example of LUAD, we found 47 SMPs with \( q < 1.1 \times 10^{-4} \). The top 5 SMPs are KRAS (\( q = 0 \)), EGFR (\( q = 1.2 \times 10^{-201} \)), BRAF (\( q = 6.1 \times 10^{-80} \)), TP53 (\( q = 2.5 \times 10^{-73} \)), and MMP16 (\( q = 9.4 \times 10^{-49} \)). Among

![Fig. 2. Mutation frequencies and distribution pattern in 16 cancer types. A, Missense mutation frequencies across 16 major cancer types. Distribution of missense mutation frequencies in whole protein sequences (All) versus protein–ligand binding-site residues (LBS) in 16 major cancer types. B, C, Cumulative distribution of deleterious mutations at protein–ligand binding-site residues, their two immediate flanking residues, and nonbinding residues. Cumulative frequencies of SIFT (B) and PolyPhen-2 scores (C) for protein–ligand binding-site residues (direct positions), \( \pm 1 \) flanking residues (D + 1 flanking positions), and nonbinding residues (outside positions). Dotted line in each graph denotes the cutoff used to define “deleterious” mutations: SIFT score \( \leq 0.05 \) and PolyPhen2 \( \leq 0.909 \). Abbreviations of 16 cancer types in Figs. 2–5, and 7 are provided in the legend of Fig. 1.](image-url)
them, four (KRAS, EGFR, BRAF, and TP53) are well-known lung cancer proteins (49). MMP16, encoding matrix metalloproteinase-16, is involved in the breakdown of extracellular matrix in normal physiological processes and disease processes (i.e. tumor metastasis) (50). Thus, MMP16 may be an under-studied mutated gene involved in LUAD metastasis. For SKCM, we found 66 SMPs with \( q < 1.2 \times 10^{-7} \). Among the top five significant proteins, BRAF \( (q = 0) \), RAC1 \( (q = 7.9 \times 10^{-179}) \), IDH1 \( (q = 2.7 \times 10^{-138}) \), HBG2 \( (q = 1.3 \times 10^{-65}) \), and FABP1 \( (q = 6.8 \times 10^{-47}) \), three (BRAF, RAC1, and IDH1) are well known cancer proteins (51, 52). HBG2 encodes hemoglobin subunit gamma-2, mediates cancer immunotherapy (53). FABP1, which encodes fatty acid-binding protein 1, plays critical roles in fatty acid uptake, transport, and cancer cell metabolism (54–56). These analyses suggested that HBG2 and FABP1 might involve in cancer cell metabolism or tumor immunotherapy in melanoma. The details of the SMPs for other cancer types are shown in Fig. 4 and supplemental Table S4.

**Identifying Novel Druggable Mutations**—One main aim of cancer genomic studies is to identify novel anticancer targets or druggable mutations for targeted therapies and, furthermore, improve patient treatment and clinical care in a fashion of precision cancer medicine. To explore the potential application of SGDriver in precision cancer medicine, we systematically searched the drugs’ pharmacological spaces by integrating the predicted SMPs in both pan-cancer (Fig. 3) and individual cancer analysis (Fig. 4). As shown in Fig. 5A, 98 SMPs (such as BRAF, ERBB2, AKT1, and DNMT1) can be targeted by 72 U.S. FDA approved anticancer drugs with binding affinity less than 10 \( \mu M \) annotated in DrugBank, BindingDB, and ChEMBL databases. For example, molecules targeting on BRAF Val600Glu mutation (Fig. 5B) have been approved by U.S. FDA in melanoma targeted therapy, and specific drugs include vemurafenib and dabrafenib. In addition, targeting EGFR Leu858Arg mutation is being clinically investigated for lung cancer therapy (57).

Furthermore, we used drug repositioning strategy (38) to expand the scope of drugs targeting SMPs. We found that 126 SMPs were targeted by 596 repurposed drugs (noncancer drugs approved for other disease treatment). Among the 126 SMPs, 36 (e.g. RAC1 and TLR4) can only be targeted by noncancer drugs. To elucidate the potential application of SGDriver for prioritizing new druggable mutations, we briefly highlighted below three proteins (SPOP, ESR1, and NR3C1) harboring the enriched druggable mutations in protein–ligand binding-site residues across different cancer types. SPOP, encoding the cullin-RING ubiquitin ligase adapter protein speckle-type POZ protein, is mutated in 8–14% of prostate and endometrial cancers (58). As shown in Fig. 3 and 4, we found that SPOP is significantly mutated in pan-cancer \( (q = 8.7 \times 10^{-12}) \) and individual cancer (prostate cancer \( (q = 1.1 \times 10^{-6}) \) and uterine cancer \( (q = 7.4 \times 10^{-5}) \)). Moreover, one driver mutation, p.Phe133Leu (58), is significantly enriched at its ligand binding-site residue (Fig. 5C). This provides potential novel strategy for targeted prostate and uterine cancer therapy.

ESR1, encoding ER, is a member of the nuclear receptor family. 70% of breast cancer samples express ER, and most of these are sensitive to ER inhibitors (41). In this study, we found that ER is significantly mutated in uterine cancer \( (q = 1.4 \times 10^{-59}) \). Two mutations: p.Tyr537Ser and p.Asp538Gly, are significantly enriched at its ligand binding-site residues (Fig. 5C). Toy et al. recently reported three highly recurrent mutations (p.Tyr537Ser, p.Tyr537Asn, and p.Asp538Gly) in metastatic breast cancers (41). To further repurpose potential small-molecule drugs that are more likely to target mutant ER than wild-type ER, we docked 40 known ER small-molecule drugs (supplemental Table S6) collected from the DrugBank database into wild-type ER and p.Tyr537Ser mutant ER protein crystal structures, respectively (see Methods). We found that some ligands (e.g. genistein and trilostane) had a higher docking binding affinity on p.Tyr537Ser mutant ER structure compared with wild-type ER structure, whereas other ligands (e.g. tamoxifen, ERA-923, and fluvestrant) showed a lower docking binding affinity on p.Tyr537Ser mutant ER structure compared with wild-type ER structure (supplemental Table S6). A recent study reported that the resistance of tamoxifen...
is associated with ESR1 somatic mutations in breast cancer derived xenografts (59). In addition, Robinson et al. found that activation of ESR1 mutations is involved in hormone-resistant (such as tamoxifen and fluoruvast) metastatic breast cancer (60). Thus, tamoxifen, ERA-923, and fluoruvast might have a high resistance risk in ESR1 mutant (e.g., p.Tyr537Ser) breast cancer because of the decreasing binding affinity of the ligand-protein complexes (supplemental Table S6). On the contrary, genistein and trilostane that have the higher predicted binding affinity with p.Tyr537Ser mutant ER protein might be potential candidates to overcome anti-endocrine therapeutic resistance in breast and uterine cancer. Although this computational result has good support in literature, future experimental validation is required. In addition, future work in virtual screening and docking of mutated ER by comparison to biological property-matched decoy molecules is necessary for better interpreting their significance and identifying potential old drugs or new molecules for targeted breast cancer treatment.

NR3C1, encoding glucocorticoid receptor (GR), regulates genes that control the development, metabolism, and immune response (61). Here, we found that GR was significantly mutated in cancers (q = 0.03). The mutation p.Met604Ile is significantly enriched in ligand binding-site residues (Fig. 5C). GR agonists or antagonists had been approved for the treatment of allergic, inflammatory nasal conditions, and hypercortisolism, such as dexamethasone (agonist) and mifepristone (antagonist) (34). Skor et al. showed that GR antagonisms provided a novel therapy for triple-negative breast cancer (62). GR agonist dexamethasone inhibits EGF-induced cell migration by modulating transcriptional activities in MCF10A mammary epithelial cells (63). Collectively, U.S. FDA approved GR agents may provide new indications by targeting ligand binding-site residue mutations such as p.Met604Ile.

We next found that 152 SMPs were targeted by 790 clinical or preclinical agents. For example, KRAS Gly12Asp/Ser/Val mutations (Fig. 5B) have been widely investigated for molecular targeting in nonsmall cell lung cancer (64), including a covalent KRAS inhibitor SML-8–73-1 in lung cancer (65). Fig. 6 shows a bipartite drug-target interaction network connecting 180 SMPs (squares) and 673 U.S. FDA approved drugs (circles). The original Cytoscape file for preparing Fig. 6 is provided in supplementary file 1. All drugs are grouped by ATC classification system codes. As shown in Fig. 6, some FDA approved agents (i.e., sorafenib, afatinib, pazopanib, and nilotinib) can target multiple SMGs. Notably, some new SMGs, such as ALB, ADRB2, CHRM2, and NR3C1, may be potential candidates for drug repurposing in targeted cancer therapy. Put together, Fig. 6 offers a rich resource for identifi-
fying new repurposing approved drugs to target cancer SMPs for future cancer therapeutic development.

How Many Patients May Benefit From the Approved Therapeutic Agents?—We next investigated the tumor sample coverage for U.S. FDA approved drugs by targeting SMPs detected in aforementioned pan-cancer and individual analyses. Among 16 cancer types, there were 13 cancer types whose patients could benefit from the approved therapeutic agents by targeting SMPs; the three cancer types that did not benefit are: KICH, LAML and PRAD (Fig. 7). In total, of the 4997 TCGA tumor samples used in this study, 659 (13%) samples might be treated with 56 FDA approved anticancer drugs. If we expanded the search space to include repurposed drugs from all 596 FDA approved drugs, this coverage would increase to 31%. Specifically, there were 1556 samples harboring ligand binding-site residue mutations in SMPs that might be targeted by 470 FDA approved drugs. When focusing on individual cancer type, we found up to 60% of SKCM patients might benefit from 214 approved therapeutic agents (29 anticancer drugs and 185 repurposing drugs). This high proportion is largely attributed to BRAF Val600Glu mutation that is a well-studied target for targeted cancer therapy (66), occurring in 89 samples. The similar feature appeared in THCA where BRAF Val600 was altered in 189 samples and 56% of patients could benefit from 9 approved therapeutic agents. However, in our analysis, few available FDA approved targeted agents could be used for KIRC and BLCA targeted therapeutics (Fig. 7B). Drug repositioning may provide great opportunities for such kind of cancer treatments (e.g. KIRC and BLCA). For example, RARG, encoding retinoic acid receptor gamma, was significantly mutated in BLCA (q = 3.0 × 10^{-7}). Several dermatological drugs such as acitretin (34), tretinon (67), alitretinoin (68), and adapalene (69) may have new indications for BLCA treatment by targeting RARG (Fig. 6). Notably, alitretinoin showed preclinical and clinical anti-tumor activities in bladder (70), breast (71), and other cancer treatments (68). Although the results above are promising, they are still exploratory and future wet-lab experimental validation is needed.

**DISCUSSION**

Massive amounts of cancer genomics data generated from next-generation sequencing have motivated investigators to develop novel computational approach for prioritizing new
Fig. 6. **Global drug-target interaction network by targeting the predicted significantly mutated proteins (SMPs) in cancer.** In this network, nodes include 180 targets (SMPs, squares) and 668 U.S. FDA approved drugs (circles), and edges denote the interactions with pharmacological profiles (IC$_{50}$, EC$_{50}$ or $K$) less than 10 $\mu$M. All drugs were grouped using the Anatomical Therapeutic Chemical (ATC) classification system codes. The original Cytoscape file is provided in supplementary file.

druggable genes or mutations for the development of precision cancer medicine. Although the results we reported here are exploratory, the prioritized list can help cancer biologists and physician scientists to find the top candidates for experimental and clinical validation. In this study, we proposed SGDriver, a structural genomics-based approach under the Bayes inference statistical framework, to identify significantly mutated proteins and druggable mutations in cancer. To verify the hypothesis of SGDriver, we examined the mutational load distribution of missense mutations at protein–ligand binding-site residues and whole protein level. We found that a higher mutation rate at protein–ligand binding-site residues than that of whole protein level in all the 16 cancer types we examined. Furthermore, we found that missense mutations located at protein–ligand binding-site residues were more likely to be deleterious when compared with other residues.
Here, “deleterious” mutations are based on prediction that might be disruptive to its protein structure or function, rather than to the organism. SGDriver is not designed for detecting the effect of gain-of-function of somatic mutations either. To expand our study, we performed gene set enrichment analysis of SMPs identified by SGDriver in two important types of cancer genes: tumor suppressor genes, whose somatic mutations tend to be loss-of-function, and oncogenes, whose somatic mutations tend to be gain-of-function in cancer. Using the tumor suppressor genes and oncogenes collected in our recent studies (21, 33), we found that the identified SMPs in individual cancer type and pan-cancer analysis were significantly enriched in both tumor suppressor genes ($p = 3.1 \times 10^{-33}$, Fisher’s exact test) and oncogenes ($p = 1.5 \times 10^{-26}$) (supplemental Fig. S3). These results further indicated that the SMPs identified by SGDriver could be either oncoproteins or tumor suppressors, which is consistent with a recent study (72).

We proposed SGDriver and applied it to 4997 tumor genomes across 16 cancer types from TCGA. Based on analyzing this large tumor cohort, we identified 298 SMPs in each cancer type. The broken line indicates the number of cancer patients who may potentially benefit from the treatment of these agents.

Fig. 7. Survey of clinical benefit rate from therapeutic agents. A, Frequency of missense mutations at the protein–ligand binding-site residues of the top 50 significantly mutated proteins (SMPs) with the lowest adjusted p value. B, Summary of U.S. FDA approved or experimental agents targeting the predicted SMPs in each cancer type. The broken line indicates the number of cancer patients who may potentially benefit from the treatment of these agents.

**Table:**
| ATC Category          | Dermatologicals | Antiinfectives | Antineoplastic | Antiparasitic | Nervous system | Metabolism | Various | Endocrine | Cardiovascular | Blood | Respiratory system | Sensory organs | Musculo−skeletal | Antiparasitic |
|-----------------------|-----------------|----------------|----------------|---------------|----------------|-------------|---------|------------|-----------------|--------|---------------------|----------------|----------------|-------------|
| Dermatologicals       | D               | A              | B              | C             | D              | A           | B       | C          | D               | E      | F                   | G              | H               | I            |
| Antiinfectives        | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Antineoplastic        | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Antiparasitic         | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Nervous system        | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Metabolism            | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Various               | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Endocrine             | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Cardiovascular        | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Blood                 | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Respiratory system    | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Sensory organs        | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |
| Musculo−skeletal      | A               | B              | C              | D             | A              | B           | C       | D          | E               | F      | G                   | H              | I               | J            |

**Figure:**
- **A:** Frequency of missense mutations at the protein–ligand binding-site residues of the top 50 significantly mutated proteins (SMPs) with the lowest adjusted p value.
- **B:** Summary of U.S. FDA approved or experimental agents targeting the predicted SMPs in each cancer type. The broken line indicates the number of cancer patients who may potentially benefit from the treatment of these agents.
Structural Genomics-based Prioritization of Mutations

Molecular & Cellular Proteomics 15.2

geted therapy, and this proportion would increase to 31% if we adopted the drug repositioning strategy. Finally, we identified hundreds of druggable mutations (e.g., SPOP, ESR1, and NR3C1), which may be potential candidates for being targeted by FDA approved anticancer drugs or repurposing drugs. In summary, SGDriver provides a new, alternative tool to efficiently identify the potential druggable targets or mutations for rapidly emerging precision cancer medicine.

There are some limitations in current study. First, this study focused on missense mutations in protein–ligand binding-site residues and excluded other types of mutations, such as nonsense mutations, insertions/deletions (indels), synonymous mutations, copy number variation, or gene fusion. All these mutations may play important roles in tumor initiation and development. For example, some approved drugs target genetic events such as indels, copy number variant and gene fusion further reflects this limitation (29). Second, mutational heterogeneity of tumor samples may lead to false positive findings in prioritizing SMGs. In this study, we attempted to incorporate gene expression information to weaken the influence of mutational heterogeneity based on previous study (14). However, this practice may cause other kind of unknown biases in our analysis. Third, we did not distinguish different functional roles for driver mutations, such as loss-of-function and gain-of-function. For traditional cancer drug discovery and development, agonists are often designed to target loss-of-function drivers (e.g., p53 activators) (73), whereas inhibitors or antagonists are often designed to target oncogenes (e.g., EGFR and BRAF inhibitors) (74). Fourth, we limited our analyses only to the significantly mutated protein–ligand binding-site residues, which are directly targeted by approved or repurposed therapeutic agents. However, targeting the interacting partners of significantly mutated proteins, even these partner proteins are wild type, may be also an important strategy for anticancer drug design and development (24). Finally, structural genomics-based approaches are limited in the number of proteins with high-resolution 3D structures available. Until now, only ~15% proteins have known 3D structures in the whole human genome. In this study, only 11,119 proteins with high-resolution 3D structures were investigated in pan-cancer and individual studies. Finally, the prioritized SMPs are based on large-scale analyses. Their clinical implications need experimental validation, especially pharmacological work, although such work is usually labor-intensive and time-consuming.

Although there exist some limitations in current study, SGDriver provides useful features and an alternative tool for speeding up the discovery of potential druggable targets for targeted cancer therapy. Furthermore, SGDriver employs protein–ligand binding-site information for identifying SMGs and druggable mutations. Thus, SMGs and mutations identified by SGDriver should be more druggable than traditional computational approaches. In total, we identified 14,471 druggable somatic mutations, including well-known driver mutations and new mutations. These mutations serve a rich resource for future experimental or computational validation.

We further employed drug repositionings strategies to identify more druggable targets or mutations for cancer targeted therapeutics (38). For instances, integrated analysis indicated that the percentage of patients benefitting from current targeted therapy would increase from 13% to 31% by drug repositioning strategy. Considering the rapid accumulation of cancer sequencing data and protein structural genomics data, we plan to improve SGDriver in the near future in the following directions: (1) to predict SMGs with higher quality through the integration of more cancer mutations (including missense mutations, small indels, and nonsense mutations) and structural genomics data; (2) to reduce the false positive rates of SGDriver using the corrected background mutation model based on gene expression and DNA replication time information (14); (3) to characterize driver mutations into loss-of-function which may be targeted by agonists or activators and gain-of-function targeted by antagonists or inhibitors; (4) to utilize molecular network and gene expression data to integrate the interacting partners of significantly mutated proteins to expand the candidate pool for drug repurposing study; (5) to identify druggable mutations related to tumor cell colony formation using improved SGDriver to reduce the anticancer drug resistance (75); and (6) to identify old drugs or new small molecules to target specific mutations or mutated proteins for precision cancer therapies by performing molecularly docking-based virtual screening on the significantly mutated proteins identified by SGDriver. In summary, this study provides a novel computational framework by uniquely incorporating large-scale cancer genomics profiles (~5000 tumor-normal matched pairs) into protein structure genomics data and identified thousands of druggable mutations for therapeutic development in cancer.

* This work was partially supported by National Institutes of Health grants (R01LM011177, P50CA095103, P50CA098131, and P30CA068485). The Robert J. Kleberg, Jr. and Helen C. Kleberg Foundation (to Z.Z.), a Vanderbilt-Ingram Cancer Center’s Breast SPORE pilot project (to Z.Z.), and Ingram Professorship Funds (to Z.Z.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
§ This article contains supplemental material.
†† To whom correspondence should be addressed: Department of Biomedical Informatics, Vanderbilt University School of Medicine, 2525 West End Avenue, Suite 600, Nashville, TN 37203. Tel.: (615) 343-9158; Fax: (615) 936-8545; E-mail: zhongming.zhao@vanderbilt.edu.¶¶ These authors contributed equally to this work.

** Competing interests:** The authors declare that they have no competing interests.

REFERENCES

1. Hudson, T. J., Anderson, W., Artez, A., Barker, A. D., Bell, C., Bernabe, R. R., Bhan, M. K., Calvo, F., Eerola, I., Gerhard, D. S., Guttmacher, A., Guyer, M., Hemsley, F. M., Jennings, J. L., Kerr, D., Klatt, P., Kolar, P., Kusada, J., Lane, D. P., Laplace, F., Youyong, L., Needleman, G., Ozen-
Structural Genomics-based Prioritization of Mutations

Jones, S., Anagnostou, V., Lytle, K., Parpart-Li, S., Nesselbush, M., Riley, J. Jr., Raphael, B. J., Dobson, J. R., Oesper, L., and Vandin, F. (2014) Identifying driver mutations in sequenced cancer genomes: computational approaches for prioritizing driver mutations and significantly mutated genes in cancer genomes. Brief. Bioinform. bbv068

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, A. S., Kiezun, A., Hammerman, P. S., McKenna, A., Drier, Y., Zhu, L., Ramos, A. M., Shi, Y., Hahn, W. W., Rosenberg, R., Helmerson, H. R., Ambrogo, L., Nickerson, E., Sheffer, E., Cortes, M. L., Auclair, D., Saksgena, S., Voet, D., Noble, M., D’Cara, D., Lin, P., Lichtenstein, L., Heiman, D. I., Fennell, T., Imielinski, M., Hernandez, B., Hodis, E., Baca, S., Dulak, A. M., Lohr, J., Landau, D. A., Wu, C. J., Melendez-Zaigia, J., Hidalgo-Miranda, A., Koren, A., McCarroll, S. A., Mora, J., Lee, R. S., Crompton, B., Onofrio, R., Parkin, M., Winckler, W., Ardile, K., Gabriel, S. B., Stegmaier, C. W., Li, W. X., Zhu, Z., Elia, S., Toz, O., Lopez-Bigas, N., Luo, R., Muthuswamy, L., Ouellette, B. F., Pearson, J. V., Quesada, V., Raphael, B. J., Sander, C., Speed, T. P., Stein, L. D., Stuwart, J. M., Teague, J. W., Tokitoki, Y., Sunzdona, T., Valencia, A., Wheeler, D. A., Wu, H., Zhao, S., Zhou, G., Lathrop, M., Thomas, G., Yoshida, T., Axton, M., Gunter, C., Miller, L. J., Zhang, H., Haider, S. A., Wang, J., Yung, C. K., Oros, A., Liang, Y., Gnaneshwar, S., Gubernian, J. H., 118–123 Chalmers, D. M., Dhar, K. W., Kaan, T. S., Lowrance, W. W., Maus, T., Rodriguez, L. L., Vergely, C., Bowtell, D. D. C., Noonan, D. F., Aziano, L. R., Etemadmoghaddam, D., Gardiner, B. B., Kench, J. G., Sutherland, R. L., Tempero, A. M., Nelder, N. J., Wilson, P. J., Gallinger, S., Tsao, M. S., Shaw, P. A., Petersen, G. M., Mukhopadhyay, D., DePinho, R. A., Thayer, S., Shazand, K., Beck, T., Sam, T., Timma, L., Ballin, V., Lu, Y., Ji, J., Zhang, X., Chen, F., Hu, X., Yang, Q., Tiao, F., Zhang, X., Liang, Z., Zhao, L., Tan, B., Brenner, P., Holcato, I., Zaridze, D., Brazma, A., Egevad, L., Prokhorthokh, E., Banks, R. E., Uhlen, M., Viksna, J., Ponten, F., Skyrbajin, K., Birney, E., Borg, A., Borresen-Dale, A. L., Caldas, C., Foekens, J. A., Martin, S., Reis-Filho, J. S., Richardson, A. L., Sotriou, C., Thoms, G., van’t Veer, L., Birnbaum, D., Blanche, H., Boucher, P., Boyault, S., Masson-Jacquier, J. D., Paupote, I., Pivot, X., Vincent-Salomon, A., Tabone, E., Theillet, C., Treilleux, L., Biocas, P., Pecac, T., Francois, G., Gharbi, S., Zhang, Y., Tan, J., Kushi, L., Kallioniemi, O. P., Ley, T. J., Wainwright, B. J., Kallioniemi, O. P., Ley, T. J., Liu, E. T., and Wainwright, B. J. (2010) International network of cancer genome projects. Nature 464, 993–998

Chin, L., Andersen, J. N., and Futreal, P. A. (2011) Cancer genomics: from international perturbations in the cancer interactome. Mol. Biol. Evol. 28, 1079–1087

Tamborero, D., Gonzalez-Perez, A., and Lopez-Bigas, N. (2013) OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. Bioinformatics 29, 2238–2244

Porta-Faro, E., and Guigo, R. (2014) e-Drugs: a rational method to identify protein regions driving cancer. Bioinformatics 30, 3109–3114

Jia, P., Wang, Q., Chen, Q., Hutchinson, K. E., Pao, W., and Zhao, Z. (2014) MSEA: detection and quantification of mutation hotspots through mutation set enrichment analysis. Genome Biol. 15, 489

Cheng, F., Jia, P., Wang, Q., Lin, C. C., Li, W. H., and Zhao, Z. (2014) Studying tumorigenesis through network evolution and somatic mutational perturbations in the cancer interactome. Mol. Biol. Evol. 31, 2156–2169

Jia, P., and Zhao, Z. (2014) VarWalker: personalized mutation network analysis of putative cancer genes from next-generation sequencing data. PLoS Comput. Biol. 10, e1003460

Leiserson, M. D., Vandin, F., Wu, H. T., Dobson, J. R., Eldridge, J. V., Thomas, J. L., Papoutsaki, A., Kim, Y., Niu, B., McClellan, M., Lawrence, M. S., Gonzalez-Perez, A., Tamborero, D., Cheng, Y., Ryslik, G. A., Lopez-Bigas, N., Ge, G., Deng, L., and Raphael, B. J. (2015) Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat. Genet. 47, 106–114

Rubio-Perez, C., Tamborero, D., Schroeder, M. P., Antolin, A. A., DeuPons, J., Perez-Llamas, C., Mestres, J., Gonzalez-Perez, A., and Lopez-Bigas, N. (2015) In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer Cell 27, 382–396

Davoli, T., Xu, A. W., Mengwa, K. E., Sack, L. M., Yoon, J. C., Park, P. J., and Elledge, S. J. (2013) Cumulative haploinsufficiency and triposensitivity drive aneuploidy patterns and shape the cancer genome. Cell 155, 948–962

Wang, K., Li, M., and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic
46. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrating models of biomolecular interaction networks. Genome Res. 13, 2489–2504

47. Cancer Genome Atlas, N. (2012) Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70

48. Arteaga, C. L., and Engelman, J. A. (2014) ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. Cancer Cell 25, 282–303

49. Cancer Genome Atlas Research, N. (2014) Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543–550

50. Delassus, G. S., Cho, H., Park, J., and Eliceiri, G. L. (2008) New pathway links from cancer-progression determinants to gene expression of matrix metalloproteinases in breast cancer cells. J. Cell. Physiol. 217, 739–744

51. Hodis, E., Watson, I. R., Kryukov, G. V., Arolt, S. T., Imielinski, M., Theurillat, J. P., Nickerson, E., Auclair, D., Li, L., Place, C., Dicara, D., Ramos, A. H., Lawrence, M. S., Cibulskis, K., Sivachenko, A., Voet, D., Sakseha, G., Stransky, N., Onorito, R. C., Winckler, W., Ardile, K., Wagle, N., Jiao, W., Chong, K., Morton, D. L., Sternke-Hale, K., Chen, G., Noble, M., Murphy, M., Lai, J., Leunig, M., Miller, A. T., Levis, J. E., Wagner, S. N., Hoon, D. S., Schadendorf, D., Lander, E. S., Gabriel, S. B., Getz, G., Garraway, L. A., and Chin, L. (2012) A landscape of driver mutations in melanoma. Cell 150, 251–263

52. Krauthammer, M., Kong, Y., Ha, B. H., Evans, P., Bacchiocchi, A., McCusker, J. P., Cheng, E., Davis, M. J., Goh, G., Choi, M., Araya, S., Narayan, D., Dutton-Regester, K., Capatana, A., Holman, E. C., Bosserman, L., Szolnoki, M., Brash, J. D., Zbinden, M. R., and Wishart, D. S. (2011) DrugBank 3.0: a comprehensive resource for ‘omics’ research on drugs. Nucleic Acids Res. 39, D1035–D1041

53. Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., and Overing, J. P. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 40, D1100–D1107

54. Liu, T., Lin, Y., Wen, X., Jorissen, R. N., and Gilson, M. K. (2007) BindingDB: a database resource of the National Center for Biotechnology Information. Nucleic Acids Res. 35, D8–D20

55. Yeh, H. H., Ogawa, K., Balatoni, J., Mukhapadhyay, U., Pal, A., Gonzalez-Lepera, C., Shavrin, A., Soghomonyan, S., Flores, L., 2nd, Young, D., Volgin, A. Y., Najjar, A. M., Krasnykh, V., Tong, W., Alaudinni, M. M., and Gelovani, J. G. (2011) Molecular imaging of active mutant L858R EGFR receptor (EGFR) kinase-expressing nonsmall cell lung carcinomas using PET/CT. Proc. Natl. Acad. Sci. U.S.A. 108, 1603–1608

56. Theurillat, J. P., Uddeshi, N. D., Ermington, W. J., Svinkina, T., Bumb, C., Cooper, C., Kitchens, R. T., Lin, A., Phommaly, C., Morigy, D., Leu-
Kunju, L. P., Talpaz, M., Schott, A. F., and Chinnaiyan, A. M. (2013) Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat. Genet. 45, 1446–1451.

61. Lu, N. Z., Wardell, S. E., Burnstein, K. L., Defranco, D., Fuller, P. J., Giguere, V., Hochberg, R. B., McKay, L., Renoir, J. M., Weigel, N. L., Wilson, E. M., McDonnell, D. P., and Cidlowski, J. A. (2006) International Union of Pharmacology. LXV. The pharmacology and classification of the nuclear receptor superfamily: glucocorticoid, mineralocorticoid, progesterone, and androgen receptors. Pharmacol. Rev. 58, 782–797.

62. Skor, M. N., Wonder, E. L., Kocherginsky, M., Goyal, A., Hall, B. A., Cai, Y., and Conzen, S. D. (2013) Glucocorticoid receptor antagonism as a novel therapy for triple-negative breast cancer. Clin. Cancer Res. 19, 6163–6172.

63. Lauriola, M., Enuka, Y., Zeisel, A., D’Uva, G., Roth, L., Sharon-Sevilla, M., Lindzen, M., Sharma, K., Nevo, N., Feldman, M., Carvalho, S., Cohen-Dvashi, H., Kedmi, M., Ben-Chetrit, N., Chen, A., Solmi, R., Wiemann, S., Schmitt, F., Domany, E., and Yarden, Y. (2014) Diurnal suppression of EGFR signalling by glucocorticoids and implications for tumour progression and treatment. Nat. Commun. 5, 5073.

64. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A., and Shokat, K. M. (2013) K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. Nature 503, 548–551.

65. Hunter, J. C., Gurbani, D., Ficarro, S. B., Carrasco, M. A., Lim, S. M., Choi, H. G., Xie, T., Marto, J. A., Chen, Z., Gray, N. S., and Westover, K. D. (2014) In situ selectivity profiling and crystal structure of SML-8–73-1, an active site inhibitor of oncogenic K-Ras G12C. Proc. Natl. Acad. Sci. U.S.A. 111, 8895–8900.

66. Cancer Genome Atlas Research, N. (2014) Integrated genomic characterization of papillary thyroid carcinoma. Cell 159, 676–690.

67. Reddy, A. P., Chen, J. Y., Zacharewski, T., Gronemeyer, H., Voorhees, J. J., and Fisher, G. J. (1992) Characterization and purification of human retinoic acid receptor-gamma 1 overexpressed in the baculovirus-insect cell system. Biochem. J. 287 (Pt 3), 833–840.

68. Parrella, E., Gianni, M., Fratelli, M., Barzagio, M. M., Raska, I., Jr., Diomedes, L., Kurosaki, M., Pisano, C., Carminati, P., Merlini, L., Dallavalle, S., Tavecchio, M., Rochette-Egly, C., Terao, M., and Garattini, E. (2006) Antitumor activity of the retinoid-related molecules [(E)-3-(4′-hydroxy-3′-adamantyl)phenyl]-4-yl]acrylic acid (ST1926) and 6-[3-(1-adamantyl)-4-hydroxyphenyl]-2-naphthalene carboxylic acid (CD437) in F9 teratocarcinoma: Role of retinoid acid receptor gamma and retinoid-independent pathways. Mol. Pharmacol. 70, 909–924.

69. Kim, M. J., Ciletti, N., Michel, S., Reichert, U., and Rosenfield, R. L. (2000) The role of specific retinoid receptors in sebocyte growth and differentiation in culture. J. Invest. Dermatol. 114, 349–353.

70. Laaksovirta, S., Rajala, P., Nurmi, M., Tammela, T. L., and Laato, M. (1999) The cytostatic effect of 9-cis-retinonic acid, tretinoin, and isotretinoin on three different human bladder cancer cell lines in vitro. Urol. Res. 27, 17–22.

71. Lawrence, J. A., Adamson, P. C., Caruso, R., Chow, C., Kleiner, D., Murphy, R. F., Venzon, D. J., Shovlin, M., Noone, M., Merino, M., Cowan, K. H., Kaiser, M., O’Shaughnessy, J., and Zujewski, J. (2001) Phase I clinical trial of altretinoin and tamoxifen in breast cancer patients: toxicity, pharmacokinetic, and biomarker evaluations. J. Clin. Oncol. 19, 2754–2763.

72. Kambarov, A., Lawrence, M. S., Polak, P., Leshchiner, I., Lage, K., Golub, T. R., Lander, E. S., and Getz, G. (2015) Comprehensive assessment of cancer missense mutation clustering in protein structures. Proc. Natl. Acad. Sci. U.S.A.

73. Khoo, K. H., Verma, C. S., and Lane, D. P. (2014) Drugging the p53 pathway: understanding the route to clinical efficacy. Nat. Rev. Drug Discov. 13, 217–236.

74. Cheng, F., Jia, P., Wang, Q., and Zhao, Z. (2014) Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. Oncotarget 5, 3697–3710.

75. Carreira, S., Romanell, A., Goodall, J., Grist, E., Ferralideschi, R., Miranda, S., Prandi, D., Lorente, D., Frenel, J.-S., and Pezaro, C. (2014) Tumor clone dynamics in lethal prostate cancer. Sci. Transl. Med. 6, 254ra125–254ra125.