

**Supplementary information**  
**for**  
**Analysis of multiple compound–protein interactions reveals**  
**novel bioactive molecules**

Hiroaki Yabuuchi<sup>1†</sup>, Satoshi Nijima<sup>1†</sup>, Hiromu Takematsu<sup>2</sup>, Tomomi Ida<sup>2</sup>, Takatsugu Hirokawa<sup>4</sup>, Takafumi Hara<sup>3</sup>, Teppei Ogawa<sup>1</sup>, Yohsuke Minowa<sup>1</sup> Gozoh Tsujimoto<sup>3</sup> & Yasushi Okuno<sup>1\*</sup>

*<sup>1</sup>Department of Systems Biosciences for Drug Discovery, <sup>2</sup>Laboratory of Membrane Biochemistry and Biophysics, Graduate School of Biostudies and <sup>3</sup>Department of Genomic Drug Discovery Science, Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida-Shimo-Adachi-cho, Sakyo-ku, Kyoto 606-8501, Japan. <sup>4</sup>Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan.*

†These authors contributed equally to this work

\*To whom correspondence should be addressed. E-mail: [okuno@pharm.kyoto-u.ac.jp](mailto:okuno@pharm.kyoto-u.ac.jp)

## Contents:

### Supplementary methods

- Calculation of chemical and protein descriptors
- Construction of the prediction model using SVM
- LBVS methods based on chemical similarity
- Five-fold CV for the CGBVS and LBVS methods
- Polypharmacological prediction for ADRB2
- Radioligand binding assays
- Comprehensive predictions of binding modes between proteins and their ligands
- Analysis of ligand promiscuity
- Calcium mobilization assay
- cAMP accumulation assay
- Off-chip Mobility Shift Assay (MSA)
- Assessment of the novelty of the compound bioactivity

### References

#### Supplementary data sets

- Supplementary data set 1. Chemical structures of 11,500 Bionet compounds (a separate flat file)
- Table S1. CPI data for 5,207 ligand–GPCR pairs (a separate Excel file)
- Table S2. Performance comparison between the CGBVS and LBVS methods using five-fold CV
- Table S3. Performance comparison between the CGBVS and SBVS methods for ADRB2
- Table S4. Compound IDs and names for supplementary information (a separate Excel file)
- Table S5. The 50 top-ranked GPCR ligands for ADRB2 predicted by CGBVS
- Table S6. Substructures characteristic of ligand promiscuity for GPCRs
- Table S7. Compound IDs, names, and scores of 11,500 Bionet compounds (a separate Excel file)

Table S8. Scores and bioactivity of Bionet compounds ranked by CGBVS for GPCR ligand screening

Table S9. Known scaffolds and active compounds related to the top-ranked compounds

Table S10. Trained compounds that share maxMCS with hit compounds

Table S11. CPI data for 15,616 inhibitor–kinase pairs (a separate Excel file)

Table S12. Substructures characteristic of ligand promiscuity for kinases

Table S13. Scores and bioactivity of Bionet compounds ranked by CGBVS for kinase inhibitor screening

Table S14. Scores and bioactivity of Bionet compounds ranked by LBVS for kinase inhibitor screening

Table S15. Scores and bioactivity of Bionet compounds ranked by SBVS for kinase inhibitor screening

Figure S1. Outline of the CGBVS method

Figure S2. Polypharmacology map between 866 compounds and GPCRs without ligand information

Figure S3. Statistics of polypharmacology map of GPCRs

Figure S4. Polypharmacology map predicted by LBVS

Figure S5. Primary ligand screening on ADRB2, NPY1R, EGFR and CDK2

Figure S6. Polypharmacology maps for kinases predicted by CGBVS

Figure S7. Statistics of polypharmacology map of kinases

Figure S8. Performance comparison between CGBVS and conventional methods on predicting kinase inhibitors

Figure S9. Max-MCS list for active compounds

Figure S10. Experimental confirmation of the *in vitro* kinase activity of Bionet compounds screened by LBVS and SBVS

Figure S11. ROC curve obtained by CGBVS with reduced protein descriptors

Figure S12. Ligand screening for the Bionet compound set using LBVS and SBVS

## SUPPLEMENTARY METHODS

**Calculation of chemical and protein descriptors.** Chemical descriptors were calculated using the DRAGONX program (version 1.2; Talete S.r.l., Milan, Italy). In this study, 929 descriptors were calculated from descriptor categories 1–10 (constitutional descriptors, topological descriptors, walk and path counts, connectivity indices, information indices, two-dimensional (2D) autocorrelations, edge-adjacency indices, Burden eigenvalue descriptors, topological charge indices and eigenvalue-based indices), categories 17 and 18 (functional group counts and atom-centered fragments), and category 20 (molecular properties). Note that descriptors with values dependent on the 3D coordinates of the molecules (descriptor categories 11–14) were not used. Additionally, descriptors that count functional groups and atom types (categories 15 and 16) were omitted due to output errors of the program. Also, charge descriptors (category 19) were omitted because their values depend on the condition of the molecules. Finally, descriptors showing little variation across the compounds were eliminated. As a result, 797 descriptors were used for the subsequent processing.

Protein descriptors were calculated from the sequences alone. Specifically, dipeptide composition-based description (a mismatch-allowed spectrum kernel) was used to represent GPCRs, providing 400 dimensions. Mismatch-allowed spectrum kernel is based on shared occurrences of (k,m) patterns in the protein sequence, where the (k,m) pattern generated by a fixed k-length sub-sequence consists of all k-length sub-sequences differing from it by, at most, m mismatches. We set (k,m) at (2,1); hence, the number of descriptors is  $20^2$  (=400) (Leslie *et al.*, 2004). These descriptors are simple, yet are known to be quite good at predicting protein structural classes, functional classes, and subcellular localizations (Bhasin and Raghava, 2004; Hua and Sun, 2001). The advantage of using such descriptors lies in their being alignment-free and the easy to handle varying sequence lengths. After completing the series of experiments on GPCRs, we found that using additional descriptors, such as sequence-derived physicochemical features together with dipeptide composition, could lead to even better prediction models. In view of this observation, we chose to use a set of descriptors consisting of 1,497 features provided by the PROFEAT Webserver (Li *et al.*, 2006) in the

next series of experiments on kinases. Additionally, calculation of these descriptors was applied not to full-length sequences, but to domain sequences, because the five-fold cross validations confirmed that the domain sequence-based descriptor exhibited better accuracy (89.76%) than the full-length based descriptor (88.08%), consistent with the fact that kinase active sites tend to be localized in their domain sequences.

**Construction of the prediction model using SVM.** The SVM represents a class of statistical learning algorithms, the theoretical basis of which was first given by Vapnik (Vapnik, 1995). Due to its high generalization ability, the SVM has been applied to pattern recognition problems in various areas. A SVM classifier constructs a hyperplane that separates two different groups with a maximum margin, which is defined as the closest distance from any point to the separating hyperplane.

To find an optimal hyperplane that discriminates the presence and absence of CPIs, we constructed numerical vectors for compound–protein pairs (both for positive and negative samples) by concatenating chemical descriptors and protein descriptors, which were input to the SVM for learning. As information about non-interaction pairs was unavailable, we randomly generated the same number of negative samples as that of positive samples from pairs with unknown interactions. We used a portion of the codes from the LIBSVM suite of programs (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>), which employs a modified version of the sequential minimal optimization (SMO) algorithm (Schölkopf *et al.*, 2004; Platt, 1999). The parameters of the SVM with the radial basis function (RBF) kernel were optimized by a grid search.

Once the SVM classifier was obtained, we could test whether other CPIs have interactions. In addition to a simple output of a yes/no decision, scores could be assigned to the samples by estimating the SVM confidence in the correctness of the predicted outputs (Platt, 2000). This scoring method is based on the idea that samples lying closer to the hyperplane have a larger probability of being misclassified than those lying far away. We ranked all the tested CPIs according to the probability scores.

**LBVS methods based on chemical similarity.** We used popular ligand-based screening methods that compute the similarity of chemical descriptors. The neighboring compounds in the space defined by the chemical descriptors tend to show similar properties and biological activities; thus, these methods can be used to help in the discovery of lead compounds (Oprea, 2002). We scored the chemical compounds according to the distance to known ligands in the principal component (PC) space (Jolliffe, 1986) of the chemical descriptors. The number of PCs was determined so that the cumulative proportion of the explained variance exceeded 80%. We scored a pair of compound A and protein B on the basis of the similarity between compound A and the known ligands of protein B. Specifically, Pearson correlation coefficients were calculated between compound A and all known ligands of protein B in the PC space, and the maximum (that is, the similarity to the nearest-neighbor known ligand) was chosen as the score. We call this method the nearest neighbor (NN) with principal components analysis (NN-PCA). We also applied NN with the Tanimoto coefficient (NN-TC) and NN with the Pearson correlation coefficient (NN-PCC) in the original space.

**Five-fold CV for the CGBVS and LBVS methods.** The prediction performance was evaluated by five-fold CV, whereby samples were randomly split into five subsets of almost equal size; a classifier was constructed using four subsets (training set) and tested on the remaining subset (test set), and this process was repeated five times. The parameters of each classifier were optimized using a grid search with nested cross-validation, i.e., the above-mentioned training set was further split into three subsets (training set) and one subset (test set), and the best parameters were selected through the four-fold CVs.

Performance measures (shown in Supplementary Table S2) were calculated as follows:

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN}),$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}),$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}),$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}),$$

where TP, TN, FP, and FN denote the number of true-positives, true-negatives, false-positives, and false-negatives, respectively. We also used the ROC curve (Hanley and McNeil, 1982), which shows the false-positive rate versus the true-positive rate: that is, 1 – specificity versus sensitivity. The threshold was chosen as the farthest point from the diagonal line indicating a random ROC. Taking into account the performance variation caused particularly by the random generation of negative samples, we repeated the five-fold CV 20 times with different positive and negative sample sets. Our proposed CGBVS was compared to the other methods in terms of the mean accuracy of the five-fold CV, and the ROC curve.

**Polypharmacological prediction for ADRB2.** We predicted the probability of interaction of 826 compounds (excluding the 40 known ADRB2 ligands) with ADRB2 using a prediction model of CGBVS, which was constructed for validation with ADRB2 functional assays. Then, we ranked the compounds according to the CGBVS scores and carried out further analysis for the 50 top-ranked compounds. First, we surveyed the literature about the interaction between ADRB2 and these compounds using SciFinder (Wagner, 2006) and PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>). Then, we verified the interactions of ADRB2 with the available compounds by *in vitro* binding assays.

**Radioligand binding assays.** Competitive binding assays were performed in the presence of 30 pM <sup>125</sup>I-labelled-(+/-)-cyanopindolol (2200 Ci/mmol; PerkinElmer Life Sciences, Inc., Boston, MA) and on a 0.02 µg sample of membranes from human embryonic kidney 293 cells expressing ADRB2 in a total volume of 150 µL binding buffer [50 mM Tris–HCl (pH 7.4) and 10 mM MgCl<sub>2</sub>]. After a 1 h incubation at 37°C, unbound and membrane-bound radioactivity were separated by filtration of the samples through a glass fiber filter, followed by 10 washing steps with 200 µL ice-cold binding buffer, and were measured by liquid scintillation counting. Nonspecific binding was determined in the presence of 0.1 µM unlabelled propranolol. Here, we decided the ligand with K<sub>i</sub> < 100 µM as active.

**Comprehensive predictions of binding modes between proteins and their ligands.** As well as the prospective screening for ADRB2, we applied CGBVS to all the remaining compound–protein pairs. The prediction scores for all of the combinations between the proteins and their ligands were displayed as a 2D color map. Then, the number of predicted common ligands between every pair of proteins was represented by the elements of a polypharmacological interaction matrix in Figure 3C.

**Analysis of ligand promiscuity.** We used extended-connectivity fingerprints (ECFP<sub>6</sub>) as implemented by Pipeline Pilot (version 6.1.5; Accelrys, San Diego) to represent chemical substructures. The propensity for ligand promiscuity was defined as the posterior probability of promiscuity against selectivity given a substructure. The probability was calculated for each substructure using the Laplacian-corrected estimator (Xia *et al.*, 2004). The substructures with high probabilities were considered characteristic of the promiscuous ligands.

**Calcium mobilization assay.** Cellular effects of compounds were measured using the FLIPR Calcium Assay 4 Kit (Molecular Devices, Sunnyvale, CA) in CHO-K1 cells stably expressing human ADRB2. Briefly,  $1.5 \times 10^4$  cells were plated in half-sized black-walled plates and incubated overnight. The medium was changed to serum-free for 4 h. After loading with the fluorescent calcium indicator, compounds were injected into each well in a FLEX station II (Molecular Devices) and time-lapse changes in fluorescence (Ex 485 nm, Ex 525 nm) were monitored. For antagonistic activity, cells were pretreated for 10 min with compounds prior to 50 nM isoproterenol stimulation. The activities of test compounds were compared with that of vehicle (DMSO), and significantly active compounds ( $P < 0.05$  by unpaired t-test,  $n = 4$ ) were selected for secondary screening to determine EC<sub>50</sub> values.

**cAMP accumulation assay.** SK-N-MC (neuroblastoma) cells were washed twice with 20 mL of cAMP buffer (145 mM NaCl, 5 mM KCl, 1 mM MgSO<sub>4</sub>, and 10 mM HEPES, pH 7.4), supplemented with 0.5% bovine serum albumin and 10 mM glucose at 37°C. Cells were removed with a rubber policeman and resuspended in 50 mL of cAMP buffer (37°C); after centrifugation (5 min,  $150 \times g$ ), the pellet was washed again and resuspended in 10 mL of culture medium (37°C), after which the cell suspension was diluted to a final concentration of



$10^6$  cells/mL. Cell suspension (1 mL) was preincubated for 5 min at room temperature with 100  $\mu$ M (final concentration) papaverine and buffer or different concentrations of the test compounds. Neuropeptide Y solution (10  $\mu$ L) of different concentrations and 10  $\mu$ L of forskolin (1.5 mM) were added and incubated for 1 h at 37°C with shaking. The incubation was stopped by the addition of 0.1 mL of 1 M HCl (15 min incubation) and centrifugation (4°C, 15 min, 2000  $\times$  g). The supernatant was diluted with 0.05 M acetate buffer from the cAMP Kit (RPA 509; Amersham, Des Plaines, IL) and assayed. The sample (100  $\mu$ L) was incubated with 100  $\mu$ L of [ $^{125}$ I]cAMP and 100  $\mu$ L of anti-cAMP antiserum for 3 h at 4°C with shaking. The second antibody was added and the solution was mixed and incubated for 10 min at room temperature with shaking. After centrifugation (10 min, 3200  $\times$  g, room temperature), the supernatant was removed and the pellet counted in a  $\gamma$ -counter. Primary screening was performed using 50  $\mu$ M of test compound. Then, potent compounds showing more than 20% activation as compared to neuropeptide Y were selected for secondary screening to determine EC<sub>50</sub> values.

**Off-chip Mobility Shift Assay (MSA).** Compound solution (5  $\mu$ L of 4 $\times$ ), 5  $\mu$ L of 4 $\times$  substrate/ATP/metal solution, and 10  $\mu$ L of 2 $\times$  kinase solution were prepared with assay buffer (20 mM HEPES, 0.01% Triton X-100, 2 mM DTT, pH 7.5) and mixed and incubated in the wells of a polypropylene 384 well microplate for 1 or 5 h at room temperature. Then 60  $\mu$ L of Termination Buffer (QuickScout Screening Assist MSA; Carna Biosciences, Kobe, Japan) was added to each well. The reaction mixture was applied to a LabChip3000 system (Caliper Life Sciences, Hopkinton, MA), and the product and substrate peptide peaks were separated and quantitated. The kinase reaction was evaluated by the product ratio calculated from peak heights of product (P) and substrate (S) peptides [P/(P + S)]. Primary screening was performed using 50  $\mu$ M of test compound. Then, potent compounds showing more than 20% inhibition were selected for secondary screening to determine IC<sub>50</sub> values.

**Assessment of the novelty of the compound bioactivity.** We carefully performed a literature search for all hit compounds to confirm the novelty of their bioactivities. First, a chemical structure search was carried out using SciFinder and PubChem. If an identical structure was

found, we looked for references to the compound in biological experiments. Finally, we classified the bioactivity of the compound into two cases; one was the case of no report for our assayed target (i.e., ADRB2, NPY1R, EGFR, or CDK2), though bioactivities for other targets were reported, and the other was the case of no report about any biological activities.

Consequently, in the ADRB2 ligands, compound **8** (1-(*tert*-butylamino)-3-[(2-methyl-1H-indol-4-yl)oxy]-2-propanol) was previously reported to antagonize ADRB2 activity. Compound **33** (1-[(3-methoxypropyl)amino]-3-[(2-methyl-1H-indol-4-yl)oxy]-2-propanol) has been found to show agonistic/antagonistic activities on other GPCRs such as serotonin receptors. Compound **9** ((2-aminophenyl)(4-methylphenyl)amine) is known to inhibit many proteins other than in the kinase family, such as alpha-glucosidase, nucleotide-binding oligomerization domain-containing protein (NOD), and signal transducer and activator of transcription (STAT). Interestingly, compounds **31** (2-(3-isopropoxyphenyl)-1-ethanamine) and **32** (2-morpholino-2-oxoacetohydrazide) were reported to exhibit virus inhibition activities without their target information. Therefore, ADRB2 activities of these four compounds (**9**, **31**, **32**, **33**) have not been reported previously. For EGFR, compound **18** (6,7-dimethoxy-N-phenyl-4-quinazolinamine) was shown to act as an EGFR inhibitor. Although compounds **35** (7-chloro-N-(3-methoxybenzyl)-4-quinazolinamine) and **36** (6,7-Dimethoxy-N-(2-thienylmethyl)-4-quinazolinamine) were known to inhibit other proteins such as NOD1 and STAT, their inhibitory activities for EGFR have not been reported.

On the other hand, no biological activities have been reported for the rest, i.e., four ADRB2 ligands (**5**, **28**, **29**, **30**), three NPY1R ligands (**12**, **34**, **15**), two EGFR inhibitors (**21** and **37**), and two CDK2 inhibitors (**23** and **26**) (See Supplementary Figure S9 for details).

## REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410
- Bhasin M, Raghava GP (2004) Classification of nuclear receptors based on amino acid composition and dipeptide composition. *J. Biol. Chem.* **279**: 23262–23266
- Fredriksson R, Lagerström MC, Lundin LG, Schiöth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**: 1256–1272
- Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**: 29–36
- Hua S, Sun Z (2001) Support vector machine approach for protein subcellular localization prediction. *Bioinformatics* **17**, 721–728
- Jolliffe IT (1986) *Principal Component Analysis*. Springer, New York, USA
- Leslie CS, Eskin E, Cohen A, Weston J, Noble WS (2004) Mismatch string kernels for discriminative protein classification. *Bioinformatics* **20**: 467–476
- Li ZR, Lin HH, Han LY, Jiang L, Chen X, Chen YZ (2006) PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence. *Nucleic Acids Res.* **34**: W32–37
- Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK (2001) Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Comput.* **13**: 637–649
- Okuno Y, Yang J, Taneishi K, Yabuuchi H, Tsujimoto G (2006) GLIDA: GPCR–ligand database for chemical genomic drug discovery. *Nucleic Acids Res.* **34**: D673–677
- Oprea TI (2002) Chemical space navigation in lead discovery. *Curr. Opin. Chem. Biol.* **6**: 384–389

- Platt JC (1999) Fast training of Support Vector Machines using sequential minimal optimization. In *Advances in Kernel Methods: Support Vector Learning*, Scholkopf B, Burges C, Smola AJ (eds) pp 185–208. MIT Press, Cambridge, Massachusetts, USA
- Platt JC (2000) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*, Smola AJ, Bartlett PL, Scholkopf B, Schuurmans D (eds) pp. 61–74. MIT Press, Cambridge, Massachusetts, USA
- Schölkopf B, Tsuda K, Vert JP (2004) *Kernel Methods in Computational Biology*. MIT Press, Cambridge, Massachusetts, USA
- Vapnik VN (1995) *The Nature of Statistical Learning Theory*. Springer, New York, USA
- Wagner AB (2006) SciFinder Scholar 2006: an empirical analysis of research topic query processing. *J. Chem. Inf. Model.* **46**: 767–774
- Xia X, Maliski EG, Gallant P, Rogers D (2004) Classification of kinase inhibitors using a Bayesian model. *J. Med. Chem.* **47**: 4463–4470

## SUPPLEMENTARY DATA SETS

The chemical structures of the 11,500 Bionet compounds are available as **Supplementary Data Set 1** file (a separate flat file in SDF format).

CPI data, enlarged polypharmacology maps (Figure 3, Supplementary Figures S2 and S6), names for the top-ranked compounds (Supplementary Tables S5, S8, S13, S14 and S15), and chemical structure images for trained/predicted compounds are available at <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/>

**Table S2. Performance comparison between the CGBVS and LBVS methods using five-fold CV.** For each method, the mean and standard deviation (in percentages) of the accuracy, sensitivity, precision, specificity, and area under the curve (AUC) are shown (See Supplementary methods for details).

	Accuracy	Sensitivity	Precision	Specificity	AUC
CGBVS	<b>91.9 ± 0.3</b>	<b>90.9 ± 1.1</b>	<b>92.7 ± 0.7</b>	<b>92.8 ± 0.8</b>	<b>96.7 ± 0.3</b>
NN-PCA	84.4 ± 0.3	88.7 ± 0.5	81.7 ± 0.6	80.2 ± 0.9	91.6 ± 0.2
NN-TC	83.6 ± 0.3	86.9 ± 1.1	81.6 ± 0.6	80.4 ± 1.0	91.1 ± 0.2
NN-PCC	84.3 ± 0.3	88.4 ± 1.8	81.7 ± 1.1	80.2 ± 1.8	91.9 ± 0.2

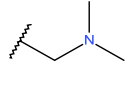
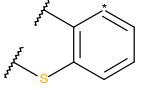
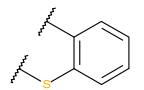
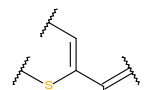
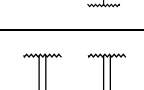
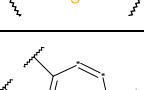
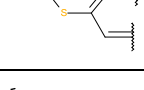
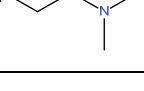
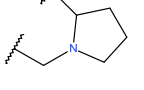
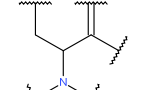
**Table S3. Performance comparison between the CGBVS and SBVS methods for ADRB2.** Enrichment factors (EFs) and hit rates (in percentages) are shown respectively for 10% and 20% of the top-ranked virtual libraries in the database (See Supplementary methods for details).

	Top 10%		Top 20%	
	EF	Hit rate	EF	Hit rate
CGBVS	<b>3.0</b>	<b>14</b>	<b>3.4</b>	<b>16</b>
Glide (Asp113)	2.7	13	2.4	11
Glide	1.5	7	1.6	8

**Table S5. The 50 top-ranked GPCR ligands for ADRB2 predicted by CGBVS.** IC<sub>50</sub> were determined by binding assays in the presence of [<sup>125</sup>I]-cyanopindolol and on membranes from human embryonic kidney 293 cells expressing ADRB2. Orange: ligands reported in the literature, green: screened active compounds, blue: inactive compounds, white: unavailable compounds. See Supplementary Table S4 for the other chemical information of the numbered compounds. This table in HTML format is available at <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/adrb2pp.php>.

Compound No.	GLIDA ID	Compound name	CGBVS score	Glide (asp113)	NN-PCA score	IC <sub>50</sub> (μM)	Reference (PubMed ID)
8	L001311	Sandoz-21-009	0.998	-8.36	0.882		7635249
38	L000835	Isamoltane	0.994	-6.88	0.903		1674359
39	L001043	Isoetarine	0.993	-7.25	0.971		6114117
3	L002023	Tropisetron	0.992	-5.24	0.794	37.7	
40	L000886	Prenalterol	0.991	-6.86	0.914		6126381
41	L001078	Metoclopramide	0.990	-6.30	0.711	67.4	
42	L000765	Methoxytyramine	0.990	-7.27	0.941	912	
43	L001080	H87-07	0.989	-4.65	0.913		6137200
44	L000596	Thioperamide	0.989	-4.30	0.683	N.D.	
45	L013367	Tertatalol	0.988	-8.45	0.837		2868822
4	L000152	BW723C86	0.988	-8.16	0.639	9.42	
46	L000084	7-OHDPAT	0.988	-6.13	0.734	N.D.	
47	L004706	A68930	0.987	-6.93	0.730	84.7	
48	L000652	RS6750	0.987	-6.64	0.373	62.6	
49	L004998	SC5311	0.987	-6.40	0.564		
50	L000725	Bupropion	0.986	-6.99	0.803		
51	L000015	(-)-AMMTC	0.985	-6.38	0.794		
52	L000082	6-chloromelatonin	0.985	-5.70	0.724		
53	L001059	Methoxamine	0.984	-7.14	0.881	12.3	
54	L000491	Practolol	0.983	-5.12	0.998		2895652
55	L000972	cis-(Z)-flupenthixol	0.983	-6.88	0.755		
2	L003700	Granisetron	0.981	-5.70	0.794	42.1	
56	L005352	SDZ 205-557	0.980	-5.56	0.687		
57	L000081	ADTN	0.979	-7.59	0.825		
58	L001989	Denopamine	0.978	-7.75	0.912		2897524
59	L001150	2,5-dimethoxy-4-ethylamphetamine	0.978	-6.62	0.880		
60	L000210	Cyanopindolol	0.977	-8.67	0.853		17028849
1	L000127	BIBP3226	0.977	-5.46	0.664	63.1	
61	L000745	Benzeneethanamine	0.977	-6.26	0.860		
62	L000061	2-me-5-HT	0.977	-6.26	0.798		
63	L001812	Xamoterol	0.976	-8.38	0.780		2862938
64	L001274	Levodopa	0.976	-6.64	0.803	89.1	
65	L000159	Carazolol	0.974	-8.67	0.782		17962520
66	L000151	BW245C	0.974	-6.90	0.615		
67	L000341	L803087	0.974	-6.86	0.604		
68	L000930	Hydroxyzine	0.972	-7.00	0.875	N.D.	
69	L015013	Ropinirole	0.971	-6.97	0.905	200	
70	L009221	SB204070	0.971	-5.35	0.598		
71	L000586	RP333	0.970	-7.79	0.832		8982677
72	L013372	Bisoprolol	0.970	-5.51	0.871		15655528
73	L001007	Octopamine	0.970	-6.58	0.976	73.7	
74	L001261	Melatonin	0.970	-5.31	0.713	N.D.	
75	L015545	Ciproxifan	0.969	-4.81	0.792		
76	L000677	Phenylephrine	0.968	-6.52	0.937		3725
77	L000938	Cisapride	0.968	-7.35	0.572	587	
78	L000647	Mexamine	0.968	-6.53	0.759		
79	L001535	Zacopride	0.968	-5.19	0.538	244	
80	L002130	N0774	0.968	-6.43	0.705		
81	L000232	Dopamine	0.967	-4.29	0.781		3725
82	L013385	Buspirone	0.967	-5.20	0.426	416	
83	L000785	Clebopride	0.967	-6.72	0.432		
84	L000078	5-hydroxytryptamine	0.967	-6.80	0.798		
85	L000091	AC-7954	0.967	-5.75	0.521	86.8	
86	L000106	Amisulpride	0.967	-6.40	0.381	105	

**Table S6. Substructures characteristic of ligand promiscuity for GPCRs.** The ten top-ranked frequent substructures represented by ECFP fingerprints are shown based on the highest posterior probabilities. See Supplementary Methods (Analysis of ligand promiscuity) for the detail calculation, and Supplementary Table S4 for the chemical names and structures of the numbered compounds.

ECFP pattern	Probability	Compound No.	Subfamily of target GPCRs
	0.988	286, 85, 287, 288, 127, 289, 290, 291, 292, 293, 294, 295, 296, 297, 141, 129, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 143, 308, 309, 310, 311, 312, 140, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323	Amines, Peptides
	0.986	324, 287, 325, 326, 327, 290, 328, 329, 330, 331, 332, 333, 296, 334, 335, 141, 336, 337, 55, 338, 300, 339, 340, 307, 341, 309, 310, 342, 343, 344, 345, 346, 347	Amines
	0.986	324, 287, 325, 326, 327, 290, 328, 329, 330, 331, 332, 333, 296, 334, 335, 141, 336, 337, 55, 338, 300, 339, 340, 307, 341, 309, 310, 342, 343, 344, 345, 346, 347	Amines
	0.986	324, 287, 325, 326, 327, 290, 328, 329, 330, 331, 332, 333, 296, 334, 335, 141, 336, 337, 55, 338, 300, 339, 340, 307, 341, 309, 310, 342, 343, 344, 345, 346, 347	Amines
	0.986	324, 287, 325, 326, 327, 290, 328, 329, 330, 331, 332, 333, 296, 334, 335, 141, 336, 337, 55, 338, 300, 339, 340, 341, 309, 310, 342, 343, 344, 345, 346, 347	Amines
	0.986	324, 287, 325, 326, 327, 290, 328, 329, 330, 331, 332, 333, 296, 334, 335, 141, 336, 337, 55, 338, 300, 339, 340, 341, 309, 310, 342, 343, 344, 345, 346, 347	Amines
	0.985	286, 85, 287, 288, 127, 289, 290, 291, 292, 293, 294, 295, 297, 141, 129, 299, 348, 301, 302, 303, 304, 305, 306, 307, 143, 308, 309, 312, 140, 313, 314, 316, 317, 318, 319, 320, 323	Amines, Peptides
	0.985	349, 350, 48, 351, 352, 83, 353, 354, 355, 356, 357, 358, 77, 359, 360, 338, 361, 362, 363, 364, 365, 366, 367, 368, 369, 70, 370, 371, 372, 373, 374	Amines, Peptides
	0.985	375, 376, 377, 378, 379, 380, 351, 381, 382, 383, 384, 385, 330, 386, 387, 388, 389, 348, 390, 391, 392, 342, 393, 394, 395, 396, 397, 398, 399, 400, 401	Amines, Peptides
	0.982	324, 287, 326, 327, 290, 328, 329, 331, 332, 333, 296, 334, 335, 336, 337, 338, 300, 339, 340, 341, 309, 310, 345	Amines



**Table S8. Scores and bioactivity of Bionet compounds ranked by CGBVS for GPCR ligand screening.** **A.** Prediction of ADRB2 ligands. The 30 top-ranked compounds, ordered by CGBVS score, are shown. EC<sub>50</sub> and IC<sub>50</sub> were determined by the FLIPR calcium assay in CHO-K1 cells stably expressing human ADRB2. **B.** Prediction of NPY1R ligands. The 20 top-ranked compounds, ordered by CGBVS score, are shown. EC<sub>50</sub> was determined by cAMP accumulation assay in SK-N-MC (neuroblastoma) cells. See Supplementary Table S4 for the other chemical information of the numbered compounds. This table in HTML format is available at <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/adrb2.php> (A) and <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/np1r.php> (B).

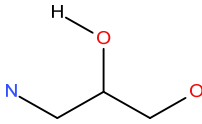
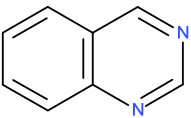
## A

Compound No.	Bionet ID	Compound name	CGBVS score	NN-PCA score	GOLD score	EC <sub>50</sub> (μM)	IC <sub>50</sub> (μM)
8	12L-933	1-( <i>tert</i> -butylamino)-3-[(2-methyl-1H-indol-4-yl)oxy]-2-propanol	0.950	0.853	40.7		0.0007
87	2W-0869	N-( <i>tert</i> -butyl)-N'-(2-methyl-1,3-benzoxazol-6-yl)urea	0.920	0.747	25.6		
88	2L-030	3-(2-hydroxyphenoxy)-2-butanone	0.917	0.907	26.2		
9	MS-2807	(2-aminophenyl)(4-methylphenyl)amine	0.912	0.863	35.4		37
31	MS-3556	2-(3-isopropoxyphenyl)-1-ethanamine	0.899	0.947	28.5		33
89	MS-3385	2-morpholino-N-(2,2,4-trimethyl-2,3-dihydro-1-benzofuran-7-yl)acetamide	0.891	0.670	37.5		
5	MS-2742	2,5-dimethyl-1-(2,2,4-trimethyl-2,3-dihydro-1-benzofuran-7-yl)-1H-pyrrole	0.888	0.529	36.4	65	
90	MS-2652	2-[2-(2-aminophenoxy)ethoxy]aniline	0.887	0.795	33.3		
29	2W-0814	N-( <i>tert</i> -butyl)-N'-(4-methoxybenzyl)thiourea	0.885	0.852	0.0	1.9	>100
91	12K-026	4-(1-methyl-2-oxopropoxy)benzenecarbonitrile	0.882	0.867	27.3		
92	8W-0277	ethyl 3-( <i>tert</i> -butyl)-1-(2-hydroxyethyl)-1H-pyrazole-5-carboxylate	0.878	0.740	26.9		
30	MS-0062	2-ethyl-2-[[2-(2-fluorobenzyl)oxy]methyl]-5,5-dimethyltetrahydrofuran	0.878	0.655	28.5	17	13
33	1M-918	1-[(3-methoxypropyl)amino]-3-[(2-methyl-1H-indol-4-yl)oxy]-2-propanol	0.878	0.765	38.8		0.0017
93	9G-920	4-[2-(ammoniooxy)acetyl]morpholine chloride	0.875	0.906	0.0		
94	3X-0356	N-(2,3-dihydro-1,4-benzodioxin-6-yl)-N'-(2-methyl-1-propenyl)urea	0.868	0.792	31.2		
95	MS-2037	4-[(propylimino)methyl]benzenol	0.866	0.930	27.0		
96	6J-911	2-(2-methoxyphenyl)-4,4-dimethyl-4,5-dihydro-1,3-oxazole	0.862	0.726	24.3		
97	MS-0628	2-(4-aminophenoxy)acetic acid hydrate	0.855	0.786	24.7		
98	KS-1145	Trimethoprim	0.854	0.660	24.5		
99	5J-328S	2,5,6-trimethyl-1H-1,3-benzimidazole	0.853	0.903	24.5		
100	MS-0090	2,2,4-trimethyl-2,3-dihydro-1-benzofuran-7-ylamine	0.852	0.836	28.3		
101	11P-700	methyl 4-[[E)-(2-hydroxy-4-methoxyphenyl) methylidene]amino]-2-methoxybenzenecarboxylate	0.852	0.520	31.5		
102	MS-2038	4-[(propylamino)methyl]benzenol	0.850	0.949	28.2		
103	MS-3213	1-[1-[4-(aminomethyl)phenyl]-1H-pyrrol-2-yl]-2-(diethylamino)-1-ethanol	0.849	0.722	37.7		
28	9X-0942	2-[2,5-dimethyl-4-(morpholinomethyl)phenoxy]acetamide	0.848	0.732	25.7		13
32	3F-004	2-morpholino-2-oxoacetohydrazide	0.847	0.886	0.0		48
104	6G-939	1-morpholino-1,3-butanedione	0.846	0.869	16.5		
105	3X-0258	N-(3,4-dimethoxyphenyl)-2,6-dimethyl-4-morpholinecarboxamide	0.845	0.761	27.8		
106	MS-2517	2-amino-N-( <i>tert</i> -butyl)benzenecarbohydrazide	0.837	0.817	0.0		
107	MS-2980	N-( <i>tert</i> -butyl)-N'-(2-furylmethyl)thiourea	0.836	0.881	0.0		

**B**

Compound No.	Bionet ID	Compound name	CGBVS score	NN-PCA score	EC <sub>50</sub> (μM)
108	8R-0629	2,4,8,10-tetramethylpyrido[2',3':3,4]pyrazolo[1,5-a]pyrimidine	0.990	0.916	
109	6W-0231	ethyl 4-[(2-methoxyethyl)amino]-3-methyl-1-(4-methylphenyl)-1H-pyrazolo[3,4-b]pyridine-5-carboxylate	0.982	0.801	
110	4R-0657	ethyl 1-(5-[[[(benzoylamino)carbothioyl]amino]-2-pyridinyl]-3,5-dimethyl-1H-pyrazole-4-carboxylate	0.980	0.692	
111	6F-929	diethyl 5-amino-3-(4-chlorophenyl)[1,2,4]triazolo[4,3-a]pyridine-6,8-dicarboxylate	0.980	0.752	
112	4R-0647	N-(4,6-dimethyl-1H-pyrazolo[3,4-b]pyridin-3-yl)thiourea	0.978	0.897	
113	4R-0661	ethyl 1-(5-[[[(aminocarbothioyl)amino]-2-pyridinyl]-3,5-dimethyl-1H-pyrazole-4-carboxylate	0.977	0.836	
114	6F-927	diethyl 5-amino-3-(4-pyridinyl)[1,2,4]triazolo[4,3-a]pyridine-6,8-dicarboxylate	0.975	0.776	
115	3H-939	diethyl 2-[2-[1-(6-fluoro-2-pyridinyl)-4-piperidinylidene]hydrazino]-6-hydroxy-3,5-pyridinedicarboxylate	0.973	0.706	
116	7T-0113	N-methyl-N'-(1,4,6-trimethyl-1H-pyrazolo[3,4-b]pyridin-3-yl)thiourea	0.973	0.913	
117	3H-940	diethyl 2-hydroxy-6-[2-[1-(4-nitrophenyl)-4-piperidinylidene]hydrazino]-3,5-pyridinedicarboxylate	0.973	0.583	
118	8P-512S	methyl 6-[4-(4,8-dimethyl-2-quinolinyl)piperazino]nicotinate	0.970	0.814	
119	1H-921	diethyl 2-(2-[1-[3-chloro-5-(trifluoromethyl)-2-pyridinyl]-4-piperidinylidene]hydrazino)-6-hydroxy-3,5-pyridinedicarboxylate	0.969	0.571	
15	3H-950	diethyl 2-(3,5-dimethyl-1H-pyrazol-1-yl)-6-hydroxy-3,5-pyridinedicarboxylate	0.967	0.827	63
120	9W-0807	ethyl 6-[(6-chloro-3-pyridinyl)methyl]-5,7-dimethylpyrazolo[1,5-a]pyrimidine-3-carboxylate	0.966	0.827	>100
121	8W-0821	6-[(6-chloro-3-pyridinyl)methyl]-2,5,7-trimethylpyrazolo[1,5-a]pyrimidine	0.965	0.914	
122	7T-0131	3-(2,5-dimethyl-1H-pyrrol-1-yl)-1,4,6-trimethyl-1H-pyrazolo[3,4-b]pyridine	0.965	0.877	
12	7W-0360	ethyl 1-(4-chlorophenyl)-4-[(4-methoxybenzyl)amino]-3-methyl-1H-pyrazolo[3,4-b]pyridine-5-carboxylate	0.964	0.848	16
123	8N-803	ethyl 3-(2-[2-[(allylamino)carbothioyl]hydrazino]-2-oxoethoxy)pyrido[1,2-a]indole-10-carboxylate	0.964	0.630	
124	5W-0344	ethyl 4-chloro-3-methyl-1-(4-methylphenyl)-1H-pyrazolo[3,4-b]pyridine-5-carboxylate	0.963	0.893	
34	6W-0328	ethyl 4-chloro-1-(4-chlorophenyl)-3-methyl-1H-pyrazolo[3,4-b]pyridine-5-carboxylate	0.963	0.834	16

**Table S9. Known scaffolds and active compounds related to the top-ranked compounds.** This table in HTML format is available at <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/scaf.php>.

Structure	Name	Activity	Reference
	Oxypopropanolamine	Beta-adrenoreceptor antagonist	Hosohata <i>et al. Gen. Pharmacol.</i> , <b>26</b> , 743-747 (1995)
	Quinazoline	EGFR inhibitor	Vieth <i>et al. Biochem. Biophys. Acta</i> , <b>1697</b> , 243-257 (2004)

**Table S10. Trained compounds that share maxMCS with hit compounds. A.** Compound list for GPCRs. **B.** Compound list for kinases. See Supplementary Table S4 for the other chemical information of the numbered compounds. This table in HTML format is available at <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/scaf.php>.

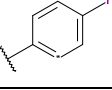
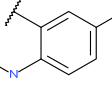
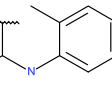
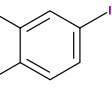
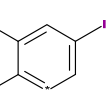
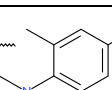
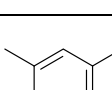
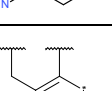
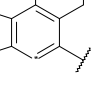
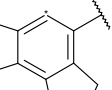
### A

Target GPCR	Active compound	Max-MCS hits in trained compounds	Known targets for the trained compounds
ADRB2	<b>9</b> ((2-aminophenyl)(4-methylphenyl)amine)	<b>10</b>	ADA1-2, 5HT1B, 5HT2
	<b>8</b> (1-( <i>tert</i> -butylamino)-3-[(2-methyl-1H-indol-4-yl)oxy]-2-propanol)	<b>8, 7, 60</b>	ADRB2-3, 5HT1D, 5HT1A, 5HT1B
	<b>33</b> (1-[(3-Methoxypropyl)amino]-3-[(2-methyl-1H-indol-4-yl)oxy]-2-propanol)	<b>8, 7, 60</b>	ADRB2-3, 5HT1D, 5HT1A, 5HT1B
	<b>29</b> (N-( <i>tert</i> -butyl)-N'-(4-methoxybenzyl)thiourea)	<b>127</b>	HRH1, 5HT2A
	<b>32</b> (2-Morpholino-2-oxoacetohydrazide)	<b>132</b>	PTAFR
	<b>28</b> (2-[2,5-Dimethyl-4-(morpholinomethyl)phenoxy]acetamide)	<b>130</b>	CCR5
	<b>30</b> (2-Ethyl-2-[[2-(2-fluorobenzyl)oxy]methyl]-5,5-dimethyltetrahydrofuran)	<b>129, 56, 72, 140, 141, 142, 143, 144</b>	ACM1-5, DRD1-5, HRH1-2, 5HT1R, ADA2B, 5HT4R, 5HT2A, AGTR1-2, ADA2A, TA2R, 5HT7R, ADRB1, 5HT6R
	<b>5</b> (2,5-Dimethyl-1-(2,2,4-trimethyl-2,3-dihydro-1-benzofuran-7-yl)-1H-pyrrole)	<b>6, 145, 146, 147</b>	OPRD, OPRK, OPRM
	<b>31</b> (2-(3-Isopropoxyphenyl)-1-ethanamine)	<b>125, 53, 58, 59, 61, 71, 148, 149, 150, 151</b>	5HT1F, ADA1, ADRB1, 5HT2R, 5HT1E, AA2AR
NPY1R	<b>15</b> (Diethyl 2-(3,5-dimethyl-1H-pyrazol-1-yl)-6-hydroxy-3,5-pyridinedicarboxylate)	<b>16, 152, 153, 154, 155</b>	CCKAR, ADA1A, 5HT2A
	<b>34</b> (Ethyl 4-chloro-1-(4-chlorophenyl)-3-methyl-1H-pyrazolo[3,4-b]pyridine-5-carboxylate)	<b>13</b>	AA1R
	<b>12</b> (ethyl 1-(4-chlorophenyl)-4-[(4-methoxybenzyl)amino]-3-methyl-1H-pyrazolo[3,4-b]pyridine-5-carboxylate)	<b>13</b>	AA1R

### B

Target kinase	Active compound	Max-MCS hits in trained compounds	Known targets for the trained compounds
EGFR	<b>35</b> (7-Chloro-N-(3-methoxybenzyl)-4-quinazolinamine)	<b>137</b>	AKT1-3
	<b>18</b> (6,7-Dimethoxy-N-phenyl-4-quinazolinamine)	<b>19, 156, 157, 158, 159, 160, 161, 162, 163, 164</b>	VGFR1-3, FGFR1-4, LCK, CSF1R, ERBB24, ABL1, YES, EGFR, JAK3, SRC, PAK1, FLT3, RIPK2
	<b>36</b> (6,7-Dimethoxy-N-(2-thienylmethyl)-4-quinazolinamine)	<b>139</b>	TGFR1
	<b>21</b> (N-{2-[(4-Chlorophenyl)sulfanyl]ethyl}-6,7-dimethoxy-4-quinazolinamine)	<b>22, 165, 166, 167</b>	AKT1-3, TIE2, VGFR2
	<b>37</b> (2-[[4-(2-Chloroacetyl)-1H-pyrrol-2-yl]methylene]malononitrile)	<b>135, 168, 169</b>	PDPK1, CDK2
CDK2	<b>26</b> ([4-Amino-2-( <i>tert</i> -butylamino)-1,3-thiazol-5-yl](4-chlorophenyl)methanone)	<b>27</b>	GSK3H, GSK3A, GSK3B
	<b>23</b> ([2-(4-Fluorophenyl)-5,6,7,8-tetrahydroimidazo[2,1-b][1,3]benzothiazol-3-yl]methanol)	<b>24</b>	TGFR1-2

**Table S12. Substructures characteristic of ligand promiscuity for kinases.** The ten top-ranked frequent substructures represented by ECFP fingerprints are shown based on the highest posterior probabilities. See Supplementary Methods (Analysis of ligand promiscuity) for the detail calculation, and Supplementary Table S4 for the chemical names and structures of the numbered compounds.

ECFP pattern	Probability	Compound No.	Subfamily of target kinases
	0.985	402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432	STE, CMGC, TK
	0.985	402, 403, 404, 405, 406, 408, 409, 410, 411, 412, 413, 414, 415, 418, 419, 420, 421, 422, 423, 425, 426, 427, 428, 429, 430, 431, 432	STE
	0.983	402, 403, 404, 405, 406, 433, 434, 408, 409, 410, 411, 412, 435, 436, 414, 415, 418, 419, 420, 421, 422, 423, 425, 426, 427, 428	STE
	0.982	402, 403, 404, 405, 406, 408, 409, 410, 411, 412, 414, 415, 417, 418, 419, 420, 421, 422, 423, 425, 426, 427, 428	STE
	0.982	402, 403, 404, 405, 406, 408, 409, 410, 411, 412, 414, 415, 417, 418, 419, 420, 421, 422, 423, 425, 426, 427, 428	STE
	0.982	402, 403, 404, 405, 406, 408, 409, 410, 411, 412, 414, 415, 418, 419, 420, 421, 422, 423, 425, 426, 427, 428	STE
	0.982	402, 403, 404, 405, 406, 408, 409, 410, 411, 412, 414, 415, 418, 419, 420, 421, 422, 423, 425, 426, 427, 428	STE
	0.977	437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453	CAMK, AGC, TKL, TK, CMGC
	0.977	437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453	CAMK, AGC, TKL, TK, CMGC
	0.975	437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 451, 452, 453	CAMK, AGC, TKL, TK, CMGC

**Table S13. Scores and bioactivity of Bionet compounds ranked by CGBVS for kinase inhibitor screening.** The 20 top-ranked compounds, ordered by CGBVS score for EGFR (A) and CDK2 (B), are shown. IC<sub>50</sub> was determined by the off-chip mobility shift assay. See Supplementary Table S4 for the other chemical information of the numbered compounds. This table in HTML format is available at <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/egfr.php#cgb> (A) and <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/cdk2.php#cgb> (B).

### A

Compound No.	Bionet ID	Compound name	CGBVS score	NN-PCA score	GOLD score	IC <sub>50</sub> (μM)
170	8X-0291	ethyl 2-[[1-(2-cyano-3-nitrilo-1-propenyl)-2-naphthyl]oxy]acetate	1.000	0.751	25.6	
37	4X-0854	2-[[4-(2-chloroacetyl)-1H-pyrrol-2-yl]methylene]malononitrile	1.000	0.878	20.3	13
171	11F-321S	2-benzyl-4-quinazolinethiol	1.000	0.848	25.5	
172	4W-0319	2-[4-(3,4-dichlorophenyl)-3,4-dihydro-1(2H)-naphthalenyliден]malononitrile	0.999	0.670	32.0	
18	11N-058	6,7-dimethoxy-N-phenyl-4-quinazolinamine	0.999	0.935	27.3	0.014
173	9M-521S	2-{1-[3-(trifluoromethyl)phenyl]ethylidene}malononitrile	0.999	0.886	24.4	
174	5P-531S	2,6-di( <i>tert</i> -butyl)-4-(5-sulfanyl-1,3,4-thiadiazol-2-yl)benzenol	0.999	0.902	27.8	
175	10T-0340	2-[(E)-1-(1,3-benzodioxol-5-yl)-3-(dimethylamino)-2-propenylidene]malononitrile	0.999	0.807	29.7	
176	12R-0634	2-[[1-(4-acetylphenyl)-1H-pyrrol-2-yl]methylene]malononitrile	0.999	0.845	29.3	
177	10E-980	2-methyl-3-(2-oxo-2-phenylethyl)naphthoquinone	0.999	0.758	31.6	
178	12W-0294	(E)-2-cyano-3-(4-oxo-4H-chromen-3-yl)-2-propenethioamide	0.999	0.883	31.9	
21	12N-063	N-{2-[(4-chlorophenyl)sulfanyl]ethyl}-6,7-dimethoxy-4-quinazolinamine	0.999	0.866	29.2	0.32
179	7X-0312	2-[[2-(allyloxy)-1-naphthyl]methylene]malononitrile	0.999	0.708	28.2	
180	2P-814	(Z)-4-(4-hydroxyphenyl)-4-oxo-2-butenic acid	0.999	0.968	23.9	
35	10N-835	7-chloro-N-(3-methoxybenzyl)-4-quinazolinamine	0.999	0.873	28.0	1.7
181	7G-002	3-(2-chlorophenyl)-2,3-diphenylacrylonitrile	0.999	0.716	36.6	
182	7X-0314	3-[2-(allyloxy)-1-naphthyl]-2-cyano-2-propenethioamide	0.998	0.750	31.0	
36	12N-055	6,7-dimethoxy-N-(2-thienylmethyl)-4-quinazolinamine	0.998	0.852	25.3	0.035
183	MS-6323	(E)-3-(4,5-dimethoxy-2-nitrophenyl)-2-(2,2-dimethylpropanoyl)-2-propenenitrile	0.998	0.780	26.2	
184	7N-759	(E)-1,4-bis(4-methoxyphenyl)-2-butene-1,4-dione	0.998	0.843	29.4	

**B**

Compound No.	Bionet ID	Compound name	CGBVS score	NN-PCA score	GOLD score	IC <sub>50</sub> (μM)
185	6T-0881	2-(2,5-dimethyl-1H-pyrrol-1-yl)-5,7-dimethyl-6-[(E)-3-phenyl-2-propenyl][1,2,4]triazolo[1,5-a]pyrimidine	0.998	0.595	-28.7	
186	2R-1182	3-(4-chlorophenyl)-3-hydroxy-1-(2-methylimidazo[1,2-a]pyridin-3-yl)-1-propanone	0.996	0.729	-31.9	
187	10R-1000	1-(3-methylimidazo[2,1-b][1,3]thiazol-2-yl)ethyl N-(4-chlorophenyl)carbamate	0.996	0.789	-27.9	
188	6R-0827	1-(6-chloro-2-pyridinyl)-4-(1H-1,2,4-triazol-1-yl)-1H-pyrazol-5-yl N,N-dimethylcarbamate	0.996	0.781	-22.4	
189	10E-944	1-Phenyl-1'-(2-hydroxyethyl)-3,3'-dimethyl-5-hydroxy[4,5']-bipyrazol	0.996	0.725	-25.4	
190	MS-6003	2-[1-(1H-1,2,4-triazol-3-ylamino)ethylidene]-1H-indene-1,3(2H)-dione	0.996	0.797	-24.8	
191	10E-909	1-{1-[4-(2,4-dichlorophenyl)-1,3-thiazol-2-yl]-5-hydroxy-3-methyl-1H-pyrazol-4-yl}-3-hydroxy-2-buten-1-one	0.993	0.527	-34.4	
192	4R-0644	1-[(2-chloro-1,3-thiazol-5-yl)methyl]-4,6-dimethyl-3-(1H-pyrrol-1-yl)-1H-pyrazolo[3,4-b]pyridine	0.991	0.702	-26.0	
193	12W-0882	N-(4,6-dimethyl-2-pyrimidinyl)-1H-1,3-benzimidazol-2-amine	0.990	0.844	-26.8	
194	10P-382S	ethyl 1-phenyl-5-(1H-pyrrol-1-yl)-1H-pyrazole-4-carboxylate	0.989	0.884	-27.7	
195	10E-975	1-(4-Chlorophenyl)-1'-H-3,3'-dimethyl-5-hydroxy[4,5']-bipyrazol	0.989	0.780	-25.7	
23	MS-2894	[2-(4-fluorophenyl)-5,6,7,8-tetrahydroimidazo[2,1-b][1,3]benzothiazol-3-yl]methanol	0.987	0.780	-30.4	19
196	2P-526S	ethyl 5-amino-1-[(4-methylphenyl)sulfonyl]-1H-pyrazole-4-carboxylate	0.985	0.753	-23.4	
197	11X-0828	2-[(2-piperidino-1,3-thiazol-5-yl)methylene]-1H-indene-1,3(2H)-dione	0.985	0.597	-30.3	
198	1R-1190	1-(2-methylimidazo[1,2-a]pyridin-3-yl)-1-ethanone	0.983	0.934	-22.8	
199	7D-110	5-nitroimidazo[2,1-b][1,3]thiazol-6-yl phenyl ether	0.982	0.836	-20.5	
200	5C-020	4-[(1,3,5-trimethyl-1H-pyrazol-4-yl)methyl]phenyl N-phenylcarbamate	0.982	0.771	-33.6	
201	10T-0814	N-(2-methoxybenzyl)-2-[3-methyl-5-(trifluoromethyl)-1H-pyrazol-1-yl]-1,3-thiazole-4-carboxamide	0.981	0.713	-22.5	
202	10T-0811	{2-[3-methyl-5-(trifluoromethyl)-1H-pyrazol-1-yl]-1,3-thiazol-4-yl}(4-phenylpiperazino)methanone	0.980	0.626	-31.1	
26	7N-773	[4-amino-2-( <i>tert</i> -butylamino)-1,3-thiazol-5-yl](4-chlorophenyl)methanone	0.980	0.763	-29.1	4.9



**Table S14. Scores and bioactivity of Bionet compounds ranked by LBVS for kinase inhibitor screening.** The 20 top-ranked compounds, ordered by LBVS score (NN-PCA score) for EGFR (**A**) and CDK2 (**B**), are shown. IC<sub>50</sub> was determined by the off-chip mobility shift assay. See Supplementary Table S4 for the other chemical information of the numbered compounds. This table in HTML format is available at <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/egfr.php#lb> (A) and <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/cdk2.php#lb> (B).

### A

Compound No.	Bionet ID	Compound name	CGBVS score	NN-PCA score	GOLD score	IC <sub>50</sub> (μM)
203	6K-003	2-[(4-methoxyphenyl)methylene] malononitrile	0.991	1.000	23.5	
204	5T-0638	6-nitro-4(3H)-quinazolinone	0.944	1.000	25.0	
205	6X-0926	4-oxo-4-phenylbutanenitrile	0.978	0.996	22.5	
206	6X-0945	4-(4-chlorophenyl)-4-oxobutanenitrile	0.997	0.996	24.0	
207	6K-004	2-(4-methoxybenzyl)malononitrile	0.978	0.995	23.7	
208	MS-3533	3-(4-fluorophenyl)propanoic acid	0.889	0.994	20.4	
209	5F-037	2-(1H-pyrrol-2-ylmethylene)malononitrile	0.933	0.988	21.2	
210	12A-107	2-(2-oxo-2-phenylethyl)malononitrile	0.997	0.987	22.6	
211	7N-749	methyl (E)-4-(4-chlorophenyl)-4-oxo-2-butenolate	0.994	0.985	26.2	
212	6C-009	1-(5-fluoro-2-hydroxyphenyl)-1-ethanone	0.976	0.982	22.2	
213	7X-0923	4-(4-methoxyphenyl)-4-oxobutanenitrile	0.996	0.981	26.2	
214	1X-0848	2-[(3-chloro-4-methoxyphenyl)methylene]malononitrile	0.996	0.980	25.1	
215	12W-5015	2,5-dihydroxy-3-methoxybenzaldehyde	0.905	0.979	18.2	
216	8M-555S	(E)-2-cyano-3-(2-furyl)-2-propenethioamide	0.845	0.979	26.4	
217	7N-761	(E)-4-(4-methylphenyl)-4-oxo-2-butenic acid	0.996	0.978	28.6	
218	9F-313S	6-methyl-2-(2-pyridinyl)-4-pyrimidinol	0.476	0.978	20.3	
219	12E-047	2-[1-amino-2-(1H-1,2,4-triazol-1-yl)ethylidene]malononitrile	0.909	0.977	14.5	
220	MS-2010	4-[(E)-3-hydroxy-1-propenyl]phenol	0.795	0.977	22.7	
221	MS-6568	2-[2-(2-fluorophenyl)hydrazono]malononitrile	0.876	0.976	23.1	
222	5P-039	2-methyl-1-naphthalenamine	0.694	0.976	26.0	

## B

Compound No.	Bionet ID	Compound name	CGBVS score	NN-PCA score	GOLD score	IC <sub>50</sub> (μM)
223	7X-0868	imidazo[1,2-a]pyridine-6-carbohydrazide	0.700	0.966	24.3	
224	7R-0016	6-chloro-2,5-dimethylpyrazolo[1,5-a]pyrimidin-7-ol	0.938	0.966	21.6	
225	11A-032	2-(1H-imidazol-1-yl)benzenecarboxamide	0.715	0.965	24.0	
226	JS-2883	5-methoxy-2-[(5-methyl-3-isoxazolyl)amino]methyl}benzenol	0.207	0.962	26.5	
227	2W-0306	methyl 3-amino-1-benzofuran-2-carboxylate	0.556	0.961	25.1	
228	7W-0880	1,3-benzodioxole-5-carbohydrazide	0.250	0.960	23.1	
229	7T-1502	methyl 1H-indole-3-carboxylate	0.719	0.959	23.3	
230	6K-546S	3-methyl-1,2,3-benzotriazin-4(3H)-one	0.704	0.958	21.9	
231	(3N-012	5-methyl[1,6]naphthyridin-2(1H)-one	0.691	0.958	28.4	
232	1B-028	1-[2-(1H-imidazol-1-yl)phenyl]-1-ethanone	0.903	0.958	24.4	
233	3L-391S	[2-(4-methylphenoxy)-3-pyridinyl]methanol	0.065	0.957	24.9	
234	3C-136	2-(methylsulfanyl)[1,2,4]triazolo[1,5-a]pyrimidin-7-amine	0.916	0.956	15.5	
235	1F-936	ethyl 3-amino-1H-indole-2-carboxylate	0.677	0.956	24.8	
236	MS-3537	2-ethoxy-N-(9H-purin-6-yl)propanamide	0.872	0.956	12.9	
237	3X-0853	ethyl 2-amino-1,3-benzothiazole-6-carboxylate	0.855	0.955	23.6	
238	1X-0007	1-ethyl-1,3-dihydro-2H-indol-2-one	0.714	0.955	23.2	
239	2X-0870	2-(1,3-benzoxazol-5-yl)acetic acid	0.909	0.955	21.0	
240	3R-0006	6-methyl[1,2,4]triazolo[4,3-b]pyridazin-8-ol	0.924	0.954	17.2	
241	9X-0852	2-methoxy-4-(1,3-oxazol-5-yl)benzenol	0.679	0.954	25.8	
242	2X-0807	2-(4-hydroxy-3-nitrophenyl)acetic acid	0.488	0.953	21.9	

**Table S15. Scores and bioactivity of Bionet compounds ranked by SBVS for kinase inhibitor screening.** The 20 top-ranked compounds, ordered by SBVS score (GOLD score) for EGFR (**A**) and CDK2 (**B**), are shown. IC<sub>50</sub> was determined by the off-chip mobility shift assay. See Supplementary Table S4 for the other chemical information of the numbered compounds. This table in HTML format is available at <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/egfr.php#sb> (A) and <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/cdk2.php#sb> (B).

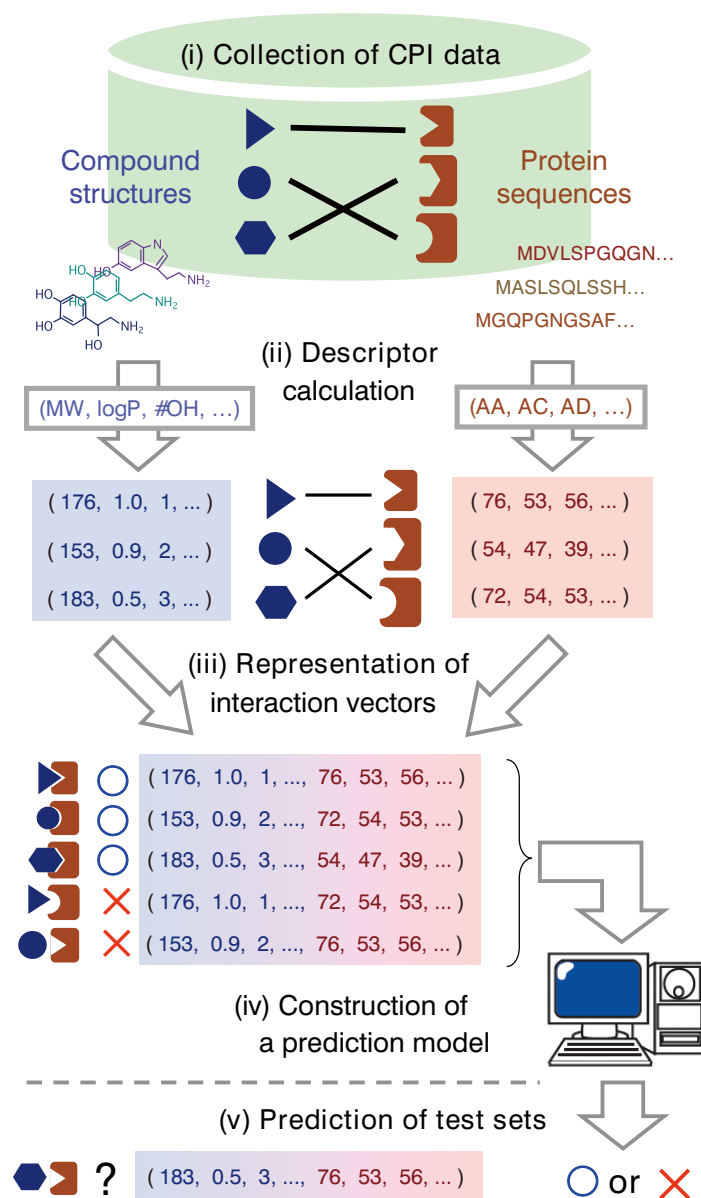
## A

Compound No.	Bionet ID	Compound name	CGBVS score	NN-PCA score	GOLD score	IC <sub>50</sub> (μM)
243	6X-0279	1-(2,5-dimethylbenzyl)-4-[5-(1-naphthyl)-1H-pyrazol-3-yl]piperidine	0.107	0.756	47.8	
244	4R-1029	(Z)-3-(dimethylamino)-1-(4-methyl-2-phenyl-1,3-thiazol-5-yl)-2-(phenylsulfonyl)-2-propen-1-one	0.400	0.778	46.1	
245	2W-0303	4-[(E)-2-(2-nitrophenyl)ethenyl]-8-phenoxy-2,3-dihydrothieno[3,2-c]quinoline	0.814	0.806	45.2	
246	6R-1523	2-[(4-chlorobenzyl)sulfanyl]-3-quinolinecarbaldehyde O-(4-methoxybenzyl)oxime	0.598	0.608	45.1	
247	7T-0323	2-[(E)-2-anilinoethenyl]-4-(2-fluorophenyl)-5-oxo-5H-indeno[1,2-b]pyridine-3-carbonitrile	0.005	0.720	44.5	
248	3W-0326	4-[(E)-2-(4-chlorophenyl)ethenyl]-8-phenoxy-2,3-dihydrothieno[3,2-c]quinoline	0.881	0.802	43.7	
249	10X-0276	4-[(E)-2-[4-(benzyloxy)phenyl]ethenyl]-6-phenoxy-2,3-dihydrothieno[3,2-c]quinoline	0.571	0.865	43.7	
250	3R-0816	2-[(4-methyl-2-phenyl-1,3-thiazol-5-yl)carbonyl]-3-phenyl-1,1-cyclopropanedicarbonitrile	0.165	0.687	43.4	
251	8M-302S	N-{3-[(4-chlorophenyl)sulfonyl]-4-methyl-6-phenyl-2-pyridinyl}-N-(4-methoxybenzyl)amine	0.398	0.786	43.4	
252	12W-0269	2-[[1-(4-chlorophenyl)-1H-pyrazol-4-yl]carbonyl]phenyl benzenesulfonate	0.034	0.743	43.3	
253	3P-066	3-(4-fluorophenyl)-1H-pyrazole-4-carbaldehyde N-(7-chloro-4-quinolinyl)hydrazone	0.098	0.597	43.0	73
254	SS-0777	6-( <i>tert</i> -butyl)-2-[(3-methoxybenzyl)sulfanyl]-4-(4-methoxyphenyl)-5,6,7,8-tetrahydro-3-quinolinecarbonitrile	0.309	0.782	42.9	
255	6P-733	3-(2,4-dichlorobenzyl)-4-hydroxy-1-[2-(1-naphthylamino)ethyl]-2(1H)-pyridinone	0.529	0.584	42.8	
256	11K-906	4-(2,6-dichlorophenyl)-2-[2-(4-fluoroanilino)vinyl]-6-methyl-3,5-pyridinedicarbonitrile	0.119	0.542	42.7	
257	1T-0830	2-[[E)-(2-nitrophenyl)methylidene]amino]-4,5-diphenyl-3-furonitrile	0.117	0.699	42.7	
258	8M-322S	4-chlorophenyl 2-(cyclohexylsulfanyl)-4-methyl-6-phenyl-3-pyridinyl sulfone	0.679	0.770	42.6	
259	10L-334S	4-[(7-benzyl-5,6-diphenyl-7H-pyrrolo[2,3-d]pyrimidin-4-yl)sulfanyl]phenyl methyl ether	0.977	0.846	42.6	
260	12P-107	2-[2-({1-[(4-methylphenyl)sulfonyl]-1H-indol-4-yl}oxy)ethoxy]-N-(2-thienylmethyl)-1-ethanamine	0.061	0.743	42.5	
261	JS-2138	1-methyl-3,5-bis[(3-nitrophenyl)methylene]tetrahydro-4(1H)-pyridinone	0.977	0.711	42.4	
262	4W-0205	(Z)-3-[3-(4-cyclohexylphenyl)-1-phenyl-1H-pyrazol-4-yl]-2-(2,4-dichlorophenyl)-2-propenenitrile	0.185	0.729	41.8	

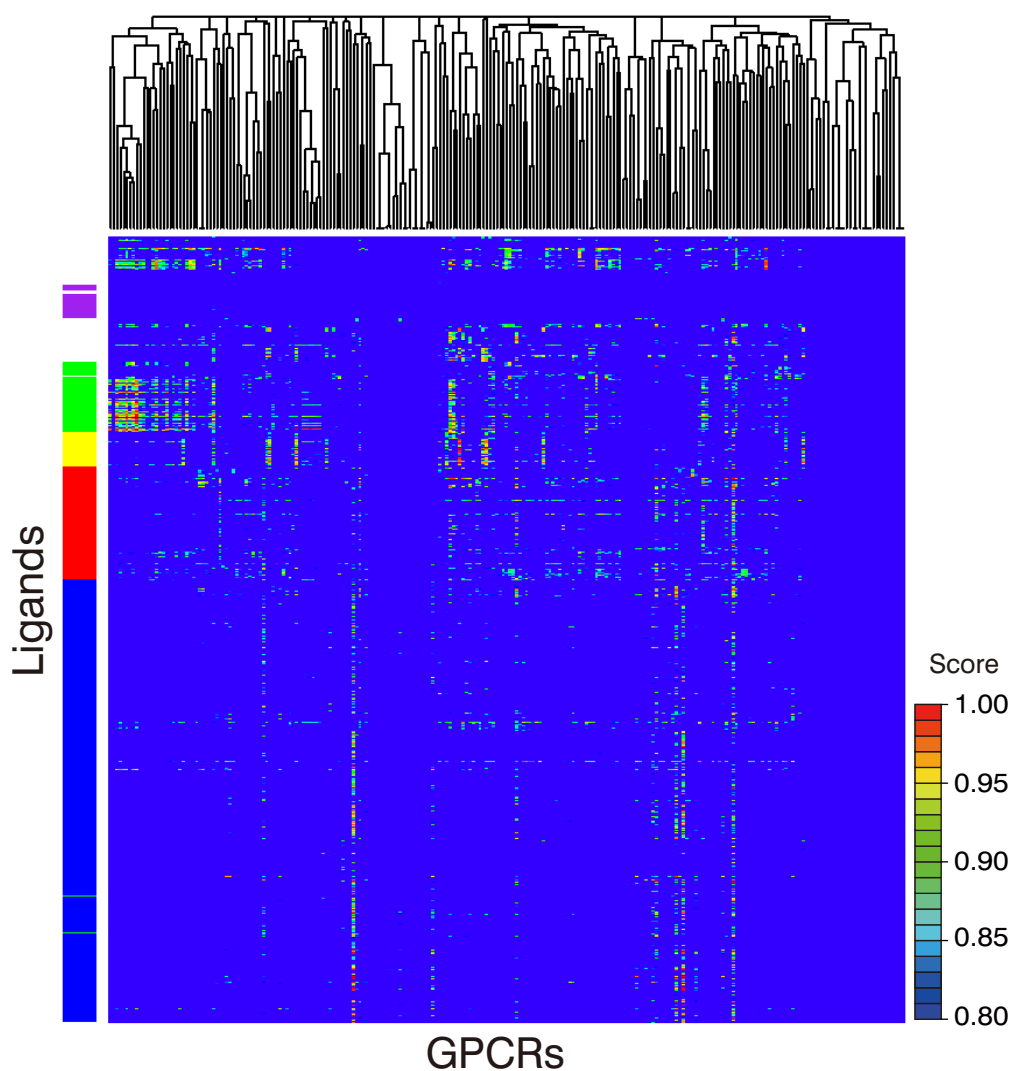
## B

Compound No.	Bionet ID	Compound name	CGBVS score	NN-PCA score	GOLD score	IC <sub>50</sub> (μM)
263	1R-0622	9-benzyl-1,6-diphenyl-9-azadispiro[2.1.2.3]decan-4-one	0.009	0.554	47.0	
264	1H-931	11,13-diphenyl-2a,3,10c,13-tetrahydro-2H-benzo[5',6']chromeno[4',3':4,5]pyrano[2,3-c]pyrazole	0.009	0.673	46.6	
265	10N-037	N-benzoyl-N'-((2,4,6-triphenyl)phenyl)thiourea	0.005	0.554	46.4	
266	10X-0276	4-((1E)-2-[4-(phenylmethoxy)phenyl]vinyl)-6-phenoxy-2,3-dihydrothiopheno[3,2-c]quinoline	0.001	0.603	46.4	
267	JS-2922	N-(4-butylphenyl)-2-[(4-[[3-(2-naphthyl)-3-oxopropyl]amino]phenyl)sulfanyl]acetamide	0.042	0.597	46.0	
268	8X-0002	methyl 4-[4-[5-(2,4-dichlorophenyl)-1H-pyrazol-3-yl]piperidino]benzenecarboxylate	0.488	0.445	45.4	
269	JS-1195	1-[phenyl(piperidino)methyl]-2-naphthyl 4-( <i>tert</i> -butyl)benzenecarboxylate	0.012	0.632	44.9	
270	RS-0182	2-[[3-(4-chlorophenyl)-3-oxo-1-phenylpropyl]amino]-9H-fluoren-9-one	0.836	0.616	44.6	
271	6R-1523	1-[2-[(4-chlorophenyl)methylsulfanyl]quinolin-3-yl]-N-[(4-methoxyphenyl)methoxy]methanimine	0.000	0.395	44.5	
272	7T-0238	N-[4-(4-pentylcyclohexyl)phenyl]-3,4-dihydro-2(1H)-isoquinolinecarboxamide	0.004	0.536	44.4	
273	1Y-0244	4-(2-phenyl-1,3-thiazol-4-yl)phenyl (E)-2-[4-(trifluoromethyl)phenyl]ethenyl sulfone	0.089	0.642	44.2	26
274	10P-927	N-benzoyl-N'-(2-methyl-9,10-dioxo-9,10-dihydro-1-anthracenyl)thiourea	0.020	0.428	44.1	
275	JS-0896	3-benzyl-5-((4-[(2,4-dichlorobenzyl)oxy]phenyl)methylene)-2-thioxo-1,3-thiazolan-4-one	0.026	0.501	43.9	
276	5F-972	2,4-dichloro-N'-(3,3-[spiro-2-adamantyl]-2,3-dihydro-1H-inden-1-ylidene)benzohydrazide	0.413	0.599	43.7	
277	8K-900	5-chloro-1-(4-nitrobenzyl)-3-phenyl-1H-indole-2-carboxamide	0.107	0.716	43.5	
262	4W-0205	(Z)-3-[3-(4-cyclohexylphenyl)-1-phenyl-1H-pyrazol-4-yl]-2-(2,4-dichlorophenyl)-2-propenenitrile	0.054	0.609	43.0	
278	MS-2481	2-nitro-N-[3-(trifluoromethyl)phenyl]dibenzo[b,f][1,4]oxazepine-10(11H)-carboxamide	0.351	0.611	42.9	
279	MS-0259	1-benzyl-1-(4-bromobenzyl)-3-[4-( <i>tert</i> -butyl)benzyl]piperidinium bromide	0.005	0.568	42.7	
280	6R-0825	[1,1'-biphenyl]-4-yl(4-methyl-3-quinolinyl)methanone	0.191	0.664	42.7	
281	3R-0060	(E)-1-[4-[(2,4-dichlorobenzyl)oxy]phenyl]-3-(2-morpholinoanilino)-2-propen-1-one	0.597	0.597	42.6	

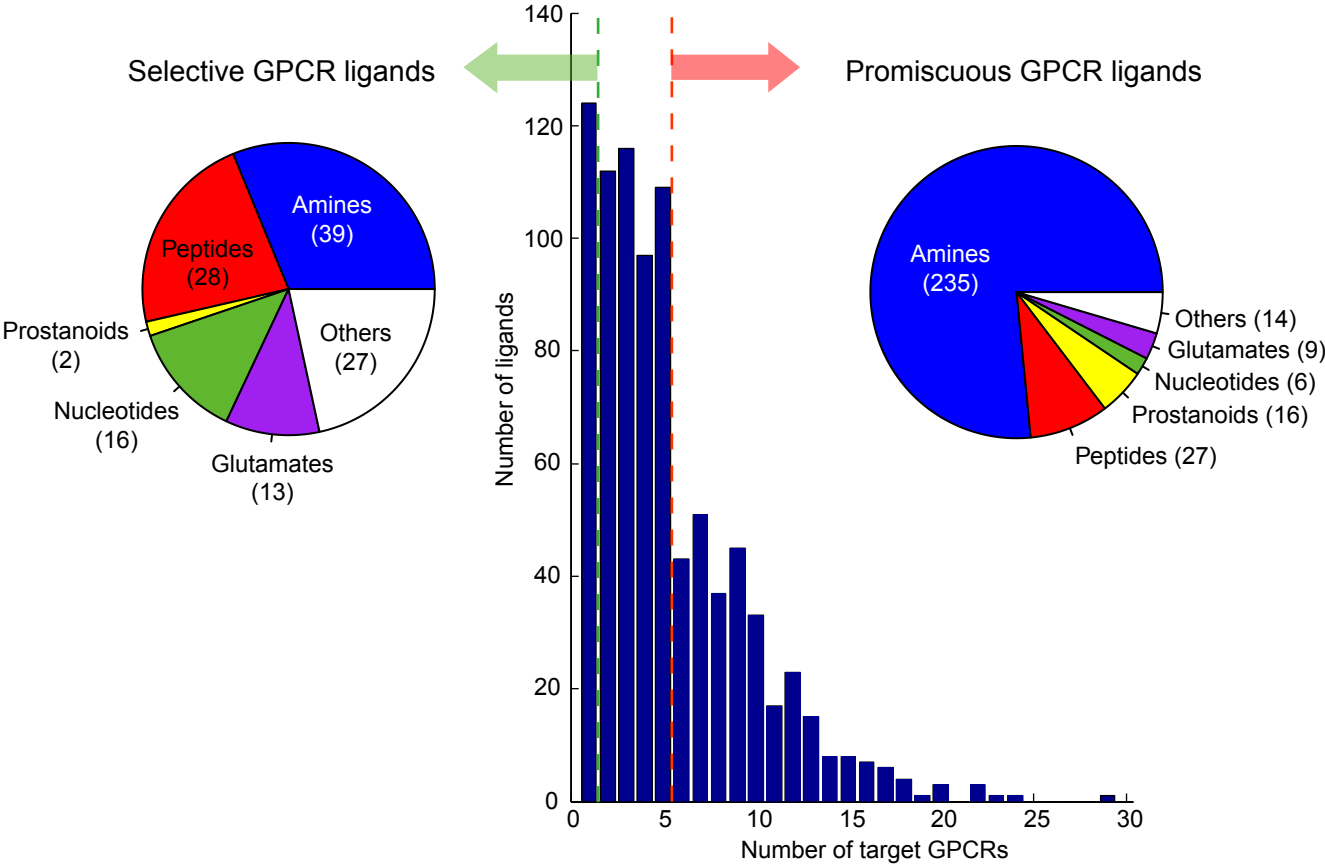
**Figure S1. Outline of the CGBVS method.** The CGBVS strategy consists of five steps: the collection of CPI data, descriptor calculation, the representation of interaction vectors, the construction of a prediction model using training sets, and the prediction of test sets.



**Figure S2. Polypharmacology map between 866 compounds and GPCRs without ligand information.** Prediction scores for all combinations of 866 compounds and 240 GPCRs without ligand information (excluding olfactory receptors) calculated by CGBVS are displayed as a 2D color map with axes representing the compounds (vertical) and the GPCRs (horizontal). The color bars indicate the classes to which the compounds belong (blue = amines, red = peptides, yellow = prostanoids, green = nucleotides, purple = glutamates). GPCRs were clustered by hierarchical clustering based on pairwise alignment scores obtained by BLAST (Altschul *et al.*, 1990). An enlarged map is available at <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/cpimap.php?fam=orphan>.

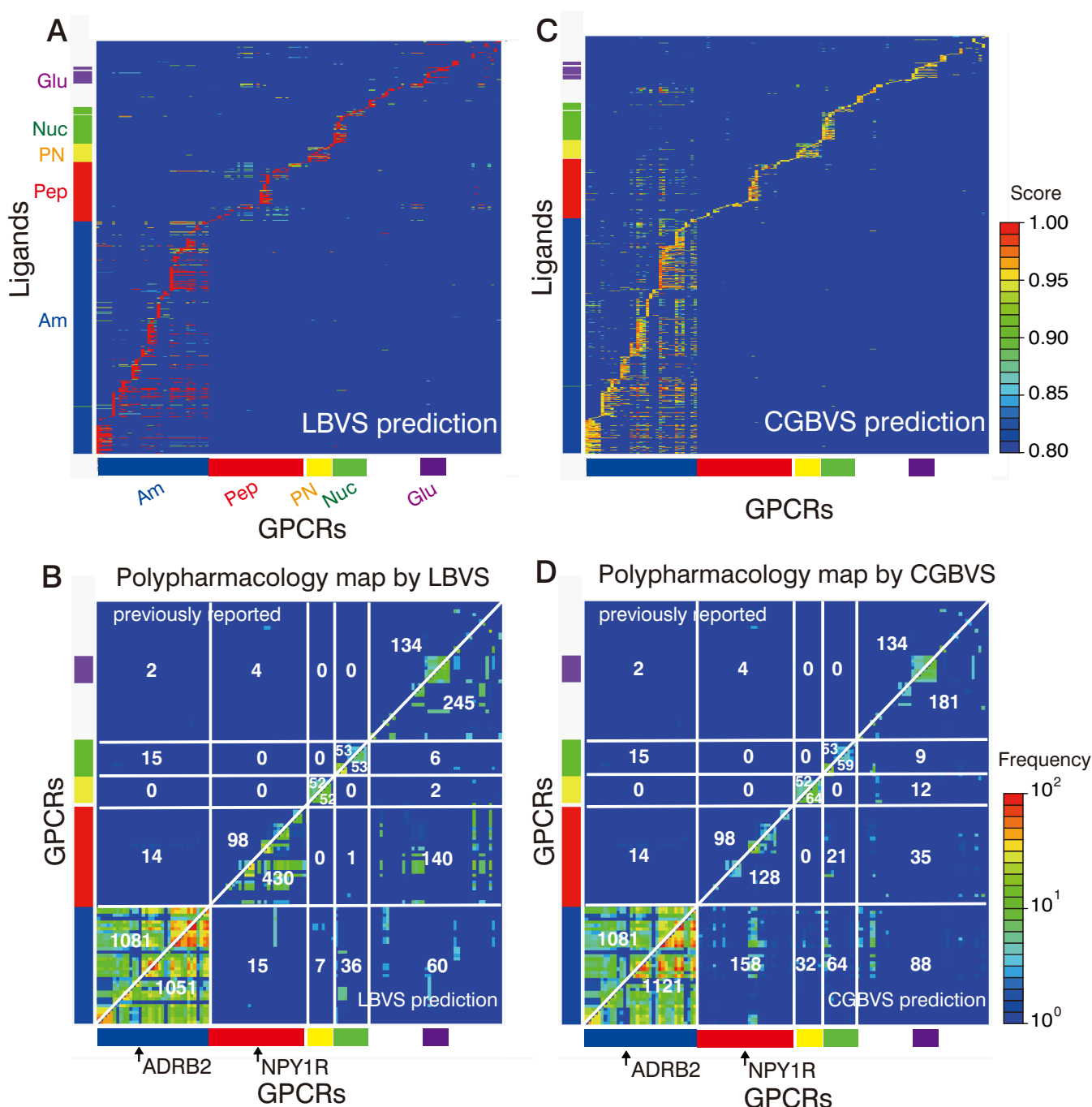


**Figure S3. Statistics of the polypharmacology map of GPCRs.** The distribution of the number of ligands with respect to the number of predicted GPCR targets (center). The breakdown of GPCR subfamilies targeted by selective ligands (with a single target) and those targeted by promiscuous ligands (with more than five targets) is shown in the left and right graphs, respectively.

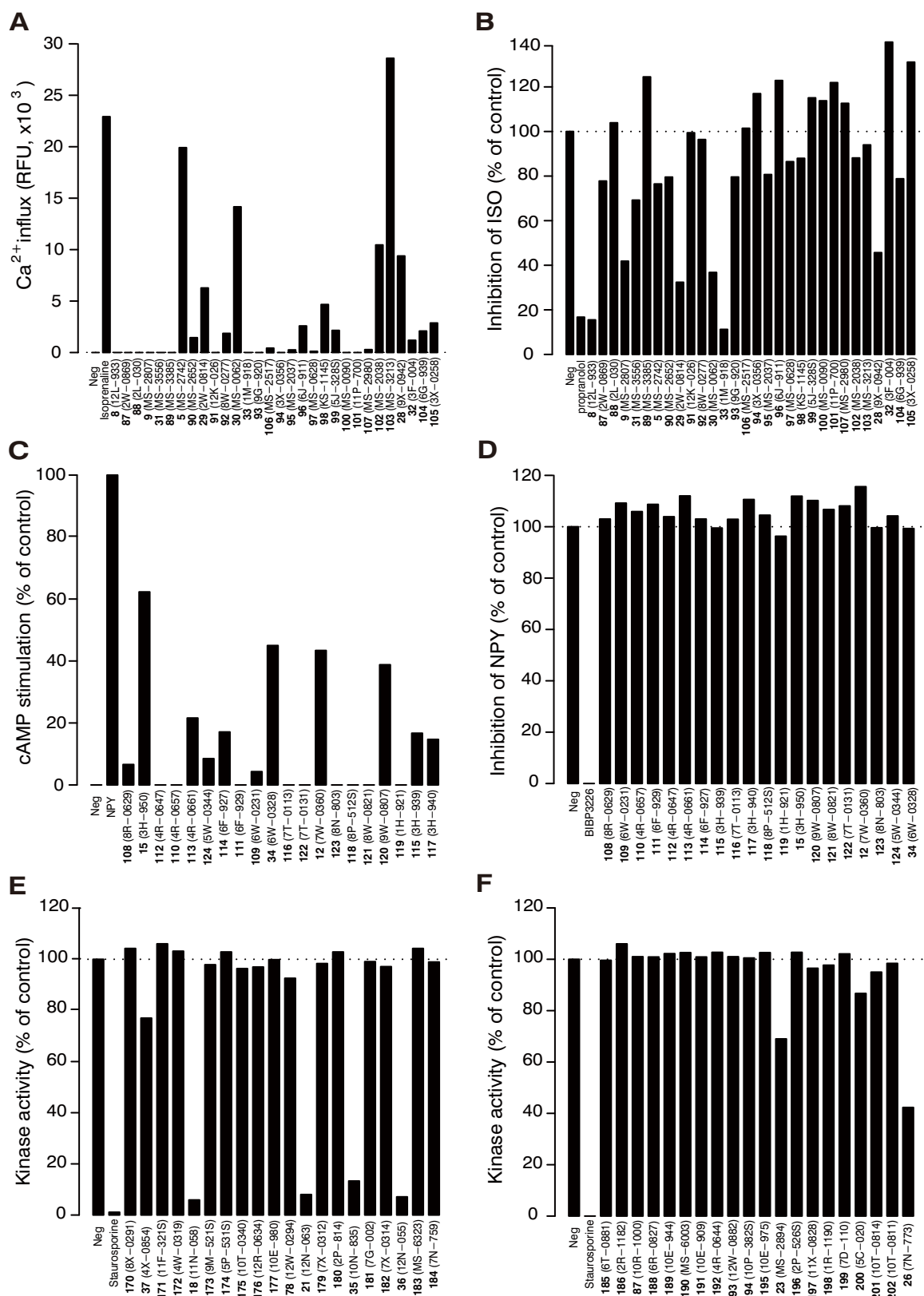




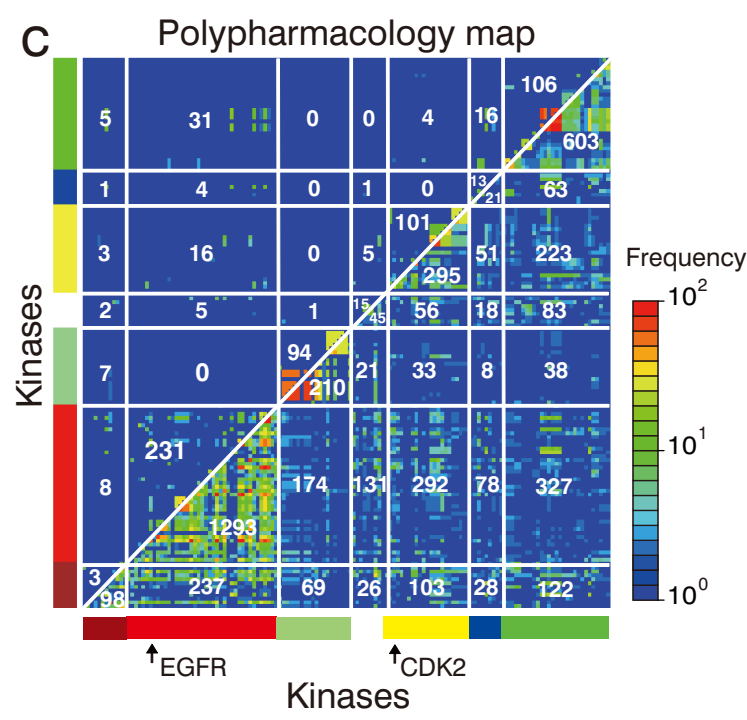
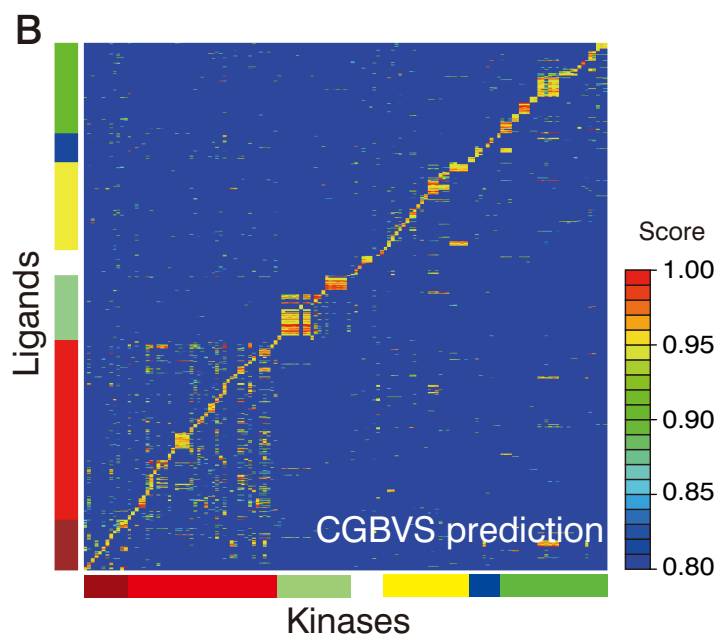
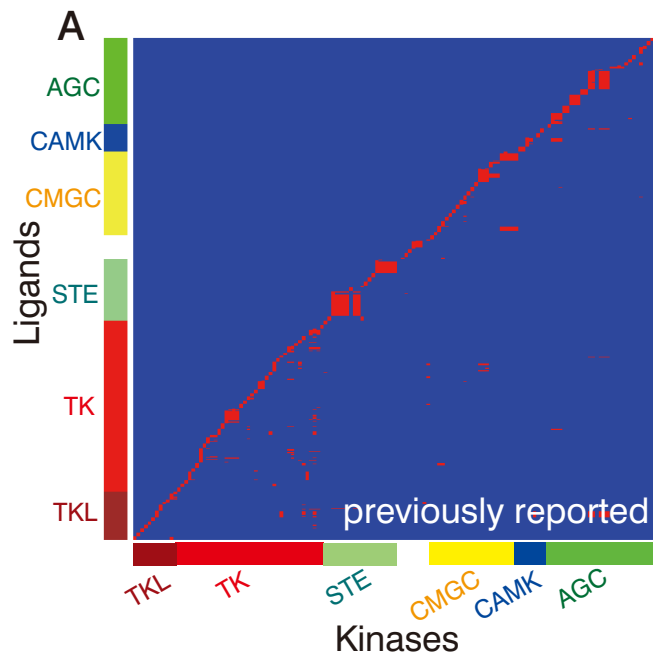
**Figure S4. Polypharmacology map predicted by LBVS. A.** Prediction scores for all compound–protein pairs calculated by NN-PCA are displayed as a 2D color map with axes representing the compounds (vertical) and the proteins (horizontal). The scores are shown in blue (low) to red (high) colors. The color bar and the order of GPCRs and ligands are the same as in Figures 3B and C. Am = amines, Pep = peptides, PN = prostanoids, Nuc = nucleotides, Glu = glutamates. **B.** Degree of intra- and inter- GPCR family promiscuity illustrated as a polypharmacological interaction matrix. The upper and lower triangular parts of the matrix represent the number of known ligands in common and the number of predicted ligands in common, respectively. **C–D.** Polypharmacology maps predicted by CGBVS (as in Figures 3B, C).



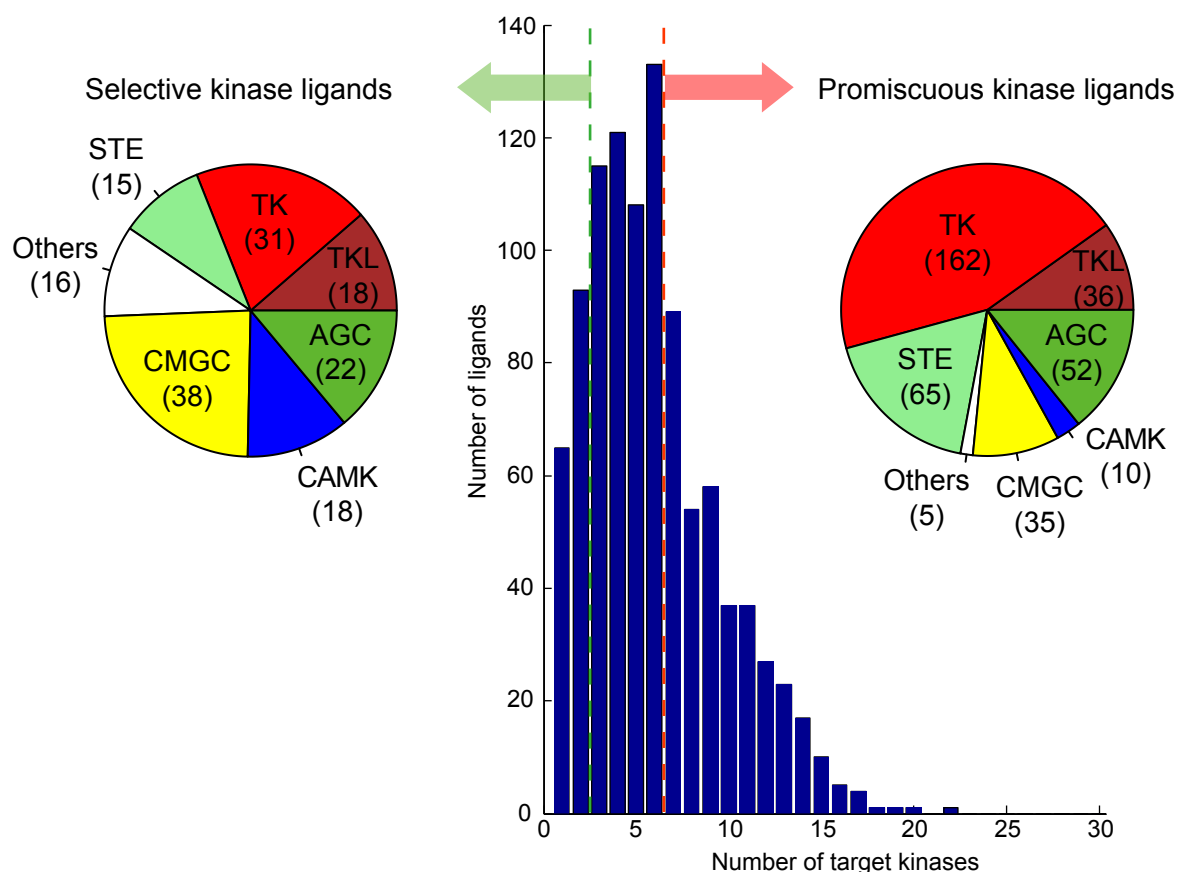
**Figure S5. Primary ligand screening on ADRB2, NPY1R, EGFR and CDK2. A.** ADRB2 agonist activity. **B.** ADRB2 antagonist activity. **C.** NPY1R agonist activity. **D.** NPY1R antagonist activity. **E.** EGFR inhibitor activity. **F.** CDK2 inhibitor activity. See Supplementary Table S4 for the other chemical information of the numbered compounds.



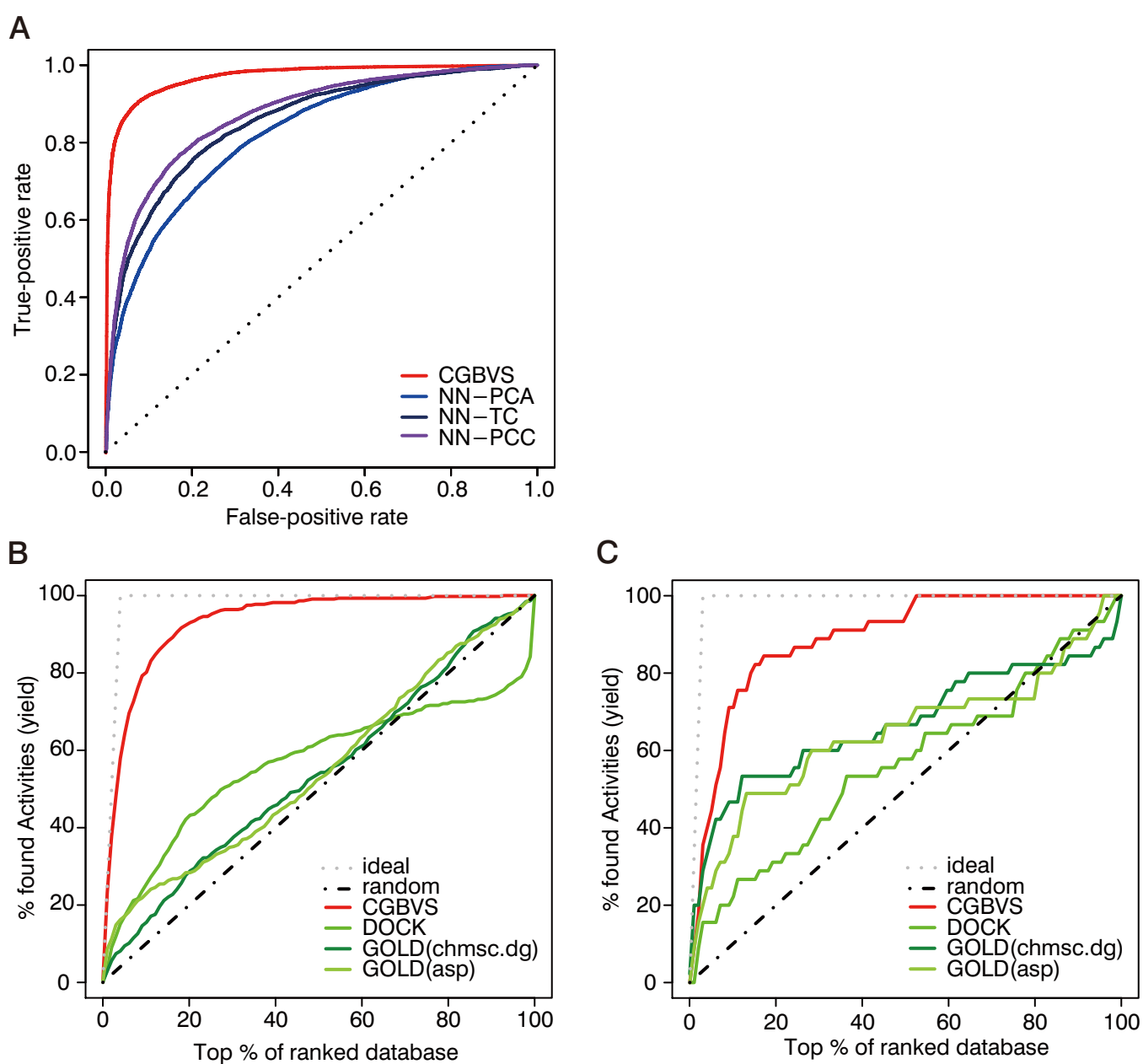
**Figure S6. Polypharmacology maps for kinases predicted by CGBVS. A.** Collective map for previously reported compound (vertical axis)–kinase (horizontal axis) interactions. The reported CPIs are displayed as red colored plots. The colored bars along with each axis indicate the classes to which the compounds and kinase belong. TKL = tyrosine kinase-like; TK = tyrosine kinase; STE = homologs of yeast Sterile 7, Sterile 11, Sterile 20 kinases; CMGC = CDK, MAPK, GSK3, CLK families; CK1 = casein kinase 1; CAMK = calcium/calmodulin-dependent protein kinase; AGC = PKA, PKC, PKC families. **B.** Prediction scores for all compound–kinase pairs calculated by CGBVS. **C.** Degree of intra- and inter-kinase family promiscuity illustrated as a polypharmacological interaction matrix. The upper and lower triangular parts of the matrix represent the number of known ligands in common and the number of predicted ligands in common, respectively. Enlarged maps are available at <http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/cgbvs/polypharm.php>.



**Figure S7. Statistics of the polypharmacology map of kinases.** The distribution of the number of ligands with respect to the number of predicted kinase targets (center). The breakdown of kinase subfamilies targeted by selective ligands (with less than three targets) and those targeted by promiscuous ligands (with more than six targets) is shown in the left and right graphs, respectively. TKL = tyrosine kinase-like; TK = tyrosine kinase; STE = homologs of yeast Sterile 7, Sterile 11, Sterile 20 kinases; CMGC = CDK, MAPK, GSK3, CLK families; CK1 = casein kinase 1; CAMK = calcium/calmodulin-dependent protein kinase; AGC = PKA, PKC, PKC families.

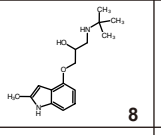
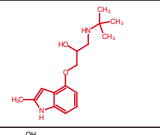
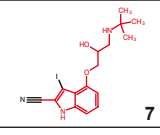
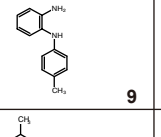
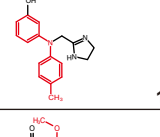
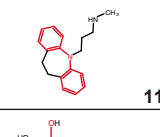
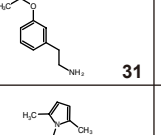
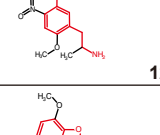
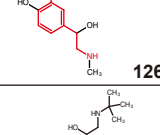
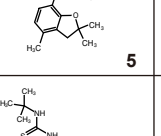
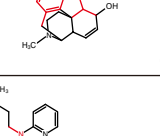
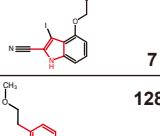
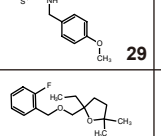
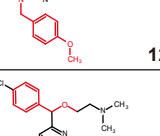
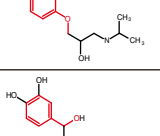
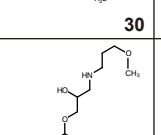
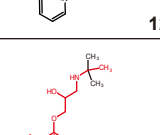
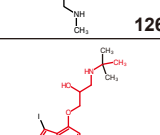
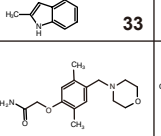
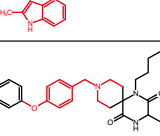
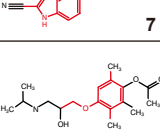
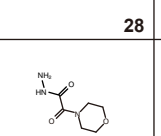
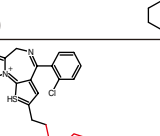
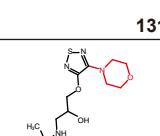
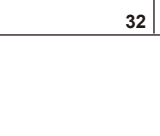
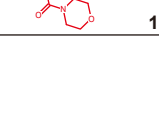
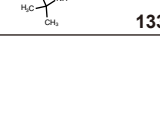


**Figure S8. Performance comparison between CGBVS and conventional methods (LVBS and SBVS) on predicting kinase inhibitors. A.** ROC curves obtained by five-fold CV using the compound–kinase CPIs for the CGBVS and LBVS methods (NN-PCA, NN-TC, NN-PCC). **B–C.** Enrichment curves obtained by the CGBVS and SBVS methods (DOCK and GOLD with two different score functions) for CDK2 (**B**) and EGFR (**C**). Ligand and decoy data sets from the Directory of Useful Decoys (DUD) were used for this evaluation.

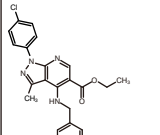
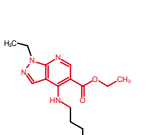
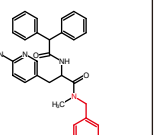
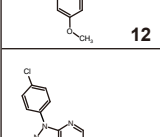
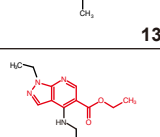
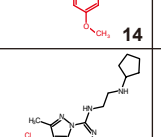
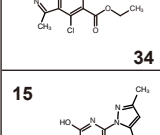
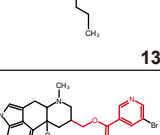
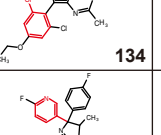


**Figure S9. Max-MCS list for active compounds.** Max-MCSs between active compounds (left) and the most relevant compounds found within the whole trained compound set (center), and that within the known ligand set of each target protein (right) are shown and colored in red. Max-MCS lists for ADRB2 (**A**), NPY1R (**B**), EGFR (**C**), and CDK2 (**D**) are shown. The columns “bioactivity” and “ligand” indicate the existence of publications regarding the compound (See the legends of Figures 4 and 5 for details). See Supplementary Table S4 for the other chemical information of the numbered compounds.

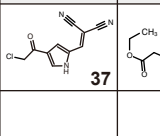
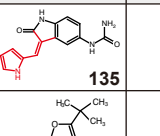
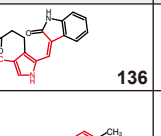
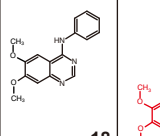
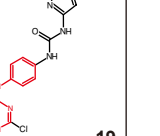
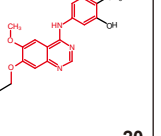
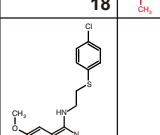
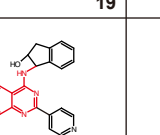
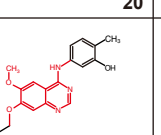
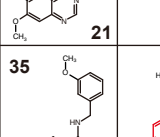
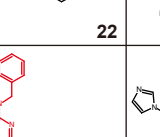
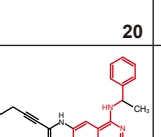
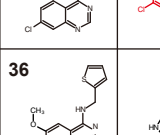
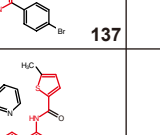
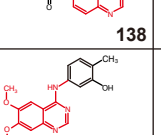
**A**

Active compound	Trained compound sharing max-MCS	ADRB2 known ligand sharing max-MCS	Scaffold-hopped	Previously-reported bioactivity	ligand
 <b>8</b>	 <b>8</b>	 <b>7</b>	No	Yes	Yes
 <b>9</b>	 <b>10</b>	 <b>11</b>	No	Yes	No
 <b>31</b>	 <b>125</b>	 <b>126</b>	No	Yes	No
 <b>5</b>	 <b>6</b>	 <b>7</b>	Yes	No	No
 <b>29</b>	 <b>127</b>	 <b>128</b>	No	No	No
 <b>30</b>	 <b>129</b>	 <b>126</b>	No	No	No
 <b>33</b>	 <b>8</b>	 <b>7</b>	No	Yes	No
 <b>28</b>	 <b>130</b>	 <b>131</b>	No	No	No
 <b>32</b>	 <b>132</b>	 <b>133</b>	No	Yes	No

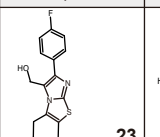
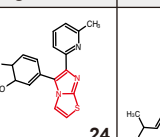
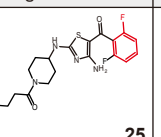
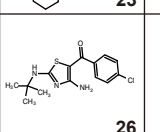
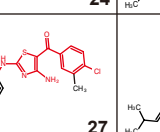
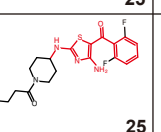
**B**

Active compound	Trained compound sharing max-MCS	NPY1R known ligand sharing max-MCS	Scaffold-hopped	Previously-reported bioactivity	ligand
 <b>12</b>	 <b>13</b>	 <b>14</b>	Yes	No	No
 <b>34</b>	 <b>13</b>	 <b>134</b>	Yes	No	No
 <b>15</b>	 <b>16</b>	 <b>17</b>	No	No	No

**C**

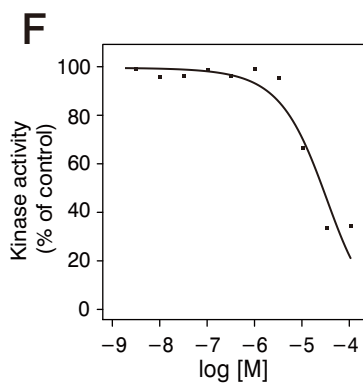
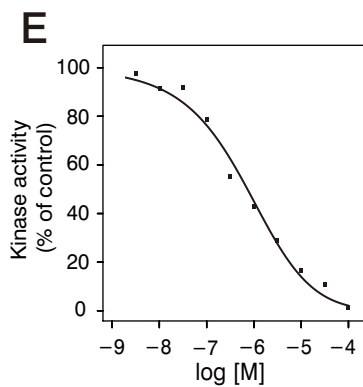
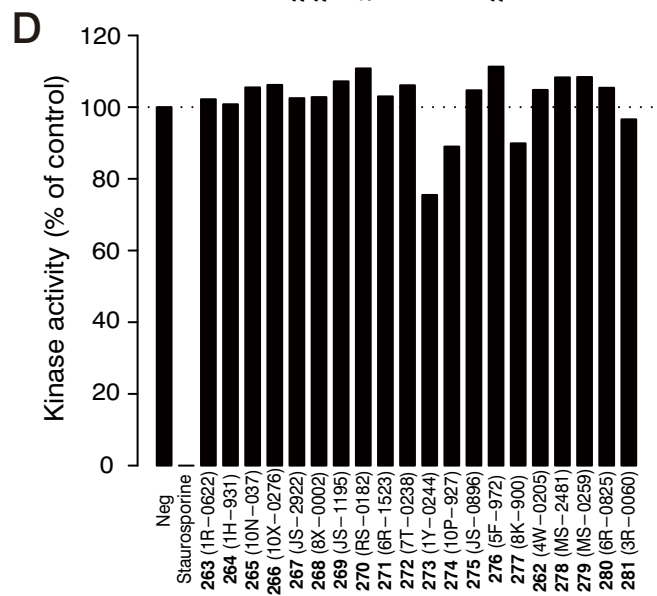
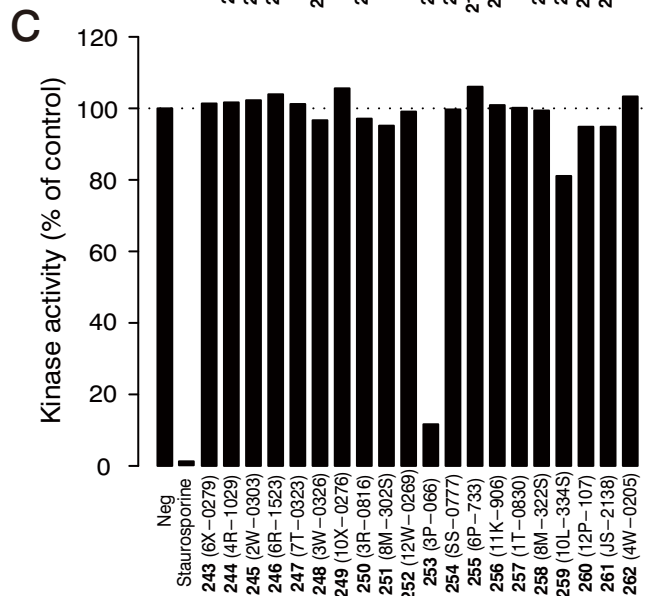
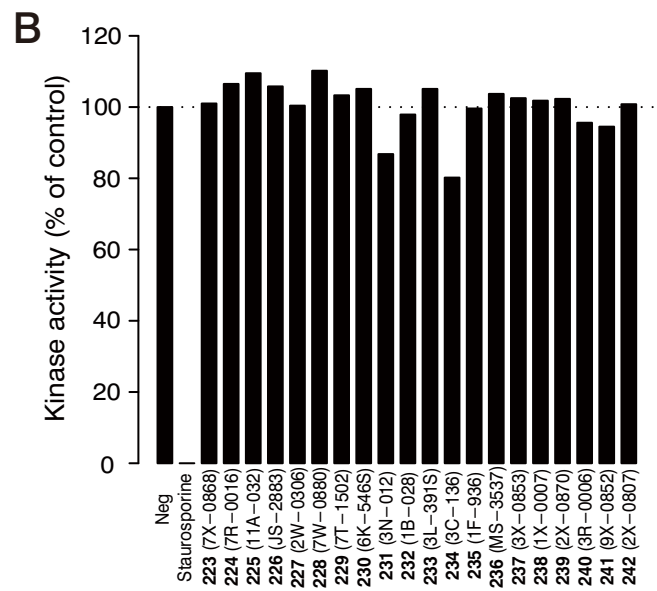
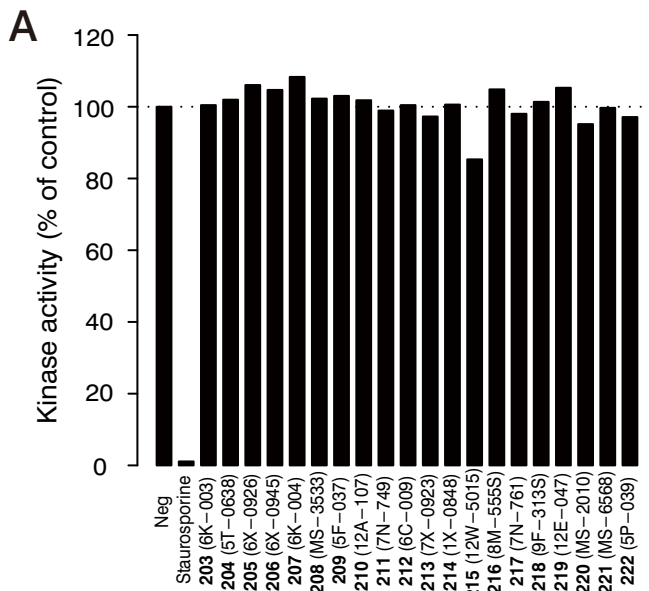
Active compound	Trained compound sharing max-MCS	EGFR known ligand sharing max-MCS	Scaffold-hopped	Previously-reported bioactivity	ligand
 <b>37</b>	 <b>135</b>	 <b>136</b>	No	No	No
 <b>18</b>	 <b>19</b>	 <b>20</b>	No	Yes	Yes
 <b>21</b>	 <b>22</b>	 <b>20</b>	No	No	No
 <b>35</b>	 <b>137</b>	 <b>138</b>	No	Yes	No
 <b>36</b>	 <b>139</b>	 <b>20</b>	No	Yes	No

**D**

Active compound	Trained compound sharing max-MCS	CDK2 known ligand sharing max-MCS	Scaffold-hopped	Previously-reported bioactivity	ligand
 <b>23</b>	 <b>24</b>	 <b>25</b>	Yes	No	No
 <b>26</b>	 <b>27</b>	 <b>25</b>	No	No	No



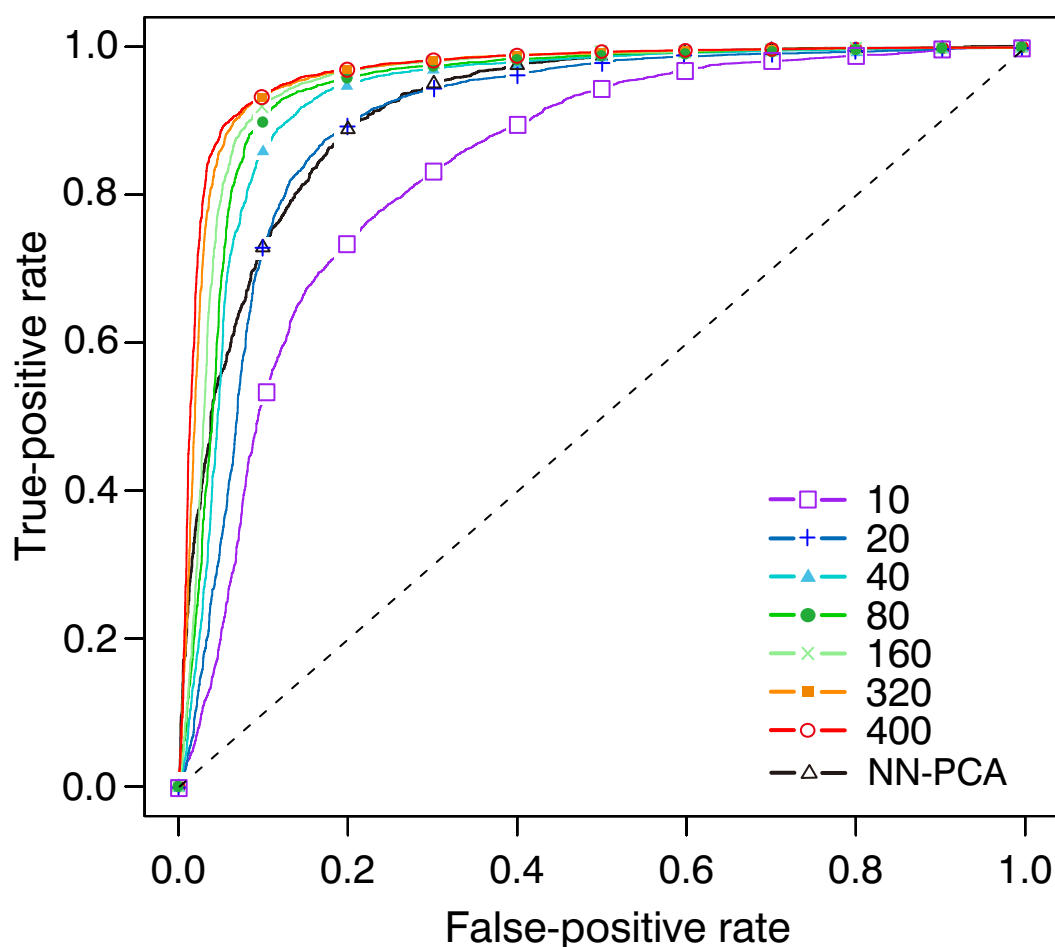
**Figure S10. Experimental confirmation of the *in vitro* kinase activity of Bionet compounds screened by LBVS and SBVS.** Primary inhibitor screening for EGFR (**A**) and CDK2 (**B**) predicted by LBVS (NN-PCA). Primary inhibitor screening for EGFR (**C**) and CDK2 (**D**) scored by SBVS (GOLD). The top 20 ranked Bionet compounds were assayed, and potent ones showing more than 20% inhibition were selected to determine IC<sub>50</sub>s. (**E**) Dose–response curves of compound **253** (3P-066) screened through SBVS for EGFR. (**F**) Dose–response curves of compound **273** (1Y-0244) screened through SBVS for CDK2. (**G**) Max-MCSs and publication status are similar to Figures 4 and 5 in the main text. See Supplementary Table S4 for the other chemical information of the numbered compounds.



**G**

	Active compound	Trained compound sharing max-MCS	EGFR/CDK2 known ligand sharing max-MCS	Scaffold-hopped	Previously-reported	
					bioactivity	ligand
EGFR	 <b>253</b>	 <b>282</b>	 <b>283</b>	No	No	No
CDK2	 <b>273</b>	 <b>284</b>	 <b>285</b>	No	No	No

**Figure S11. ROC curves obtained by CGBVS models with reduced protein descriptors.** Sets of 10 (purple colored open square), 20 (blue colored plus), 40 (light blue colored closed triangle), 80 (green colored closed circle), 160 (light green colored cross), and 320 (orange colored closed square) descriptors were sampled randomly from a total of 400 protein descriptors (red colored open circle) and used to construct each CGBVS models with 5,207 CPIs from GLIDA (Okuno *et al.*, 2006). Each ROC curve is the result of five-fold CV for each model. For comparison, the ROC curve obtained by using NN-PCA (black colored open triangle; same as in Figure 1B) was plotted in black. The NN-PCA is a ligand-based virtual screening method (LBVS) using only chemical descriptors without protein descriptors. The prediction performance by NN-PCA is superior to CGBVS when using descriptors based on only 10 or 20 amino acids. Therefore, descriptors based on 10 amino acids do not work better and might add noise to the prediction model (NN-PCA) that uses only chemical descriptors.



**Figure S12. Ligand screening for the Bionet compound set using LBVS and SBVS.** LBVS (NN-PCA) and SBVS (GOLD) scores for **(A)** ADRB2, **(B)** CDK2, and **(C)** EGFR are shown. The colored dots indicate the assayed compounds with the top 30 (A) or top 20 (B, C) scores by CGBVS (shown in Supplementary Tables S8A, S13A and S13B). Green: active compounds, blue: inactive compounds. Dotted lines show the score of the top30<sup>th</sup> / 20<sup>th</sup>-ranked compound using each method.

