Supplements

Quality assessment of pharmacology data

The Quality Assessment (QA) flag is an output from Combenefit, which was used for the synergy score and monotherapy curve fitting (see Online Methods). Scores are defined as the following:

Flag	**Meaning**
0	No data was found in a combination folder or NaN was found in a combination file.
-1	At least one of the measured drug effects was above 125% of starting cell count. This is unlikely to be genuine and a major experimental issue is suspected.
-2	No flag '-1' but measured effects below -10% were found. By definition, effects should always be positive because cell viability is being measured. Very small negative values are sometime encountered due to quantification problems at high concentrations. These were tolerated up to -10% below which major issues were suspected.
-3	No flags '-1' or '-2' but combination dose-response showed very strong fluctuations. The combination dose-response was smoothed and compared to the original non-smoothed version. If differences above 25% were found the experiment was flagged as measurements likely to be unreliable.
1	None of the previous problems were encountered. Data is supposed to be ok.

All experiments were observed to have some level of synergy or antagonism and had non-zero synergy scores. Most of these were due to random variation in the experiments and had synergy within +/-1, and only 404 experiments with low variability non-zero experiments due to quality issues in the assay and were flagged accordingly (Fig. S1B). Only high quality data (QA=1) were included in the testset, while experiments also with low quality were made available for training.

Drug combination synergy is variable across cells, but reproducible across replicates

367 had a replicate experiment where the same drug combination, cell line and concentration ranges tested (Fig. S2A). The Spearman correlation between synergy scores across the replicates was 0.56, which is comparable to the correlation of 0.63 from the 315 replicate combinations screening experiments (Fig. S2B) completed by O'Neil *et al*[1]. We observed that in all instances where the synergy and antagonism were not measured by both replicates, the quality of one or both replicates had been flagged as low (see above QA flags). Notably, the variance in synergy scores in DREAM (Fig. S2C) was larger than the variance the dataset from O'Neil et al (Fig. S2D).

Best ranked teams' Methods:

Best ranked teams' methods are below, who performed consistently well across all 3 sub-challenges, or outstanding well for at least one sub-challenge.

Yuanfang Guan

This method begins by limiting the feature space to those features (expression, CNV, methylation, mutations) mapping to genes that are putative targets of any drug in a given sub-challenge. If a drug is less specific and has multiple targets, all of them are included.

For each drug combination a separate classifier is built for predicting synergy. Each drug combination classifier is made of several random forests, and each using as distinct data type. For example, in sc1A three classifiers are created per drug combination, and with their predictions are averaged (table 1). This use of one classifier for each data type was motivated to improve the stability of the predictions, in the case that one of the feature sets contains outliers.

Sc1A Classifier	Data type
1	Mono-therapy data of Drug A and B
2	Drug A, B count data
3	All DC's data with either Drug A or B
Sc1B Classifier	Data type
1	CNV, mutation (and mean values for both), count on drug A,
2	CNV, mutation, count (and mean) on drug B
3.	Normalized CNV, mutation (and mean), count on drug A and B
Sc2 Classifier	Data type
1	Drug A, B, count,
2	Normalized All DC's data with either Drug A and B
3	Drug space as a simple parsing of original table provided in the Challenge
4	Mono-therapy data of Drug A and B

Table 1

The choice of classifier construct, randomForest, compared to using an SVM, regression, or boosting, was chosen for expedience, not any perceived advantage in accuracy, though the randomForest does facilitate prediction in a nonlinear context [2]. Instead, greater prediction accuracy is achieved by creating new features that consist of scaling the features provided by AZ-Dream with a posterior probability from a predefined functional network[1].

Network Based Feature Scaling:

Feature scaling is motivated by the observation that when we use the original features as input, we found the prediction values are similar for the same cell line, regardless of drug perturbations. This is due to the fact the genomic and expression data are static for a given cell line across all drug combinations effectively leaving just three parameters for modeling drug synergy (drug A, drug B and cell line). Scaling cell line features by the functional network of a drug target creates a dynamic parameter space for each cell line that allows for better modeling synergy.

For gene expression, methylation, and copy number variation data, features are adjusted based on their probability of a functional relationship with drug target genes. Feature x_i, associated with gene *i*, is scaled to x'_i by:

$$\{if\ g_i \in (DT_A, DT_B), x'_i = 0\ else, x'_i = x_i \times (1 - max\ (e_{ij})\ , \forall\ g_j \in (DT_A, DT_B)\ \}$$

where DT_A and DT_B are the genes targeted by drug *A* and *B* respectively, and e_{ij} is edge between genes *i* and *j* in the predefined functional relationship network[1].

For gene mutations, features are modified using the edge directly. Mutation feature x_i, associated with gene *i*, is changed to x'_i:

$$\{if\ g_i \in (DT_A, DT_B), x'_i = x_i\ else, x'_i = x_i \times max\ (e_{ij})\ , \forall\ g_j \in (DT_A, DT_B)\ \}$$

The purpose of generating this new series of features is to simulate the effective values of expression, methylation, CNV and mutations post treatment. After scaling, predictive features are different for each cell line across different drug-combinations. We reduce the effective values of drug target in expression, methylation and CNV to zero, and reduce the values of other genes according to their connections to the drug target. For mutations, we assumed that the effects of drugs are equivalent to adding in new mutations to the system, with the effect values of drug targets being 1 (similar to mutated genes), while the other genes are increased in values according to their connections to the drug target.

A biological interpretation of this approach is that scaling cell line features with the functional network's edges, allows the RandomForest to model a drug's propagation through a targeted pathway cascade.

The weighting of different set of features was primarily done by cross-validation. However, we found that a single set of genomic features is often sufficient to achieve a similar performance as the entire set.

Mikhail Zaslavskiy

The model is an ensemble of three individual models trained exclusively on categorical features describing drug and cell line identities and one model trained on categorical identity features plus corresponding drug MonoTherapy results. There are two main ideas at the core of the proposed model. First, it is very easy to overfit when dealing with biological data, so the key factor is a proper design of the cross-validation scheme which covers not only meta parameter estimation but also model selection steps. In addition to avoiding overfitting pitfalls, the cross-validation design is important to address precisely the sub-challenge questions defined by corresponding training/test splits. Second, a rich sampling of the experimental space provides an excellent support for a competitive model even without additional features describing biological entities under consideration. When we have enough data on drug/drug and drug/cell line combinations (like in sc1A and sc1B) we can derive the information on drug and cell line similarities without using additional features. Of course, it is impossible to know in advance if the experimental results alone are enough to reach the maximum performance, so the model building process is to start with the set of baseline features (drug and cell line ids) and then add step-by-step more complex features (MonoTherapy results, drug features, cell line mutations, copy number variations et c.) verifying at each step if the addition of new features lead to an improvement of the cross-validation score.

The three individual models used to predict drug synergy from durg/cell line identities are a gradient boosting tree model (xgboost package) and an svm model (e1071 package) trained on original identity features represented by a binary matrix, and an elastic net model (R/glmnet package) model trained on average scores of drug-cell line combinations with the same drug combination and a different cell line, average scores of drug-cell line combinations with the same cell line and one drug in common, average scores of drug-cell line combinations with the same cell line and no common drug. The fourth model is another gradient boosting tree model trained on drug MonoTherapy results and counts of categorical identity features. All four models were trained using 5-fold cross-validation, the final score was computed as a simple average of the individual models.

North Atlantic DREAM (NAD)

North Atlantic Dream team's solutions used different tree based models (Random Forest Regression and Extreme Gradient Boosting Trees, XGBoost [3]) to incorporate the presumed important interactions between cellular (mutations, copy number alterations etc.) and drug specific (drug targets, affected pathways etc.) features. For better representation of the similarities between cell lines and drug combinations, new sets of features were engineered using prior knowledge and also the monotherapy data.

The monotherapy data (IC50, Einf, Hill slope) for a drug combination was dependent on the ordering of the drugs in the combination, which seemed to be hard to represent in the machine learning model. To overcome this problem, North Atlantic Dream's model used monotherapy features that are independent of the ordering (such as min/max/absolute difference etc. type features from the original data). Also the expected volume under the dose-response surface (in case of additivity) was calculated using the original Loewe model [4].

As the number of training examples for a given drug combination was relatively low (about a dozen for most combinations), it was crucial to find similarities between combinations beyond the trivial ones (same drug / same target). North Atlantic Dream created different feature sets based on Gene Ontology [5] / KEGG Pathways [6]and a directed signaling network [7]. For Gene Ontology based features a set of “cancer related” GO terms were selected (based on [8]), and for each drug a GO vector was created based on the association of GO terms with the target of the drug. For KEGG Pathway based features, KEGG Pathways containing the target genes were selected, and for each drug a KEGG vector was created, giving 1 values for pathways containing the target of drug, 0 otherwise. The GO/KEGG features for drug combinations were the sum of the two drug vectors of the combination. Based on the directed signaling network, for each drug combination the “similar” drug combinations were selected. Two drug combinations were defined similar, if the two targets of combination A are direct upward from the two targets of combination B. Based on this rule a similarity vector was created for each combination. To create these GO/KEGG/signaling networks based features, in case of “DNA targeting drugs” (i.e. chemotherapy drugs), the respective DNA damage response molecule was used as indirect target (based on Woods & Turchi, 2013 [9]).

For cellular features mutations, copy number variations and gene expression were used. The main problem with cellular features was their large number. To overcome this problem, North Atlantic Dream team used a pre-assembled gene list (including target genes, known oncogenes and tumor suppressors, genes related to drug monotherapy resistance etc. [10] [11] [12]). Genes with mutations / copy number alterations were selected from this gene list. The low number of training examples for a given drug combination made it hard to use traditional feature selection/reduction methods. However, based on the drug similarities defined above, it was possible to select molecular features associated with the observed synergy scores for a given, similar set of drug combinations. During the original Challenge gene expression features were selected (from the expression of target genes and their direct neighbors in signaling network) using this method by Randomized Lasso Feature selection. In the later, collaborative phase of the Challenge, a similar method was used for mutation and copy number variation features.

For the final prediction different XGBoost models were created using subsets of the above defined features and/or different model parameters. The final submitted predictions were the ensembles of these models, either as simple averages or using hillclimbing [13]on out-of-fold predictions. For the various tasks through the Challenge R, Python, SAS, and JMP were used.

NAD feature layer importance:

NAD's Random Forest Regression model was trained using different pairs of cell line and drug combination specific features. The tested cell line features included cell line label, mutations (pre-filtered for 469 cancer related gene) and CNV (pre-filtered for 292 gene). Combination related features tested here were drug label, drug target, Gene Ontology and KEGG pathway based features (feature size: 407 and 140, respectively) and signalling network based features (601). Baseline model used cell line and drug label as features, while in the other models the respective feature was either swapped with the corresponding baseline feature (e.g.: in *CNV model* CNV and drug label, in *target model* cell line label and drug target was used as features) or added to the the features of baseline model (e,g. in *+target model* cell line label, drug label and drug target was used). Ensemble model in this case is the simple average of the prediction of these models. With all the used models 10 random cross-validation was performed, and the mean weighted Pearson correlation was calculated for each cross-validation run. For the cross-validations all the training and leaderboard data of the Challenge was used, and the size of the cross-validation set for each combination resembled the size of the test set of the Challenge.

NAD biomarker selection:

NAD ranked their biomarkers for each drug combination based on the number of times the feature was used for predicting the given combination in the Random Forest Regressor models. For each combination a separate Random Forest model was built (using mutation, CNV, drug target and KEGG features) where all the Challenge training and leaderboard data was used without the data of the actual combination. With this model the left out combination was predicted for all of the cell lines. The number of times a given cellular feature (mutation or CNV) was used (based on the internal structure of the Random Forest trees) was recorded for each combination - feature pair. For each drug combination - feature pair Mann-Whitney U test was performed to calculate the probability that the given feature is used more often for the given combination than other features, and that the given feature is used more often for the given combination than for other combinations. The final score of the feature for the combination was the product of these two probabilities. For each combination the used cellular features (mutation and CNV) was sorted decreasing order, and the top 5 features was used for further analysis.

DMIS

Support Vector Regression (SVR) were used as prediction model. The main difficulty was the high dimensionality problem of our feature space. To address this, novel literature-based approach was used. 200 genes were identified that most frequently occur in the context of cancer in the literature, and used only the mutations in those 200 genes as the features. To perform this gene selection task, the Biomedical Entity Search Tool (BEST) (http://best.korea.ac.kr)[14] was applied. BEST finds an entity relevant to a query based on the number of co-occurrences between the query terms and the entity in the PubMed corpus, the authority of journals, the recency of articles, and the term frequency inverse document frequency (TF-IDF) weighting. BEST was queried by using the query term “cancer” and the top 200 cancer-related genes were collected. For CNV features, cBioPortal was used to collect 13 gene sets of cancer-related pathways. During the creation of various types of features, it was of essence to identify the best combinations of feature groups was a necessary part. However, testing all possible combinations of feature groups would require a considerable amount of computing resources. To address this problem, a high performance computing pipeline using HTCondor was constructed. HTCondor is an open-source computing framework for coarse-grained distributed parallelization of computationally intensive tasks. We ran our pipeline using 1,764 cores from Amazon Web Service, and selected the best combination of the feature groups. Through this process, the following features for sub-challenge 1B were selected: 118 drug IDs, 99 drug targets, 94 CNVs, 241 mutations, and maximum concentrations of the dosages for each sample.

The SVR model showed a good performance on AZ dataset. However, because the original model represents target and mutation features as sparse binary vectors, it is not appropriate to apply to O'Neil et al dataset with unseen cell lines and drugs. For translatability, a dense vector was created to capture and generalize characteristics of cell lines and drugs, and constructed a deep learning model which could utilize these vectors as input.

For the post-hoc analysis, an additional deep learning model was generated, which was composed of 6 layers including a preprocessing layer as the first layer. The second layer had 4 modules and the first module gets mutation features, and the second module gets target feature. These two modules embed sparse feature vectors as dense vectors, and generate a single vector using a convolutional neural network. Pre-trained mutation vectors were used to leverage mutation information from TCGA and Mikolov’s Word2Vec algorithm[15] for mutation embedding. In addition, Asgari’s public protein dense vector[16] for target embedding was used. The third module gets drug or monotherapy-related features, and the last module gets cell line-related features. Each module generates a single vector and the vectors are concatenated and are inputted into the next layer. The rest of layers are fully-connected layer. The output layer generates a single value, which is the predicted synergy score.

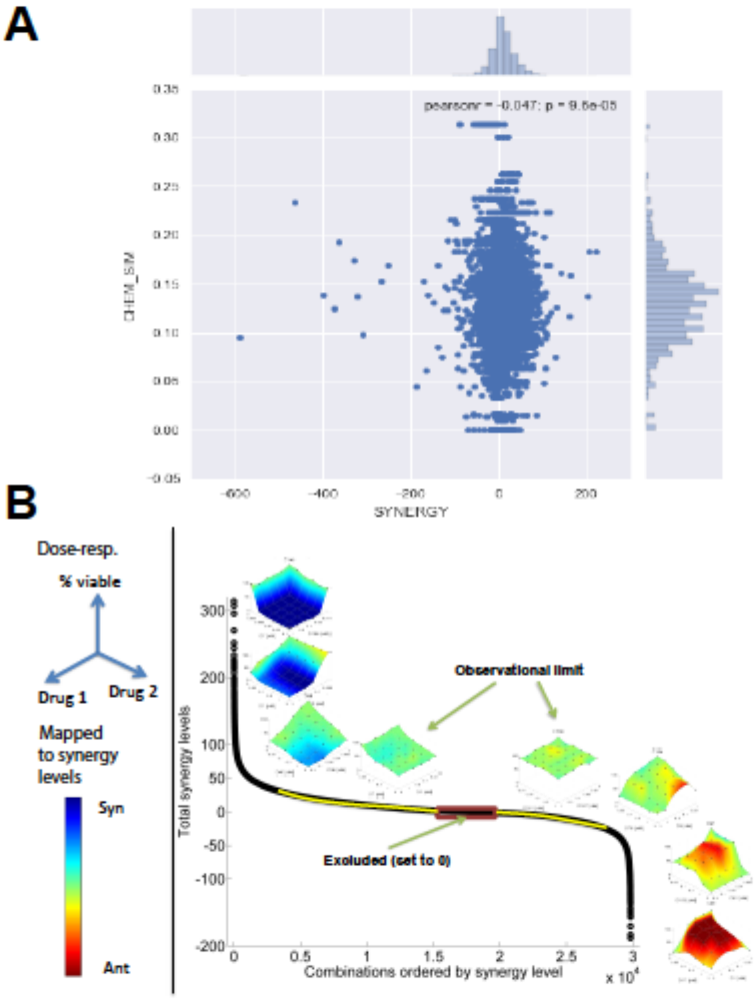
DMIS feature layer importance:

In the main Challenge, the Support Vector Regression (SVR) was used as machine learning model. For extracting feature layer importance, the accuracy of the model was estimated after randomly permuting the values of the feature. How much the permutation decreases the primary score of the SVR model is an estimate of feature layer importance.

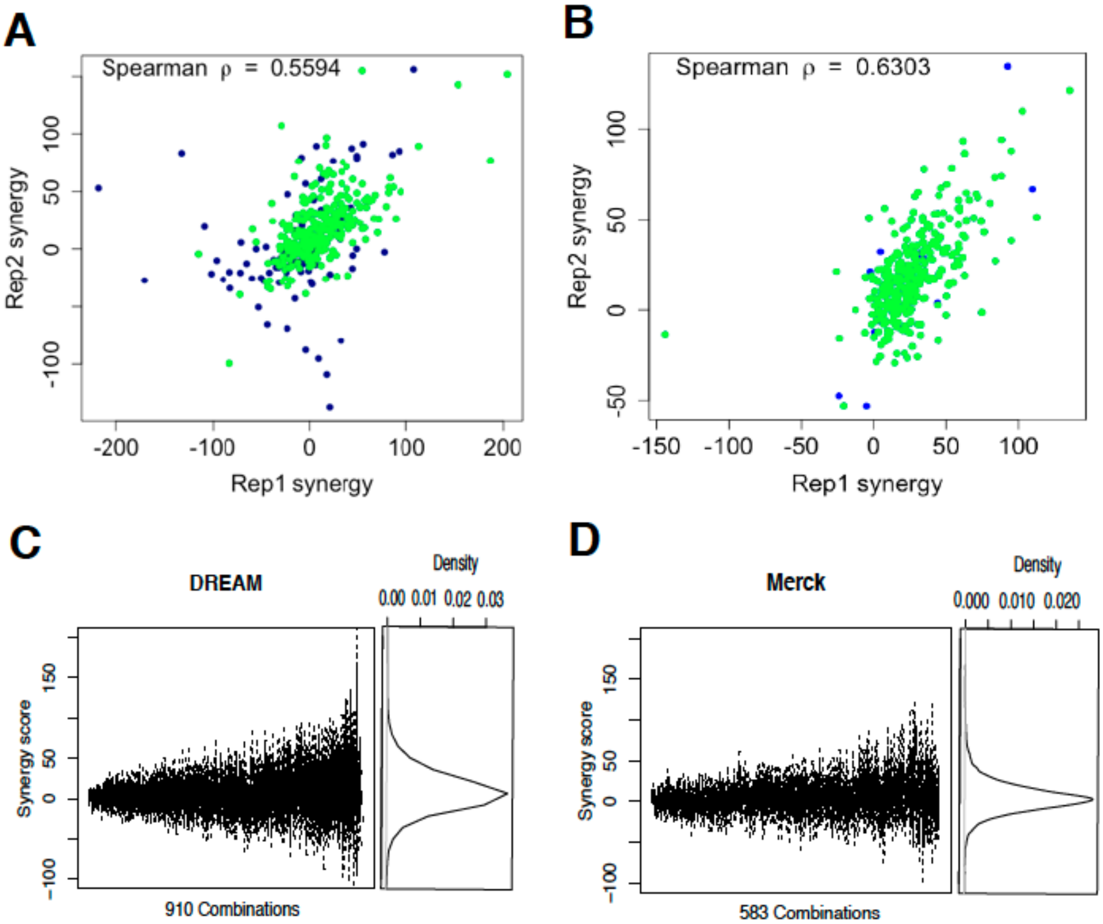
DMIS biomarker selection:

In order to get biomarker indications, important features were extracted for each drug combination individually. Therefore, samples were grouped by drug combination and for each combination and a random forest model was build. The method called “mean decrease impurity” was applied to obtain feature rankings. Random forest consists of multiple decision trees. Each node in a tree correspond to a feature and the same feature can appear in multiple trees. Each node in a tree splits the samples so that similar synergy scores group together. At each node, we can compute how much the split reduces the variance in the sample by taking the difference between the variance-before-split and the variance-after-split (this can be measured by weighted average of the two split groups). Finally, features were ranked based on their mean variance reduction, i.e., mean decrease impurity. The number of trees was limited to 200 and each tree randomly selected up to 204 features (a third of total 612 features available).

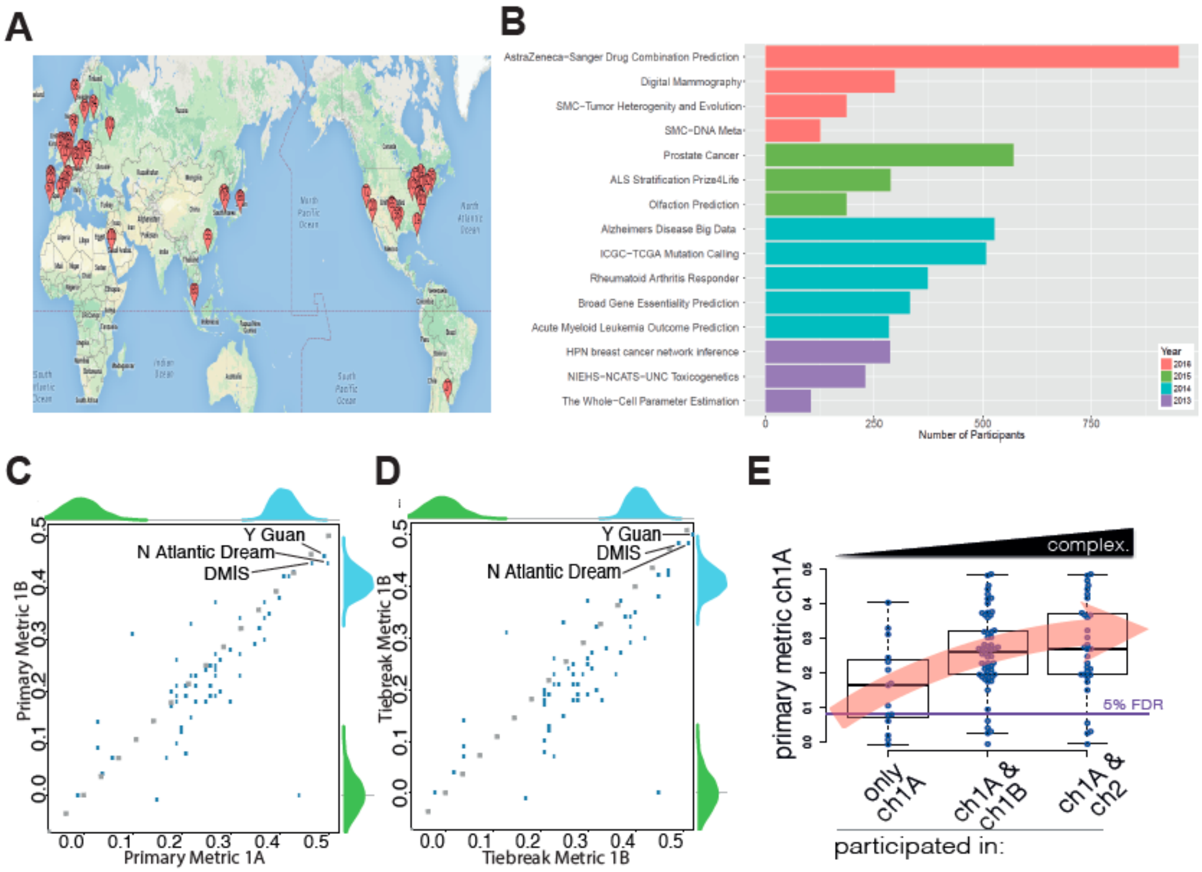
Supplemental figures



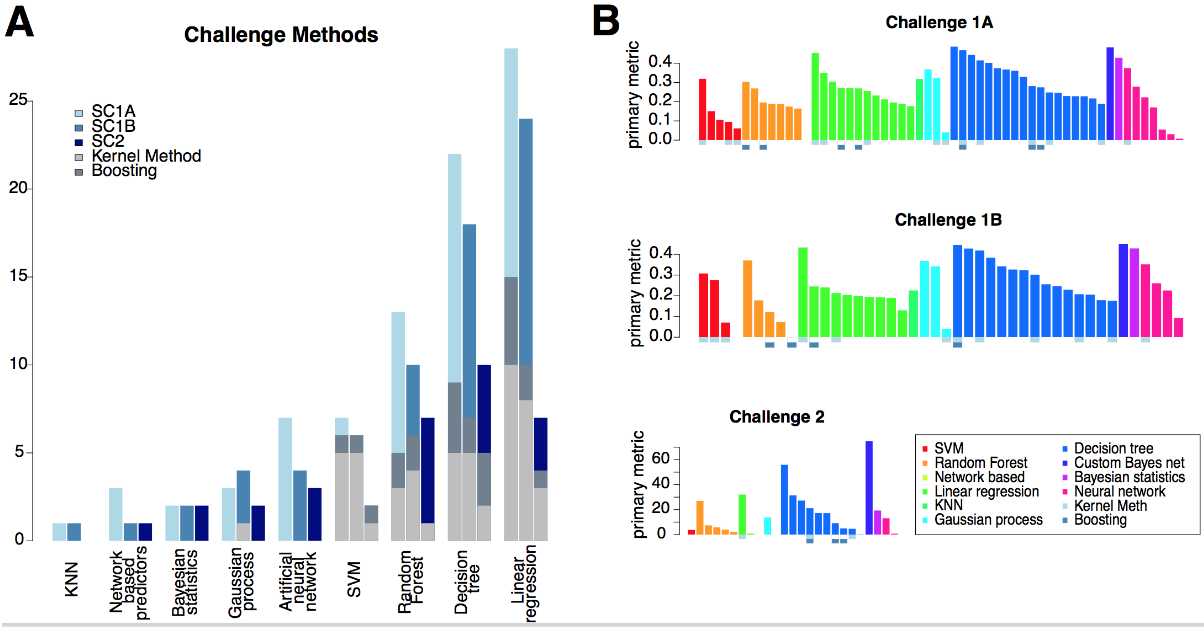
Supp. Figure 1: Distribution of synergy scores across all combinations. (A) ty of compounds in each combination is plotted against their synergy scores. (B) Chemical similar synergy scores of combinations are ordered from lowest to highest. 3D synergy heatmaps show additional cells killed (Syn) or not killed (Ant) beyond the additive effect of the two drugs at each dose. Two examples of combinations with total synergy scores of +/-20 show the limit at which synergy and antagonism can be visually confirmed. Experiments where non-zero total synergy was due to random variation were set to zero.



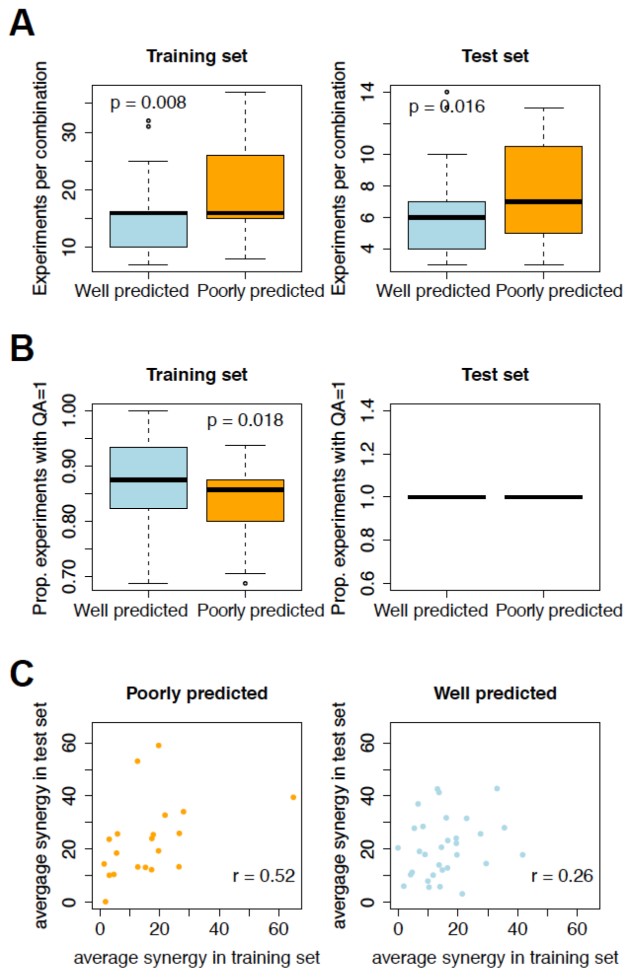
Supp. Figure 2: Reproducible of synergy scores. Shows correlation of drug combinations replicates from (A) DREAM Challenge and (B) external combination screen (O'Neil et al. 2016). In green are high quality data points, while in blue are points not passing the QC from CombeneFit (QC score=1). (C, D) Distribution of synergy scores across combinations screened for the DREAM Challenge and independent set of combinations screened by Merck.



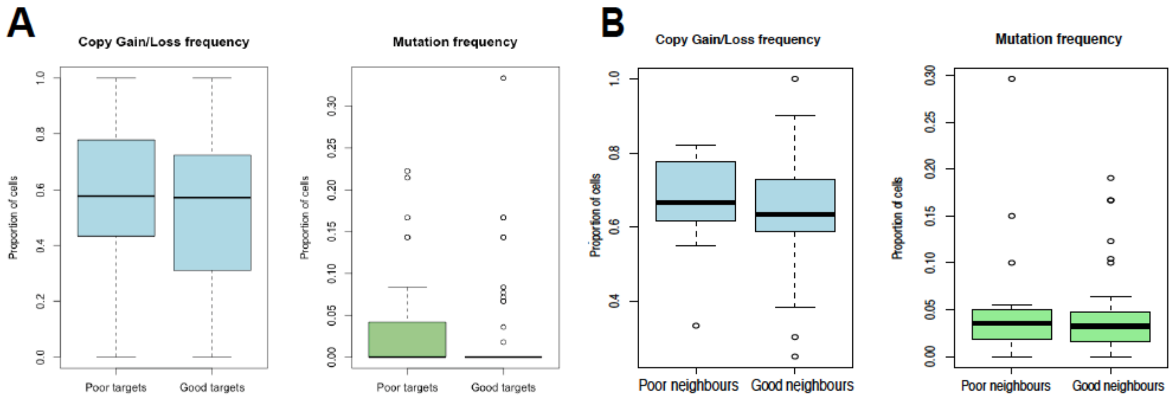
Supp. Figure 3: Participation in the Combinations Prediction DREAM Challenge. (A) Nearly 800 participants were located across five continents. (B) Comparison of participants across different DREAM Challenges. (C) Performances of sub-challenge 1A plotted against 1B based on the primary metric, average weighted Pearson correlation. (D) Performances of sub-challenge 1A plotted against 1B based on the tie-break metric, average weighted Pearson correlation of combinations with cases of synergy > 20. Dotted grey line shows 1:1 relationship between 1A and 1B performance. (E) Performance of participants in sub-challenge 1A grouped by whether they participated in one or more sub-challenges.



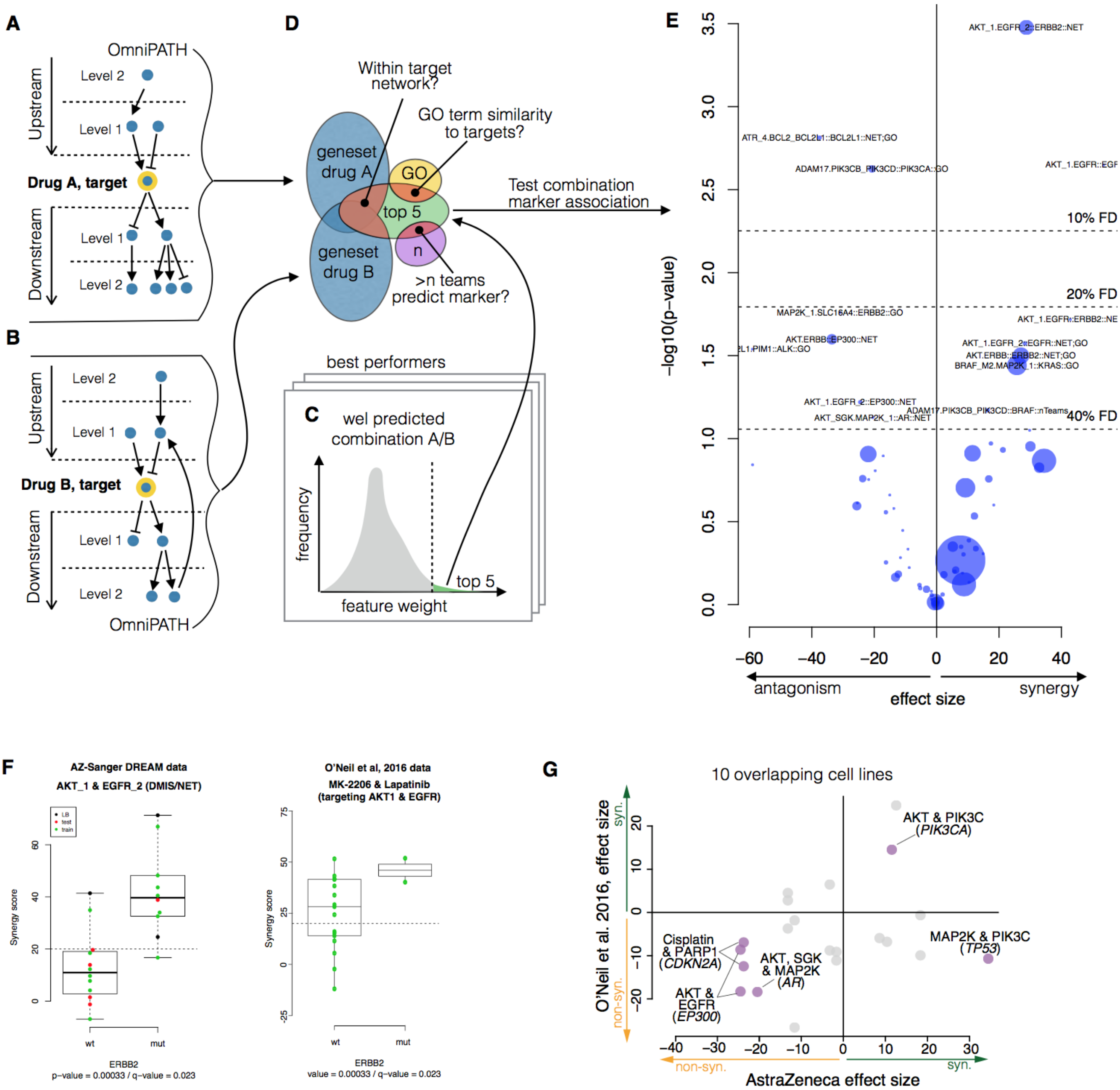
Supp. Figure 4: (A) Performance across different types of methods used by participants in each sub-challenge. Each bar represents occurrence of the method in subch 1A, 1B, and 2. (B) Performance of individual teams coloured by the primary type of machine learning method used in each sub-challenge. Indicators below each bar show cases where kernel and boosting techniques were used with the primary method.



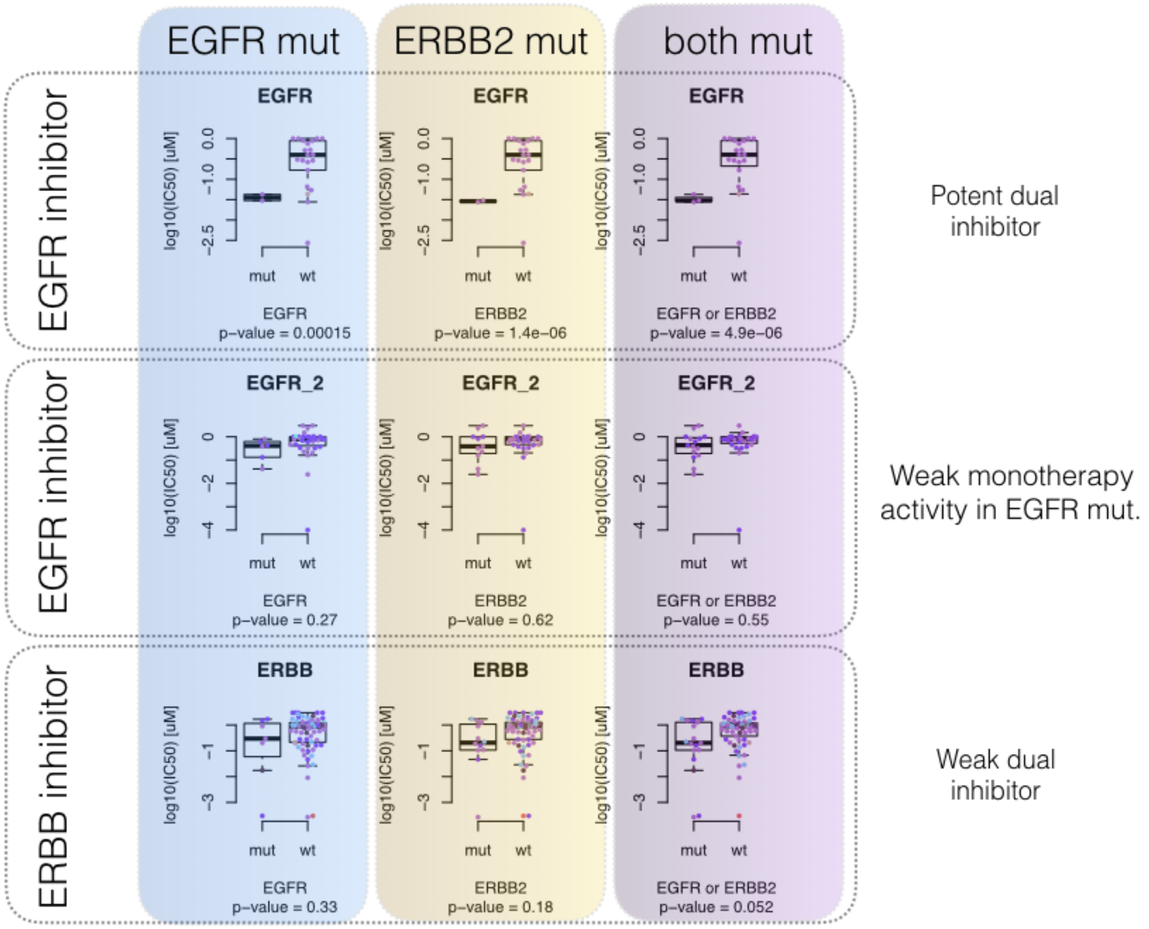
Supp. Figure 5: Training and test set differences between well and poorly predicted combinations based on (A) number of experiments for each combination, (B) proportion of experiments with high quality (QA=1), and (C) average synergy score for each combination.



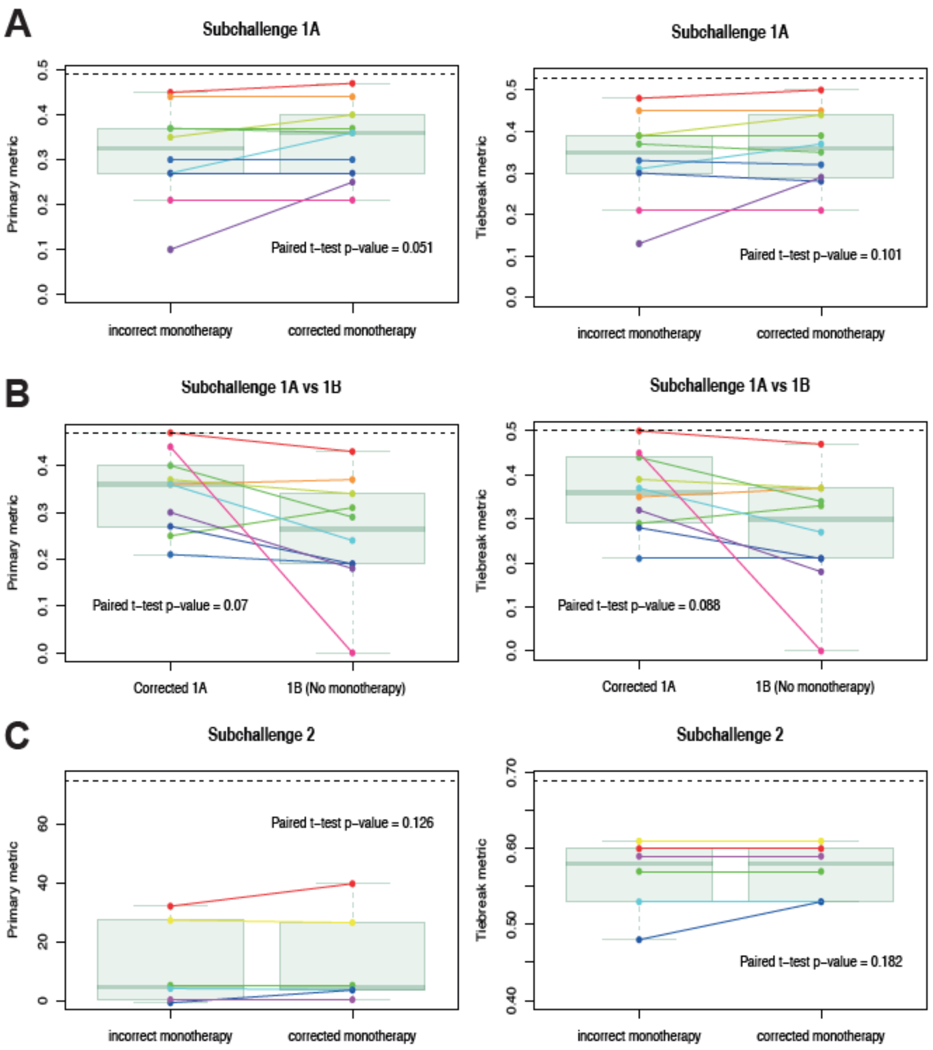
Supp. Figure 6: Alteration frequency in drug targets and nearest neighbours. (A) Copy number and somatic mutation frequency in the gene targets of drug combinations. (B) Copy number and somatic mutation frequency in the genes that are nearest interacting neighbors of the targets, as determined by *OmniPath*.



Supp. Figure 7: Post-hoc analysis of synergy biomarkers. (A) and (B) shows the target centric exploration of 2 levels up- and downstream of the putative drug targets from each combination. (C) The the top 5 ranked features from well predicted models were chosen for exploring their target enrichment. (D) Additionally, the top 5 features were further investigated if they had GO term similarity to the putative target larger than 0.5, or two independent teams relied on the same features. (E) This putative gene-to-combination association set was tested with an ANOVA model. Her exemplified with (F) AKT combined with EGFR in ERBB2 mutants showed synergy and independently validated with O'Neil et al. 2016. (G) General validation of synergy biomarkers in only overlapping cell lines between AZ-DREAM and O'Neil et al. 2016.



Supp. Figure 8: Monotherapy markers of EGFR and EBB2 inhibitors. Exploration of EGFR, ERBB mutations and copy number changes alone and in combinations.



Supp. Figure 9: Value of monotherapy on prediction performance. (A) Prediction performance of teams (dots) in sub-challenge 1A when given correct and incorrect monotherapy data. (B) Comparison of performance between sub-challenge 1A and 1B for teams that had correct monotherapy data but could use it in 1A and could not in 1B. (C) Prediction performance of teams in sub-challenge 2 when given correct and incorrect monotherapy data. Horizontal dashed line indicates the level of the top performing team.

Supplemental tables

Supplementary Table 4: Prediction performance on independent combinations screen

		Performance: Average weighted Pearson Correlation			
Team	Method	All experiments	Same Cells	Similar Drug	Similar Combination
Mikhail	1A model	0.04	0.06	0.05	0.1
Mikhail	1B model	-0.05	-0.07	-0.02	0
NorthAtlanticDream	1A model	0.05	0.05	0.05	0.03
NorthAtlanticDream	1B model	0.03	0.07	0.05	0.07
DMIS	new 1A model (Deep Learning)	0.11	0.13	0.11	0.1
DMIS	1A model	0.08	0.12	0.08	0.03
DMIS	1B model	-0.03	0.01	0	-0.05
Ensemble	Average 1A model	0.13	0.17	0.13	0.11

Supplemental source code

1. Winning method (code freeze)
2. Scoring code is available online at https://www.synapse.org/#!Synapse:syn4991619

Supplementary references

1. O'Neil, J. *et al.* An Unbiased Oncology Compound Screen to Identify Novel Combination Strategies. *Mol. Cancer Ther.* **15,** 1155–1162 (2016).
2. Breiman, L. Random Forests. *Mach. Learn.* **45,** 5–32 (2001).
3. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016).
4. Yadav, B., Wennerberg, K., Aittokallio, T. & Tang, J. Searching for Drug Synergy in Complex Dose–Response Landscapes Using an Interaction Potency Model. *Comput. Struct. Biotechnol. J.* **13,** 504–513 (2015).
5. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43,** D1049–56 (2015).

6. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44,** D457–62 (2016).

7. Babur, Ö. *et al.* Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. *Genome Biol.* **16,** 45 (2015).

8. Sun, Y. *et al.* Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. *Nat. Commun.* **6,** 8481 (2015).

9. Woods, D. & Turchi, J. J. Chemotherapy induced DNA damage response: convergence of drugs and pathways. *Cancer Biol. Ther.* **14,** 379–389 (2013).

10. An, O., Dall'Olio, G. M., Mourikis, T. P. & Ciccarelli, F. D. NCG 5.0: updates of a manually curated repository of cancer genes and associated properties from cancer mutational screenings. *Nucleic Acids Res.* **44,** D992–9 (2016).

11. Wagner, A. H. *et al.* DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Res.* **44,** D1036–D1044 (2016).

12. Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41,** D955–61 (2013).

13. Caruana, R., Niculescu-Mizil, A., Crew, G. & Ksikes, A. Ensemble Selection from Libraries of Models. in *Proceedings of the Twenty-first International Conference on Machine Learning* 18– (ACM, 2004).

14. Lee, S. *et al.* BEST: Next-Generation Biomedical Entity Search Tool for Knowledge Discovery from Biomedical Literature. *PLoS One* **11,** e0164680 (2016).

15. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. Distributed Representations of Words and Phrases and their Compositionality. in *Advances in Neural Information Processing Systems 26* (eds. Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z. & Weinberger, K. Q.) 3111–3119 (Curran Associates, Inc., 2013).

16. Asgari, E. & Mofrad, M. R. K. Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. *PLoS One* **10,** e0141287 (2015).

Consortium Authors

Bàrbara Schmitz Abecassis[23], Nanne Aben[51,56], Tero Aittokallio[68], Bissan Al-lazikani[94], Tanvir Alam[18], Amin Allam[73], Mariana Pelicano de Almeida[23], Vinicius Alves[55], Benedict Anchang[48], Albert A. Antolin[94], Wail Ba-alawi[18], Moeen Bagheri[102], Vladimir Bajic[18], Gordon Ball[97], Pedro J. Ballester[2,15,65,66], Delora Baptista[14], Christopher Bare[86], Denis Bertrand[17], Bhagya[64], Gianluca Bontempi[71,75], Keith A. Boroevich[13], Evert Bosdriesz[10,51], Salim Bougouffa[18], Gergana Bounova[51], Krishna Bulusu[4], Alberto Calderone[6], Stefano Calza[40,43], Stephen Capuzzi[55], Daniel Carlin[42], Hannah Carter[42], Luisa Castagnoli[24], Gianni Cesareni[24], Hyeokyoon Chang[30], Guocai Chen[100], Lijun Cheng[64], Jaejoon Choi[5], Kwanghun Choi[30], Elizabeth Coker[94], Miklos Cserzo[47,74], Cuong C. Dang[2,15,65,66], Tjeerd Dijkstra[76], Sorin Draghici[37,46], Jonathan Dry[4], Michel Dumontier[69], Friederike Ehrhart[23], Bo van Engelen[22], Hatice Billur Engin[42], Iwan de Esch[50], Chris Evelo[23], Sherif Farag[55], Kathleen Fisch[12], Åsmund Flobak[80,91], Stephen Friend[86], Mehmet Gonen[1516,88], Jussi Gillberg[29], Lizzy Godynyuk[23], Anna Goldenberg[60], David Gomez-Cabrero[78,97], Chris de Graaf[50], Maxim Grechkin[103], Yuanfang Guan[99], Justin Guinney[86], Emre Guney[72], Benjamin Haibe-Kains[82], Di He[81], Liye He[68], Manuela Helmer-Citterich[24], Daniel Hidru[102], Ya-Chih Hsueh[102], Justin K Huang[7], Laszlo Hunyady[47,74], Jinseub Hwang[32], Woochang Hwang[5], Yongdeuk Hwang[36], Olexandr Isayev[55], Oliver Bear Don't Walk IV[25], Minji Jeon[30], Jiadong Ji[26,34,89], Yousang Jo[21], Piotr J. Kamola[13], Georgi K. Kanev[50], Jaewoo Kang[30,70], Loukia Karacosta[48], Samuel Kaski[29], Marat Kazanov[1,90], Abdullah M Khamis[18], Suleiman Ali Khan[68], Narsis A. Kiani[3], Jinhan Kim[92], Kiseong Kim[21], Sunkyu Kim[30], Yongsoo Kim[51,53], David Knowles[39,48], Melissa Ko[27], Albert J. Kooistra[50], Martin Kuiper[80], Mijin Kwon[21], Astrid Lægreid[80], Twan van Laarhoven[20], Simone Lederer[20], Heewon Lee[70], Eemeli Lepp_aho[29], Lang Li[95], Joshua Low[23], Artem Lysenko[13], Yosvany LÑpez[13,19,41], Daniel Machado[14], Ana Belen Malpartida[23], Hiroshi Mamitsuka[8], Francesco Marabita[97], Pekka Marttinen[58], Mike J Mason[86], Alireza Mazaheri[3], Arfa Mehmood, Ali Mehreen, Michael P. Menden[4], Magali Michaut[51], Ryan A. Miller[23], Costas Mitsopoulos[94], Rajiv Movva[39], Sebastian Muraru[23], Eugene Muratov[55], Niranjan Nagarajan[17], Elias Neto[86], Tin Nguyen[31], Zheng Ning[40], Thea Norman[86], Baldo Oliva[93], Catharina Olsen[71,75], Antonio Palmeri[24], Bhawan Panesar[102], Jaesub Park[21], Sungjoon Park30, Yudi Pawitan[40], Daniele Peluso[24], Jian Peng[98], Livia Perfetto[24], Stefano Pirrò[24], Sylvia Plevritis[48], Regina Politi[55], Hoifung Poon[77], Ricardo Ramnarine[102], Linda Rieswijk[23,49], Miguel Rocha[14], Carmen Rodriguez-Gonzalvez[94], Julian R. de Ruiter[51,52], Julio Saez-Rodriguez[54,85], Zhaleh Safikhani[82], Moritz Schlichting[22], Ming-Mei Shang[97], Alok Sharma[13,67,101], Hari Sharma[102], Motoki Shiga[38], Moonshik Shin[21], Kevin Shopsowitz[57], Dylan Skola[7], Petr Smirnov[82], Sang Ok Song[92], Othman Soufan[18], Boris Steipe[102], Gustavo Stolovitzky[61,62], Chayaporn Suphavilai[17,35], Bence Szalai[47,74,84], David Tamborero[83], Jing Tang[68], Zia-ur-Rehman Tanoli[68], Jesper Tegner[9,97], Liv Thommesen[80], Seyed Ali Madani Tonekaboni[82], Amy Truong[86], Tatsuhiko Tsunoda[13,19,41], Gabor Turu[47,74], Guang-Yo Tzeng[102], Giovanni Di Veroli[4], Daniel Vis[51], Ashley Wang[102], Hong-Qiang Horace Wang[11], Sheng Wang[98], Dennis Wang[96], Krister Wennerberg[68], Lodewyk Wessels[10,51,56], Bart A. Westerman[45], Egon Willighagen[23], Russ Wolfinger[87], Tom Wurdinger[44], Lei Xie[33], Hua Xu[100], Bhagwan Yadav[59], Huwate Yeerna, Jia Wei Yin[102], Michael Yu[7], MinHwan Yu[30], Thomas Yu[86], So Jeong Yun[92], Alexey Zakharov[79], Mikhail Zaslavskiy, Hector Zenil[3], Frederick Zhang[102], Pengyue Zhang[63], Wei Zhang[42], Wenjin (Jim) Zheng[100], Remzi Çelebi[28]

1. A.A.Kharkevich Institute for Information Transmission Problems Moscow Russia

2. Aix-Marseille University UM105 France
3. Algorithmic Dynamics Lab Unit of Computational Medicine Department of Medicine Solna SciLifeLab Center for Molecular Medicine Karolinska Institute
4. AstraZeneca
5. Bio-Synergy Research Center Daejeon Republic of Korea.
6. Bioinformatics and Computational Biology Unit Department of Biology University of Rome Tor Vergata Rome 00133 Italy
7. Bioinformatics and Systems Biology Program University of California San Diego La Jolla CA
8. Bioinformatics Center Institute for Chemical Research Kyoto University Japan
9. Biological and Environmental Science and Engineering Division KAUST
10. Cancer Genomics Netherlands
11. Cancer Hospital of Chinese Academy of Sciences Hefei China
12. Center for Computational Biology and Bioinformatics University of California San Diego La Jolla CA
13. Center for Integrative Medical Sciences RIKEN Japan
14. Centre Biological Engineering (CEB) University of Minho
15. CNRS UMR7258 Marseille France.
16. College of Engineering Koç University Istanbul Turkey
17. Computational and Systems Biology Genome Institute of Singapore
18. Computational Bioscience Research Center (CBRC) KAUST
19. CREST JST Japan
20. Data Science Radboud University Netherlands
21. Department of Bio and Brain Engineering Korea Advanced Institute of Science and Technology Daejeon Republic of Korea.
22. Department of Bioinformatics - BiGCaT NUTRIM Maastricht University Maastricht 6229 ER Maastricht The Netherlands
23. Department of Bioinformatics - BiGCaT NUTRIM Maastricht University The Netherlands
24. Department of Biology University of Rome Tor Vergata Rome Italy
25. Department of Biomedical Informatics Columbia University in the City of New York
26. Department of Biostatistics School of Public Health Shandong University China.
27. Department of Cancer Biology Stanford University Stanford
28. Department of Computer Engineering Ege University Turkey
29. Department of Computer Science Aalto University
30. Department of Computer Science and Engineering Korea University Seoul Korea
31. Department of Computer Science and Engineering University of Nevada Reno NV
32. Department of Computer Science and Statistics Daegu University South Korea
33. Department of Computer Science Hunter College and The Graduate Center The City University of New York
34. Department of Computer Science Hunter College The City University of New York NY
35. Department of Computer Science National University of Singapore
36. Department of Computer Science The University of Suwon Suwon Republic of Korea.
37. Department of Computer Science Wayne State University Michigan
38. Department of Electrical Electronic and Computer Engineering Gifu University
39. Department of Genetics Stanford University

40. Department of Medical Epidemiology and Biostatistics Karolinska Institute Stockholm Sweden
41. Department of Medical Science Mathematics Medical Research Institute Tokyo Medical and Dental University Japan
42. Department of Medicine University of California San Diego La Jolla CA
43. Department of Molecular and Translational Medicine University of Brescia Brescia Italy
44. Department of Neurosurgery Cancer Center Amsterdam HZ Amsterdam The Netherlands
45. Department of Neurosurgery Cancer Center Amsterdam The Netherlands
46. Department of Obstetrics and Gynecology Wayne State University Michigan
47. Department of Physiology Faculty of Medicine Semmelweis University Budapest Hungary
48. Department of Radiology Stanford University
49. Division of Environmental Health Sciences School of Public Health University of California Berkeley CA United States
50. Division of Medicinal Chemistry AIMMS Vrije Universiteit Amsterdam The Netherlands
51. Division of Molecular Carcinogenesis NKI Amsterdam The Netherlands
52. Division of Molecular Pathology NKI Amsterdam The Netherlands
53. Division of Oncogenomics NKI Amsterdam The Netherlands
54. EMBL-EBI
55. Eshelman School of Pharmacy University of North Carolina at Chapel Hill
56. Faculty of EEMCS Delft University of Technology The Netherlands
57. Faculty of Medicine University of British Columbia
58. Helsinki Institute for Information Technology HIIT Department of Computer Science Aalto University Finland
59. Hematology Research Unit Helsinki Department of Clinical Chemistry and Hematology University of Helsinki and Helsinki University Hospital Comprehensive Cancer Center Helsinki Finland
60. Hospital for Sick Children Toronto Ontario Canada
61. IBM Research
62. Icahn School of Medicine at Mt. Sinai
63. Indiana University - Purdue University Indianapolis
64. Indiana University School of Medicine
65. INSERM U1068 Marseille France
66. Institut Paoli-Calmettes Marseille France
67. Institute for Integrated and Intelligent Systems Griffith University Australia
68. Institute for Molecular Medicine Finland FIMM University of Helsinki Finland
69. Institute of Data Science Maastricht University Maastricht Netherlands
70. Interdisciplinary Graduate Program in Bioinformatics Korea University Seoul Korea
71. Interuniversity Institute of Bioinformatics in Brussels (IB) Belgium
72. Joint IRB-BSC-CRG Program in Computational Biology Institute for Research in Biomedicine Barcelona Spain
73. King Abdullah University of Science and Technology (KAUST)
74. Laboratory of Molecular Physiology Hungarian Academy of Sciences and Semmelweis University (MTA-SE) Budapest
75. Machine Learning Group (MLG) Department d' Informatique Universite libre de

Bruxelles (ULB) Brussels 1050 Belgium

76. Max Planck Institute for Developmental Biology Tuebingen Germany
77. Microsoft Research Redmond
78. Mucosal and Salivary Biology Division King’s College London Dental Institute London SE1 9RT UK
79. National Center for Advancing Translational Sciences National Institutes of Health Rockville MD
80. Norwegian University of Science and Technology
81. Ph.D. Program in Computer Science The Graduate Center The City University of New York
82. Princess Margaret Cancer Centre University Health Network Toronto Ontario Canada
83. Research Unit on Biomedical Informatics University Pompeu Fabra UPF Barcelona Spain
84. RWTH Aachen University Faculty of Medicine Joint Research Center for Computational Biomedicine Aachen Germany
85. RWTH-Aachen
86. Sage Bionetworks
87. SAS Institute Inc. Cary NC
88. School of Medicine Koç University Istanbul Turkey
89. School of Statistics Shandong University of Finance and Economics Jinan China
90. Skolkovo Institute of Science and Technology Moscow Russia
91. St. Olav's University Hospital
92. Standigm Inc. Seoul Korea
93. Structural Bioinformatics Group GRIB IMIM Department of Experimental and Life Sciences Universitat Pompeu Fabra Barcelona Catalonia Spain
94. The Institute of Cancer Research 15 Cotwold Road London SM2 5NG UK
95. The Ohio State University College of Medicine Department of Biomedical Informatics
96. The University of Sheffield
97. Unit of Computational Medicine Department of Medicine Solna SciLifeLab Center for Molecular Medicine Karolinska Institute
98. University of Illinois at Urbana-Champaign Urbana
99. University of Michigan Department of Computational Medicine and Bioinformatics
100. University of Texas Health Science Center
101. University of the South Pacific Fiji
102. University of Toronto Toronto Canada
103. University of Washington Seattle