Stem Cell Reports, Volume *16*

Supplemental Information

Automatic identification of small molecules that promote cell conversion and reprogramming

Francesco Napolitano, Trisevgeni Rapakoulia, Patrizia Annunziata, Akira Hasegawa, Melissa Cardon, Sara Napolitano, Lorenzo Vaccaro, Antonella Iuliano, Luca Giorgio Wanderlingh, Takeya Kasukawa, Diego L. Medina, Davide Cacchiarelli, Xin Gao, Diego di Bernardo, and Erik Arner

Automatic identification of small molecules that promote cell conversion and reprogramming

Francesco Napolitano[1,2]*, Trisevgeni Rapakoulia[2,3]*, Patrizia Annunziata[4], Akira Hasegawa[5], Melissa Cardon[5], Sara Napolitano[1], Lorenzo Vaccaro[4], Antonella Iuliano[1], Luca Giorgio Wanderlingh[1], Takeya Kasukawa[5], Diego L. Medina[1,6], Davide Cacchiarelli[4,6#], Xin Gao[2#], Diego di Bernardo[1,7#], Erik Arner[5,8#]

[1]Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli (NA) 80078, Italy
[2]Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia.
[3]Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, 14195 Berlin, Germany
[4]Telethon Institute of Genetics and Medicine (TIGEM), Armenise/Harvard Laboratory of Integrative Genomics, Pozzuoli (NA) 80078, Italy.
[5]RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, 230-0045 Japan
[6]Department of Translational Medicine, University of Naples Federico II, Naples, Italy
[7]Department of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II, 80125 Naples, Italy.
[8]Graduate School of Integrated Sciences for Life, Hiroshima University, Kagamiyama, Higashi-Hiroshima, 739-8528 Japan

[*]Contributed equally to this work.
[#]Correspondence: d.cacchiarelli@tigem.it, xin.gao@kaust.edu.sa, dibernardo@tigem.it, erik.arner@riken.jp

Supplementary Materials

Supplementary Methods

Conversion to pathway-based profiles

To harmonize the two datasets, we converted the ranked lists of genes from both cell-types (FANTOM5) and drug treatments (LINCS) into *pathway-based expression profiles* (PEPs). A PEP is a transcriptomic profile expressed in terms of pathways as opposed to genes. PEPs were introduced in our previous work (Napolitano et al., 2016) and their efficacy for drug discovery applications was also proved (Napolitano et al., 2018). To convert FANTOM5 and LINCS Gene Expression Profiles (GEPs) to PEPs, we applied the gep2pep Bioconductor package (Napolitano et al., 2019) using all the 14,645 gene sets from 16 different gene set collections included in the MsigDB v6.1 (Liberzon et al., 2015). The gep2pep package iteratively performs Gene Set Enrichment Analysis (GSEA)(Subramanian et al., 2005) to compute Enrichment Scores for each gene set and each expression profile. A PEP is then defined as a ranked list of pathways, each of which is associated with an Enrichment Score (and the corresponding p-value). Once FANTOM5 and LINCS GEPs are converted to PEPs, they can be directly compared (**Supplementary Figure 4A**).

Various pathway-based profiles for the same gene expression profile can be obtained based on the chosen pathway database. In our case, as previously mentioned, we tried 16 different pathway collections available at the MSigDB database. We then evaluated which one out of these 16 collections best captured cell-type similarities, with respect to the Cell Ontology (Bard et al., 2005). To this aim, we used the Cell Ontology annotation of cell-types created by the FANTOM5 consortium (Lizio et al., 2015). In order to obtain a numerical score for each pair of cell-types *i* and *j* in the ontology, we used the Jaccard Index as follows:

$$D_{CO}(i,j)=1-\frac{|C_i \cap C_j|}{|C_i|+|C_j|-|C_i \cap C_j|} \text{ (Jaccard index)}$$

where C_i are the ontology ancestors of cell type *i*, C_j are the ontology ancestors of cell type *j*, $1 \leq i, j \leq 145$, $i \neq j$. Then we defined the PEP-based distance between cell types *i* and *j* using the Manhattan distance as follows:

$$D_P(i,j) = |P_i - P_j|$$

P_i is the PEP of cell type *i*, P_j is the PEP of cell type *j*, $1 \leq i, j \leq 145, i \neq j$.

Finally, we compared the cell distances computed on the PEPs with the same cell distances based on the Cell Ontology (**Supplemental Figure 4B**). The PEPs based on the C2 collection (Canonical Pathways) achieved the highest agreement with the ontology-based similarities, capturing more accurately the known cell hierarchy, even when compared to a previously developed gene-based approach (Iorio et al., 2010).Thus, pathway-based profiles obtained with C2 collection, which includes 250 pathways, were chosen for all further analyses.

Merging of Pathway-based Expression Profiles

As previously proposed (Iorio et al., 2010) we merged multiple expression profiles elicited by the same drug treatment in order to obtain a single "consensus-profile" for each drug, thus enhancing drug-specific effects while reducing unrelated ones. The gep2pep package (Napolitano et al., 2019) supports this operation by averaging the Enrichment Scores over multiple profiles and applying the Fisher method to aggregate their p-values. Using this approach, we merged together all the LINCS profiles induced by the same drug in the same cell line across different concentrations and treatment durations. An additional profile for each drug was generated by averaging all conditions, including different cell lines (termed "independent"). We used both approaches to obtain both cell-specific and cell-independent meta-profiles. We ended up with 17,259 drug-induced PEPs.

Additivity of the drugs at the pathway level

We showed in previous work that the transcriptional response to combinatorial drug treatment at promoters and enhancers is effectively described by a linear combination of the responses of the individual drugs (log2FC values) (Rapakoulia et al., 2017). We used our previous dataset to test if this additive relationship also applies to PEPs. Accordingly, we performed multivariable linear regression analysis, where PEPs of individual drugs were considered as explanatory variables and the PEP of combinatorial drug action as the response variable. We applied our analysis to five pathway databases, Biological Process (BP), Molecular Function (MF), Cellular Component (CC), Transcription Factor Targets (TFT) and Canonical Pathways (C2_CP). **Supplementary Table 6** demonstrates the performance of the linear regression model after ten-fold cross-validation in all the three drug combinations and the four pathway collections. The results show that the linear model using PEPs can describe the relation between single and combinatorial treatment.

To validate whether both single drug PEPs contribute to the model, we performed the same regression analysis 100,000 times with random permutations of one of the single drug PEP. The Pearson

correlation between the observed and predicted values after the permutations was significantly lower for all combinations compared to the regression model based on the non-permuted individual drug PEPs (**Supplementary Figure 2D-F**).

In silico validation with the Pluripotency Score

While the DECCODE framework is based on an unbiased, data-driven approach, we devised a pluripotency-specific method to score gene expression profiles based on prior knowledge about genes involved in the conversion to hIPSCs. We then compared these scores with DECCODE scores to validate the predictions. The pluripotency score (PS) is based on genes that were identified as differentially expressed during reprogramming. In particular, we used the "early pluripotency", "late pluripotency", "early somatic", and "late somatic" gene sets previously identified (Cacchiarelli et al., 2015) that characterize gene expression dynamics in the corresponding stages of conversion from human fibroblasts (HiF-T) to hIPSCs. The original study included also other six sets from the same context, which we used as statistical background. For each of the ten sets and for each drug-induced gene expression profile, we computed an Enrichment Score (ES) using the gep2pep tool. We then ranked them from 1st to 10th according to their ESs, thus obtaining a PEP profile. Finally, we computed the Pluripotency Score (PS) for each profile *p* as:

$$PS(p) = \log\left(\frac{R_{early\ pluripotent}(p)+R_{late\ pluripotent}(p)}{R_{early\ somatic}(p)+R_{late\ somatic}(p)}\right),$$

where $R_x(p)$ is the rank of the gene set *x* within the profile *p*. The score is positive (negative) when genes associated with pluripotency stages tend to be more up-regulated (down-regulated) than genes associated with somatic stages.

Computational validation of the obtained combinations was assessed using PSs (**Figure 1B**). In particular, for any drug combination the median of the corresponding PSs was used. Moreover, the top 30 solutions were considered for a given drug combination size, thus obtaining 30 median PS values. In order to obtain a corresponding null distribution, the same calculation was performed also for random drug combinations of equal size. The random selection was repeated 100 times for each size, thus obtaining 100 times 30 median PSs. **Figure 1B** summarizes this analysis by reporting the obtained 30 versus 300 median PSs for drug combinations of size 1 to 10.

Selection of the drugs for the experimental validation

In order to validate the method experimentally, we selected two lists of drugs: the first using the single-drug approach and the second using the combined approach.

For single drugs, we selected 25 drugs from the top of the DECCODE ranking, plus 20 non-top drugs for comparison. In particular, to build the set of non-top drugs, we chose 10 drugs from the middle of the ranking and 10 drugs from the bottom. In case of an overlap between top and non-top drugs due to the same drugs being profiled across multiple cellular contexts in the LINCS database, we removed the repeated drugs from the non-top sets and chose the next one in the ranking. In all cases, only the drugs included in the Prestwick-FDA library or in the SelleckChem Kinase inhibitors library were considered.

For the drug pairs, we applied further filtering in order to obtain a heterogeneous collection also taking into account the results from the single drug experiments. First, we reran the DECCODE algorithm for drug combinations directly considering only the available drugs in Prestwick-FDA library or in the SelleckChem Kinase inhibitors library. For each drug PEP, we fitted a linear regression model as previously described. We then picked the top 30 drugs and for each of them we selected 20 drugs whose addition to the linear models improves best the Spearman correlation with the hIPSCSs profile. Duplicated solutions were removed, resulting in 522 unique drug combinations. From the remaining pairs, we excluded those containing at least one drug that had already been tested in single drug experiments and showed negative outcome (FC < 1 for number of colonies or covered area). We then selected for experimental validation the top eight combinations having the highest DECCODE ranking and passing the above filters. Since Tazobactam showed particularly encouraging results in primary and secondary reprogramming, we considered two additional drug combinations that included Tazobactam (one such pair was already among the top eight, Tazobactam+Motesanib). Therefore, we finally obtained a list of ten drug pairs covering 16 drugs with Tazobactam included in three different pairs and another two drugs appearing twice (Motesanib and Afatinib).

Colony quantification

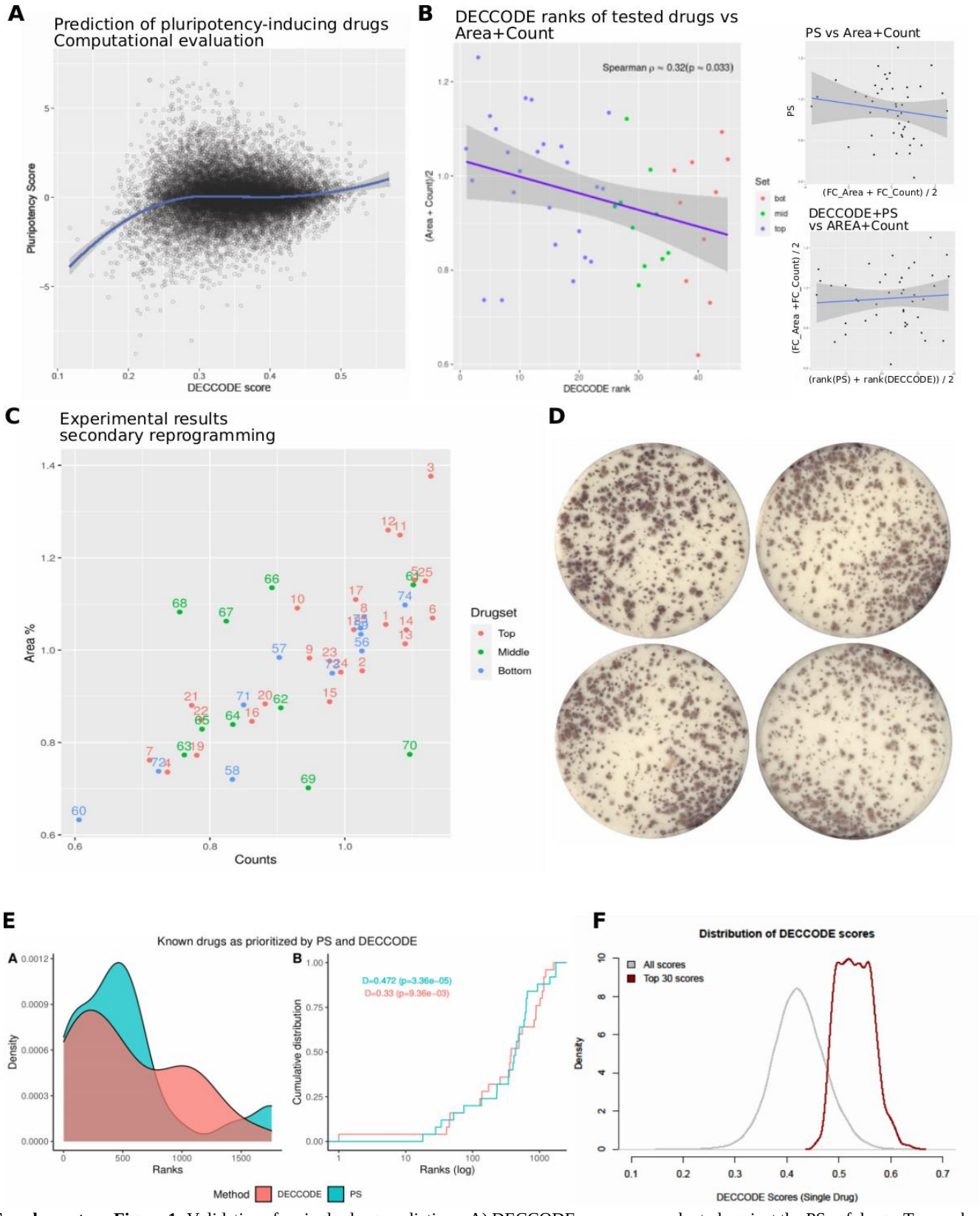
To quantify colony number and size in an unbiased and reproducible way, a completely automated procedure was developed, which is divided in two phases. The first phase was performed through a Matlab script which identifies each well inside all the plate scans, applies a 3X contrast, and saves each of them to a separate image file. The second phase was performed by an ImageJ macro that loads the well images produced by the previous step and performs the final counting and area estimation on each of them. Both Matlab and ImageJ source code, together with the high resolution plate scan images, are available online (Napolitano et al., 2020) (DOI: 10.5281/zenodo.3732772).

In secondary reprogramming experiments, colony count and area values were averaged across the three replicates of the same treatment and across the two controls on the same plate. Average fold change of treatments versus controls were then obtained accordingly (**Figure 2A**, small panels). In order to summarize both count and covered area values together, the corresponding fold changes were averaged (**Figure 2A**, main panel). In the primary reprogramming experiment, counts and areas for tazobactam treatment and controls were compared directly (**Figure 2B**, small panels). Two controls were excluded according to the Bonferroni Outlier Test ($p < 0.0118$ and $p < 0.0106$ respectively). In the case of primary reprogramming results, in order to summarize counts and covered areas together, all the absolute values were normalized dividing by the corresponding mean of the controls (**Figure 2B**, main panel). The same was done for drug combination experimental results (**Figure 3A**).

Computation of DECCODE scores for all the FANTOM5 cell types

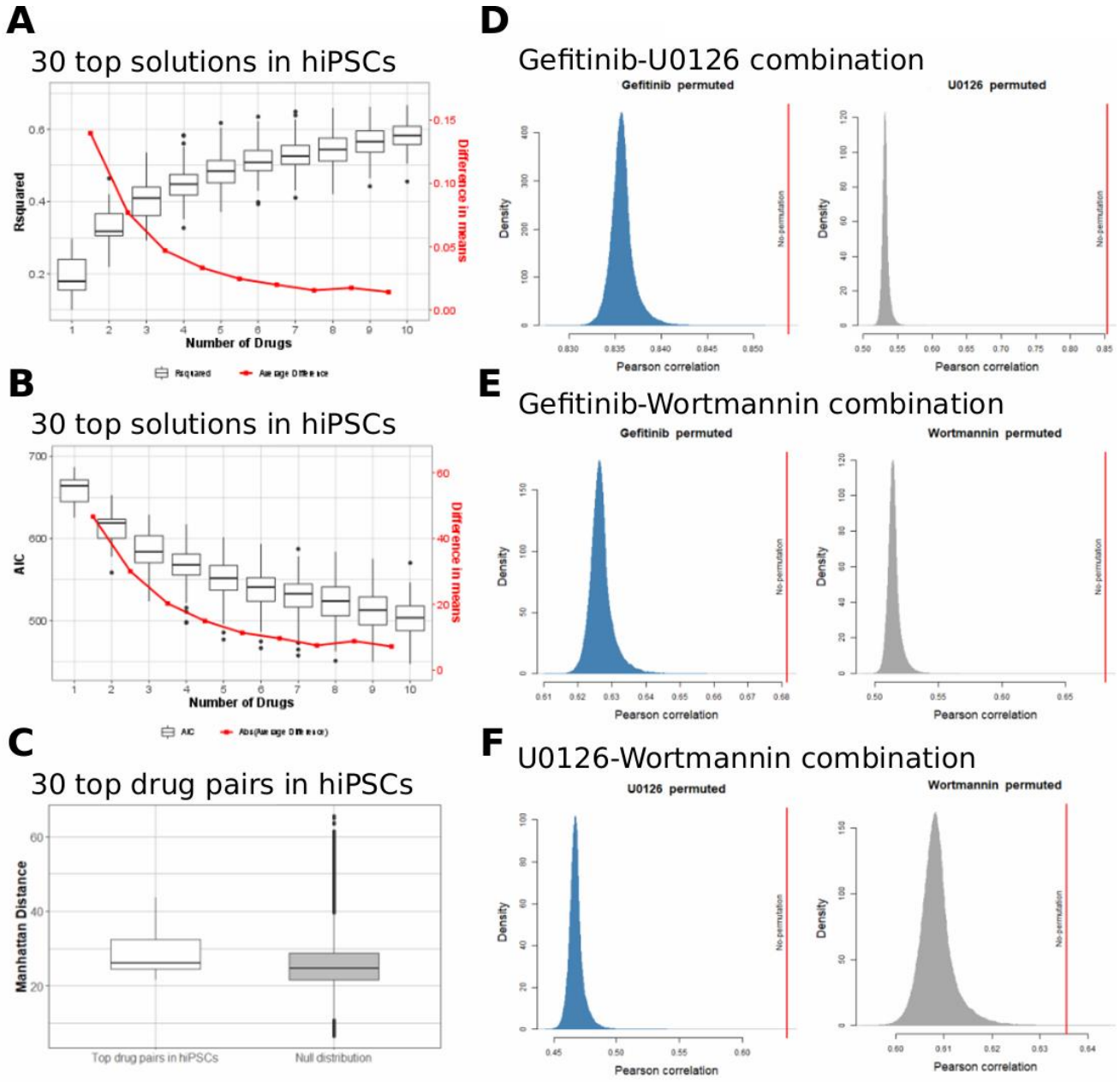
The FANTOM5 cell types include sub-types that are very similar, thus the corresponding expression profiles are not different enough to produce sub-type specific predictions. Therefore, we merged similar cell types to form a single meta-cell profile (see methods subsection “Merging of Pathway-based Expression Profiles”). In order to systematically select which cell-type profiles to merge, we took advantage of the previously computed PEP-based and ontology-based cell type distances (refer to subsection “Conversion to pathway-based profiles”). We applied the Affinity Propagation algorithm (Frey and Dueck, 2007) individually to each of the two pairwise distances to obtain two different clusterings of the same cell types (**Supplementary Figure 3**). Affinity Propagation clustering was performed using the "apcluster" R package (Bodenhofer et al., 2011). Finally, we built a consensus clustering by assigning two cell types to the same cluster if and only if they were assigned to the same cluster by both the ontology-based and PEP-based clusterings. Meta-cell profiles are obtained by averaging all the profiles included in the same cluster. We then computed single-drug and multiple-drug DECCODE scores for all the meta-cell profiles.

Supplementary Figures

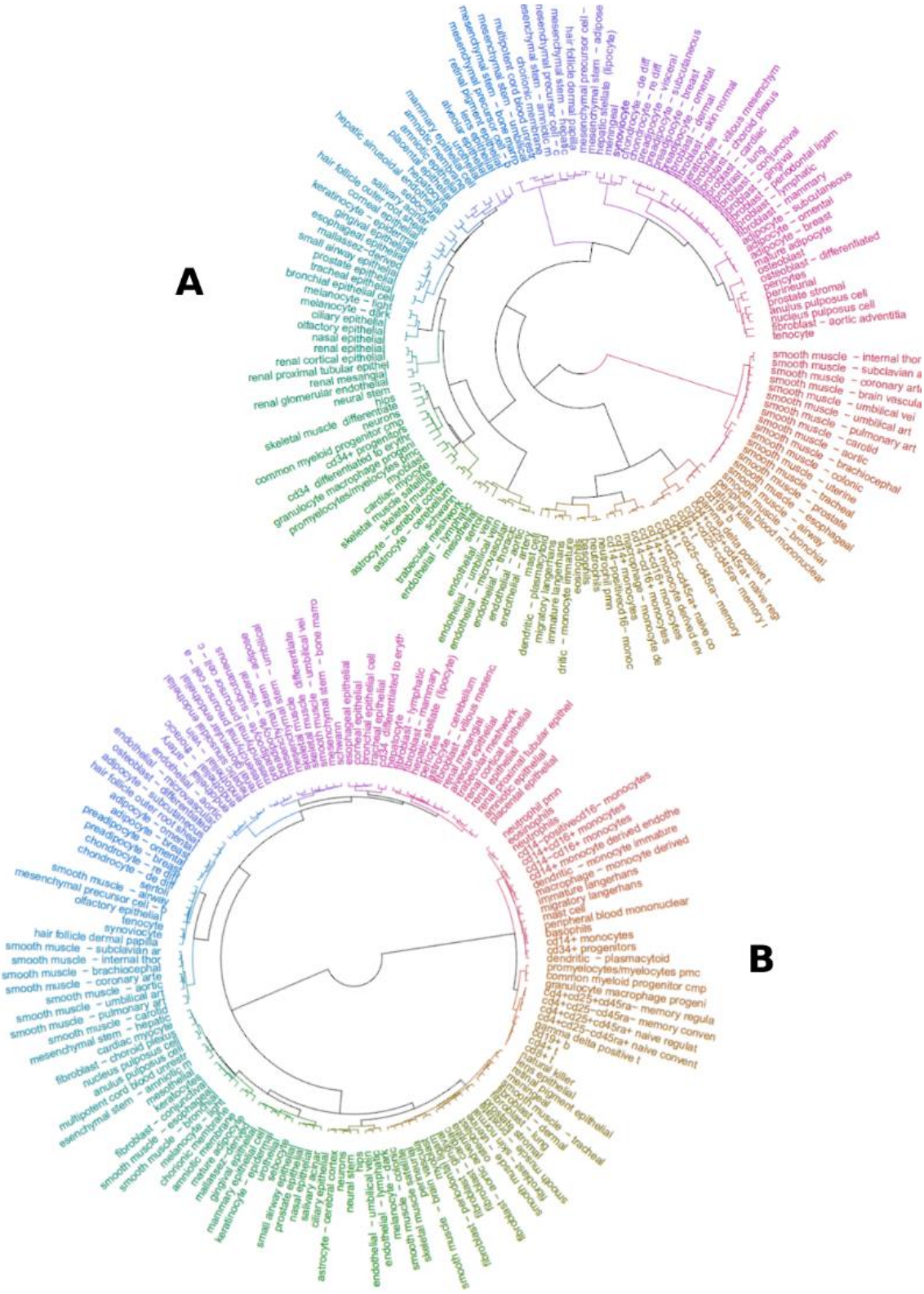


Supplementary Figure 1: Validations for single-drug predictions. A) DECCODE scores are evaluated against the PSs of drugs. Top-ranked (higher DECCODE scores) drugs exhibit higher PSs while bottom-ranked drugs exhibit lower PSs. B) Efficacy of the 45 single drug treatments experimentally tested (colony area and counts) versus DECCODE ranking (left), PS score (top-right) and combined DECCODE

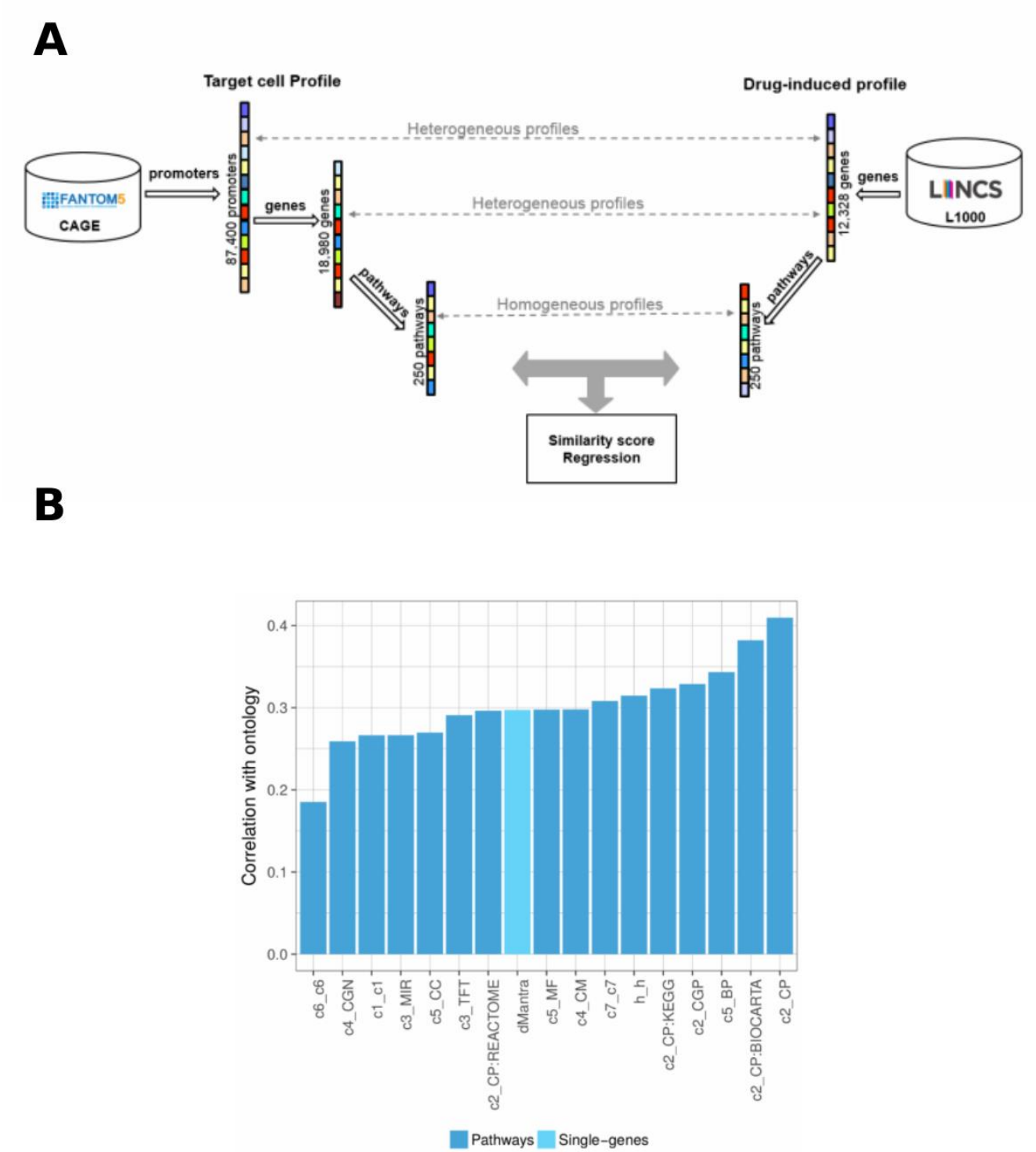
and PS ranking (bottom right). C) Experimental validation of drugs enhancing conversion to hIPSCs: number of colonies formed versus % of covered area. Tazobactam (ID: 3) shows the highest performance for covered. All ID codes are explained in Supplementary Table 2. D) Imaging of wells treated with tazobactam and OSKM (left) against the controls (only OSKM) for the specific plate (right). The fold change in number of colonies and area covered by the colonies for each drug treatment was computed against the control experiments of the corresponding plate. E) Density and Cumulative distribution of the ranks assigned in the small molecules reported in (Chen et al., 2020) based on PS and DECCODE scores. F) Distribution of the top 30 DECCODE drug scores and distribution of all the 1,768 DECCODE drug scores for all meta-cell cluster profiles.



Supplementary Figure 2: Validations for drug combinations. A) Rsquared and B) AIC criterion of the top 30 regression solutions for the hIPSC target profile as more drugs are added to the regression models. Red line highlights the average incremental improvement. C) Distribution of the distances between the drug profiles for the top 30 selected drug pairs in hiPSCs by DECCODE. Null distribution was created by random sampling 1000 drug profiles from LINCS dataset and computing their pairwise distances (499500 distances). D-F) Density plots of the Pearson correlation coefficients between observed and predicted values after the permutations of individual profiles for Gefitinib_U0126 (D), Gefitinib_Wortmannin (E), and U0126_Wortmannin (F) drug combinations using the C2_CP PEPs. The Pearson correlation coefficient achieved without permutation is also reported. Similar results were obtained for all the pathway collections of Supplementary Table 6.



Supplementary Figure 3: Hierarchical clustering visualization of cell types based on the ontology distance (A) and pathway distance (B). Affinity Propagation algorithm (Frey and Dueck, 2007) was applied for the used clustering.



Supplementary Figure 4: A) Harmonization of expression profiles. Promoter-based target cell type profiles are converted to gene-based profiles. Gene-based profiles for both primary cells and drug treated cell lines are then converted to pathway-based expression profiles (PEPs). B) Spearman correlation between pathway distances and ontology distance. Pairwise cell similarity obtained by the different pathway collections are evaluated against cell similarity obtained by the Cell Ontology annotation. Mantra distance (Iorio et al., 2010) computed on single gene ranks is also tested against the ontology distance.

Supplementary Tables

Supplementary Table 1: Top-ranked drugs by DECCODE for the hIPSCs target profile. Several drugs have been already associated with enhancement of the reprogramming process.

Drugs	Indication	ID
Motesanib (Chen et al., 2014)	treatment in solid tumors	1
Fluticasone	activating glucocorticoid receptors, inhibiting nuclear factor kappa b and inhibiting lung eosinophilia in rats	2
Tazobactam	bacterial β-lactamase inhibitor	3
Cyclizine	histamine H1 antagonist	4
Etofenamate (Yang et al., 2011)	nonsteroidal anti-inflammatory drug (NSAID), COX inhibitor	5
Pentoxifylline	modulates immunologic activity by stimulating cytokine production.	6
Irsogladine	anti-inflammatory agent	7
Leflunomide	pyrimidine synthesis inhibitor/chemotherapeutic	8
Dexfenfluramine	serotonergic anorectic drug/studied in obesity	9
Paroxetine	selective serotonin reuptake inhibitor (SSRI) drug commonly known as Paxil	10
Afatinib	tyrosine kinase inhibitor /ErbB family blocker	11
Doramapimod	highly potent p38 MAPK inhibitor	12
Nalbuphine	anticonvulsant effect/inhibited breast cancer cell growth and tumorigenesis	13
PIK-93	PI4KIIIβ inhibitor	14
Glycopyrrolate	synthetic anticholinergic agent	15
SGX523	MET receptor tyrosine kinase inhibitor.	16
Dasatinib (Lin and Wu, 2015)	Src family tyrosine kinase inhibitor	17
SB-203580 (Di Stefano et al., 2016)	inhibitor of p38α and p38β	18
Doxycycline (Chang et al., 2014)	antibacterial agent	19
Saracatinib (Zhang et al., 2014)	inhibitor of the Src/abl family	20
Levetiracetam	plays a role in the control of regulated secretion in neural and endocrine cells	21
Tranylcypromine (Di Stefano et al., 2016)	belongs to a class of antidepressants monoamine oxidase inhibitors (MAOIs).	22
HMN-214	PLK inhibitor	23
histamine	immune responses, neurotransmitter	24
dabrafenib	chemotherapeutic, inhibitor of the associated enzyme B-Raf	25

Supplementary Table 2: Small molecules facilitating reprogramming to hIPSCs reported in (Chen et al., 2020), having an available profile in LINCS database, and their ranking based on PS and DECCODE scores. The DECCODE rankings for both hIPSCs cells and meta-cell cluster 34 (see Supplementary Table 4) which includes hIPSCs as a target profile are reported.

Pert name	**PS ranking**	**DECCODE ranking (hIPSCs)**	**DECCODE ranking (hIPSCs meta-cell)**
BAY-K8644	1434	884	581
BIX-01294	136	1157	783
CHIR-99021	638	1160	368
D-4476	18	369	405
LY-294002	346	494	121
PD-0325901	52	498	53
RG-108	230	554	664
Y-27632	74	378	42
Curcumin	438	258	124
Dexamethasone	589	1022	834
Dovitinib	34	1239	1480
EPZ004777	455	1628	1557
Forskolin	659	129	761
Lenvatinib	1757	355	477
Motesanib	936	1	3
Nintedanib	28	46	168
Pazopanib	633	74	32
Quercetin	613	890	655
Resveratrol	489	41	26
Sorafenib	402	833	710
Sunitinib	415	174	242
Tivozanib	504	138	50
Tranylcypromine	232	45	34
Valproic-acid	1758	1104	1026
Vandetanib	343	359	589
Mean ranking	528.52	553.2	471.4

Supplementary Table 5: Small molecules that were experimentally proved to facilitate various cell conversions and were predicted among the top drug profiles for the corresponding Meta-cells from the DECCODE single and multi-drug approach.

DECCODE Sing Drug Approach		
Target Meta-cell	**Small Molecule**	**Rank**
Astrocyte cells- Cerebral Cortex	Tranylcypromine (Tian et al., 2016)	44
Hepatocyte cells	RG108 (Zhu et al., 2014)	55
hIPS cells - Neural Stem cells	PD0325901 (Lin et al., 2009; Wang et al., 2011; Zhu et al., 2010)	53
hIPS cells - Neural Stem cells	Tranylcypromine (Li et al., 2009; Zhu et al., 2010)	34
Mesenchymal Stem cells - Amniotic membrane - Multipotent Cord Blood Unrestricted Somatic Stem cells	PD0325901 (Lai et al., 2017)	59
Neurons	Y27632 (Hu et al., 2015)	7
Neurons	PD0325901 (Dai et al., 2015)	19

DECCODE Multidrug Approach		
Target Meta-cell	**Small Molecule**	**Rank**
Cardiac Myocyte cells	BIX01294 (Cao et al., 2016)	6
hIPS cells - Neural Stem cells	Tranylcypromine (Li et al., 2009; Zhu et al., 2010)	18
Neurons	PD032590 (Dai et al., 2015)	8
Neurons	Y27632 (Hu et al., 2015)	19

Supplementary Table 6: Pearson and Spearman correlation between fitted and observed PEPs in combinatorial treatment (Rapakoulia et al., 2017). The multivariable linear regression model was applied in five different pathway collections. The values shown in the table are the mean performance after tenfold cross validation.

	BP (4436 pathways)		**MF (901pathways)**		**CC (580 pathways)**		**TFT (615 pathways)**		**C2_CP (250 pathways)**	
	Pearson	**Spearman**	**Pearson**	**Spearman**	**Pearson**	**Spearman**	**Pearson**	**Spearman**	**Pearson**	**Spearman**
Gefitinib-U0126	0.8038	0.7883	0.8685	0.8554	0.8401	0.8401	0.7929	0.8129	0.8538	0.8550
Gefitinib-Wortmannin	0.7460	0.7226	0.7538	0.7203	0.7947	0.6297	0.7276	0.7304	0.6814	0.6751
U0126-Wortmannin	0.6430	0.6211	0.6197	0.6095	0.6826	0.6502	0.6758	0.6336	0.6355	0.6080

References

Bard, J., Rhee, S.Y., and Ashburner, M. (2005). An ontology for cell types. Genome Biol. *6*.

Bodenhofer, U., Kothmeier, A., and Hochreiter, S. (2011). APCluster: an R package for affinity propagation clustering. Bioinformatics *27*, 2463–2464.

Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrawoot, S., Zhang, X., et al. (2015). Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. Cell *162*, 412.

Cao, N., Huang, Y., Zheng, J., Spencer, C.I., Zhang, Y., Fu, J.-D., Nie, B., Xie, M., Zhang, M., Wang, H., et al. (2016). Conversion of human fibroblasts into functional cardiomyocytes by small molecules. Science *1502*.

Chang, M.Y., Rhee, Y.H., Yi, S.H., Lee, S.J., Kim, R.K., Kim, H., Park, C.H., and Lee, S.H. (2014). Doxycycline enhances survival and self-renewal of human pluripotent stem cells. Stem Cell Reports *3*, 353–364.

Chen, G., Xu, X., Zhang, L., Fu, Y., Wang, M., Gu, H., and Xie, X. (2014). Blocking autocrine VEGF signaling by sunitinib, an anti-cancer drug, promotes embryonic stem cell self-renewal and somatic cell reprogramming. Cell Res. *24*, 1121–1136.

Chen, G., Guo, Y., Li, C., Li, S., and Wan, X. (2020). Small Molecules that Promote Self-Renewal of Stem Cells and Somatic Cell Reprogramming. Stem Cell Rev. Reports *16*, 511–523.

Dai, P., Harada, Y., and Takamatsu, T. (2015). Highly efficient direct conversion of human fibroblasts to neuronal cells by chemical compounds. J. Clin. Biochem. Nutr. *56*, 166–170.

Frey, B.J., and Dueck, D. (2007). Clustering by Passing Messages Between Data Points. Science (80-.). *315*, 972–976.

Hu, W., Qiu, B., Guan, W., Wang, Q., Wang, M., Li, W., Gao, L., Shen, L., Huang, Y., Xie, G., et al. (2015). Direct Conversion of Normal and Alzheimer's Disease Human Fibroblasts into Neuronal Cells by Small Molecules. Cell Stem Cell *17*.

Iorio, F., Bosotti, R., Scacheri, E., Belcastro, V., Mithbaokar, P., Ferriero, R., Murino, L., Tagliaferri, R., Brunetti-Pierri, N., Isacchi, A., et al. (2010). Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc. Natl. Acad. Sci. *107*, 14621–14626.

Lai, P.-L., Lin, H., Chen, S.-F., Yang, S.-C., Hung, K.-H., Chang, C.-F., Chang, H.-Y., Lu, F.L., Lee, Y.-H., Liu, Y.-C., et al. (2017). Efficient Generation of Chemically Induced Mesenchymal Stem Cells from Human Dermal Fibroblasts. Sci. Rep. *7*, 44534.

Li, W., Zhou, H., Abujarour, R., Zhu, S., Joo, J.Y., Lin, T., Hao, E., Schöler, H.R., Hayek, A., and Ding, S. (2009). Generation of Human Induced Pluripotent Stem Cells in the Absence of Exogenous *Sox2*. Stem Cells *27*, N/A-N/A.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

Lin, T., and Wu, S. (2015). Reprogramming with Small Molecules instead of Exogenous Transcription Factors. Stem Cells Int. *2015*, 794632.

Lin, T., Ambasudhan, R., Yuan, X., Li, W., Hilcove, S., Abujarour, R., Lin, X., Hahm, H.S., Hao, E., Hayek, A., et al. (2009). A chemical platform for improved induction of human iPSCs. Nat. Methods *6*, 805–808.

Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S., et al. (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. *16*, 22.

Napolitano, F., Sirci, F., Carrella, D., and Di Bernardo, D. (2016). Drug-set enrichment analysis: A novel tool to investigate drug mode of action. Bioinformatics *32*, 235–241.

Napolitano, F., Carrella, D., Mandriani, B., Pisonero-Vaquero, S., Sirci, F., Medina, D.L., Brunetti-Pierri, N., and di Bernardo, D. (2018). gene2drug: a computational tool for pathway-based rational drug repositioning. Bioinformatics *34*, 1498–1505.

Napolitano, F., Carrella, D., Gao, X., and di Bernardo, D. (2019). gep2pep: a bioconductor package for the creation and analysis of pathway-based expression profiles. Bioinformatics.

Napolitano, F., Rapakoulia, T., Annunziata, P., Hasegawa, A., Cardon, M., Napolitano, S., Vaccaro, L., Iuliano, A., Wanderlingh, L.G., Kasukawa, T., et al. (2020). Automatic identification of small molecules that promote cell conversion and reprogramming - plate scans, colony quantification scripts, and DECCODE ranking.

Rapakoulia, T., Gao, X., Huang, Y., De Hoon, M., Okada-Hatakeyama, M., Suzuki, H., and Arner, E. (2017). Genome-scale regression analysis reveals a linear relationship for promoters and enhancers after combinatorial drug treatment. Bioinformatics *33*.

Di Stefano, B., Collombet, S., Jakobsen, J.S., Wierer, M., Sardina, J.L., Lackner, A., Stadhouders, R., Segura-Morales, C., Francesconi, M., Limone, F., et al. (2016). C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. Nat. Cell Biol. *18*, 371–381.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550.

Tian, E., Sun, G., Sun, G., Chao, J., Ye, P., Warden, C., Riggs, A.D., and Shi, Y. (2016). Small-Molecule-Based Lineage Reprogramming Creates Functional Astrocytes. Cell Rep. *16*.

Wang, Q., Xu, X., Li, J., Liu, J., Gu, H., Zhang, R., Chen, J., Kuang, Y., Fei, J., Jiang, C., et al. (2011). Lithium, an anti-psychotic drug, greatly enhances the generation of induced pluripotent stem cells. Cell Res. *21*, 1424–1435.

Yang, C.S., Lopez, C.G., and Rana, T.M. (2011). Discovery of nonsteroidal anti-inflammatory drug and anticancer drug enhancing reprogramming and induced pluripotent stem cell generation. Stem Cells *29*, 1528–1536.

Zhang, X., Simerly, C., Hartnett, C., Schatten, G., and Smithgall, T.E. (2014). Src-family tyrosine kinase activities are essential for differentiation of human embryonic stem cells. Stem Cell Res. 379–389.

Zhu, S., Li, W., Zhou, H., Wei, W., Ambasudhan, R., Lin, T., Kim, J., Zhang, K., and Ding, S. (2010). Reprogramming of Human Primary Somatic Cells by OCT4 and Chemical Compounds. Cell Stem Cell *7*, 651–655.

Zhu, S., Rezvani, M., Harbell, J., Mattis, A.N., Wolfe, A.R., Benet, L.Z., Willenbring, H., and Ding, S. (2014). Mouse liver repopulation with hepatocytes generated from human fibroblasts. Nature *508*, 93–97.