Supplementary Information

Supplementary Data 1. *Reversion of transcriptional signatures of familial Alzheimer's disease (fAD) mutations.* This dataset contains the results of the search for LINCS molecules that might *revert* transcriptional signatures of fAD mutations in SH-SY5Y cells, and the analyses of transcriptional signatures for the tested compounds. We first provide the signatures obtained by differential gene expression measurements of PSEN1 or APP mutated cells relative to WT, and also the additional signatures obtained as indicated in Methods (see *Computational legend* sheet for details). Second, we provide the connectivity (reversion) score of all the previous signatures for each compound in D1 space. Next, we provide the results of the experimental gene expression reversion with the tested compounds, namely Noscapine, Palbociclib, and AG-494 (see the *Reversion legend* sheet for details). In particular, we analyze the best-reversed genes with a normalized score ranging from 1 (an up-regulated gene that is highly down-regulated upon treatment) to -1 (a down-regulated gene that is highly up-regulated upon treatment). Finally, we provide the vendors for the commercial compounds used in this section.

Supplementary Data 2. *Small-molecule analogs of biologics.* This dataset contains the signature 'matching' search to identify small-molecule drugs that might resemble biologics against IL-2 receptor, IL-12 and EGF receptor. Overall, we provide a score for the mimicking of transcriptional signatures obtained from silencing experiments (D1), matching of pathways (C3), biological processes (C4) and network environments (C5) related to the biologic target, and a P-value and a rank for all the comparisons (see the *Legend* sheet for details). Finally, we provide the vendors for the commercial compounds used in this section.



Supplementary Figure 1. *Data pre-processing.* Schematic representation of preprocessing pipelines for each of the 25 CC spaces. Details are given in the Methods.



Supplementary Figure 2. The human-centered scope of the CC covers a significant portion of the known bioactivity space. The figure displays the data available in ChEMBL, grouped by "Organism" and "Target type" (cell-line, protein family, etc.). Size of the circles quantifies the number of targets, and color corresponds to the median number of active molecules per target (10 μ M). The top four organisms in each subplot (Eukaryotes, Bacteria, etc.) are shown based on the number of targets available. The rest of organisms are grouped in the "Other" category. The red arrows indicate the categories covered by our resource (i.e. human data and binding data of all organism types).



Supplementary Figure 3. *Dimensionality reduction (type I signature).* (A) An illustrative scree plot (C3: Signaling pathways), showing the dimension that keeps 90% of the variance (bold dot) and the elbow of the curve (white dot). (B) In the left panel, number of dimensions of the signature type I (bold dot), the elbow point (white dot) and original (raw) dimensions (cross). In the right panel, values are normalized by the original dimensions. Original dimensions correspond to the size of the feature space of the raw data (e.g. number of targets, number of bits in a 2D fingerprint, number of measured genes, number of morphological features, number of yeast mutants, etc.).



Supplementary Figure 4. *MoA and ATC validations.* (A) ROC curves measuring the association between type I signature cosine similarities and the fact that two drugs share a MoA or an ATC code (level 3). (B) Likewise, odds ratio and significance (*** P < 0.0001, ** P < 0.001 and * P < 0.01) of a one-tailed Fisher's exact test performed on a contingency table classifying drug pairs as having the same MoA/ATC code or not, and as belonging to the same cluster or not (see Methods).



Supplementary Figure 5. *Correlations between CC spaces.* Four dataset correlation measures are shown, as described in the Methods.



Supplementary Figure 6. Deeper dive into the partial correlation between D2 and D3 observed in Figure 1C (A and B). Pairwise distances of compounds sharing

(purple) and not sharing MoA (yellow), based on D2 and D3 similarities, respectively. (C and D) Number of compound pairs that capture shared MoA at different specificity and sensitivity cutoffs. In purple, compounds similarities shared between D2 and D3 are counted. Compounds pairs unique to D2 or D3 are counted in the red and blue areas, respectively.



Supplementary Figure 7. *Similarity network embedding (type II signature).* ROC curves of 'link prediction' exercises, performed by removing 20% of the similarity links (P < 0.01). Performance is evaluated at different embedding dimensions.



Supplementary Figure 8. *Similarity-based label prediction.* (A) Proportion of molecule-label pairs (black), molecules (red) and labels (blue) that can be considered

for prediction in each of the CC spaces. We require a label to be annotated with at least 5 molecules in a given CC space. (B) Overlap between true positives across CC spaces. The matrix is read row-wise, e.g. 191 of the 597 true positives obtained by A1 are also found by B1. The color scale is normalized to the diagonal in each row.



Supplementary Figure 9. *Correlation between GDSC drug sensitivity and signature reversion scores.* This plot certifies that transcriptional signature reversion with CC signatures performs as expected with cancer cell lines, i.e. drugs that more strongly reverse transcriptional traits of certain cancer cell lines are indeed more potent against these cells. Upper panels rank, for each cell line, drugs according to their ability to mimic (blue) or revert (red) basal gene expression profiles of the cells, based on D1 CC data. These 'connectivity' scores are correlated to cell sensitivity (expressed as IC50 or AUC of the response curve). Likewise, lower panels rank, for each drug, cell lines according to their basal gene expression 'connectivity' to the drug signature. The Fisher z-transformation is applied to correlation scores to correct for the different number of cell lines drug and drugs per cell line.



Supplementary Figure 10. CRISPR/Cas9-mediated generation of SH-SY5Y cells expressing fAD mutations. (A) Strategy designed to introduce APP^{V717F} and PSEN1^{M146V} mutations in the genome of SH-SY5Y cells. Guide RNA sequence (sgRNA) and single-stranded DNA template are indicated. Silent mutations are introduced to protect the mutated alleles from new CRISPR/Cas9-mediated cuts and also to introduce restriction sites to identify the mutated alleles. Sequences can be found in Supplementary Table 2. (B) Mutated cells can be identified by digestion with the indicated restriction enzymes after reverse transcription followed by PCR amplification of the PSEN1 or APP mutated regions. Clones were routinely tested to confirm their genotype. Representative gels are shown (all replicas (n=3) yielded similar results). (C) Independent APP^{V717F} and PSEN1^{M146V} mutant clones have significantly increased AB₄₂/AB₄₀ ratios compared with control wild-type SH-SY5Y cells, reflecting a relative increase in the generation of A β_{42} . Mean ± SD of 2-5 independent experiments are shown, as illustrated by the dots in each barplot. All clones were compared with APP-PSEN1 WT cells (APP-PSEN1^{WT/WT}). One-sided ttest applied. (D) Normalized $A\beta_{40}$ and $A\beta_{42}$ secretion in differentiated WT SH-SY5Y cells treated with the indicated compounds. Mean ± SD of three independent experiments are shown. One-sample one-sided t-test comparing column means to the reference value of 100. *** P < 0.001, ** P < 0.01, * P < 0.05.



Supplementary Figure 11. *Experimentally tested AD candidate compounds, mapped to the CC.* Candidates for signature reversion in the AD signature reversion experiment are located in 2D projections of the CC. In purple, we show the candidates for the APP^{V717F} mutant, and in green the candidates for the PSEN1^{M146V} mutant.



Supplementary Figure 12. *Experimentally tested biologic mimetics.* Candidates selected in the IL2R (purple), IL-12 (green) and EGFR (orange) computational screening are located in 2D projections of the CC.







Supplementary Figure 14. *Proliferation and viability of resting PBMC treated with PHA*. (A) Resting PBMC were labelled with CFSE and stimulated with PHA. CFSE decay was assessed by flow cytometry three days later. Cell proliferation was quantified as the percentage of cells with reduced CFSE compared with non-stimulated lymphocytes. Values were normalized as percentage of proliferation compared to cells stimulated in the absence of drug. Values were compared with the corresponding vehicle controls (DMSO 0.1% for the left panel and DMSO 0.5% for the right panel). (B) In the same experiment described in (A), viability was quantified by measuring the percentage of events detected in the gate corresponding to living cells. Values were normalized as percentage of viability compared with cells stimulated in the absence of drug. Mean \pm SD of 3-5 independent experiments are shown, as illustrated by the dots in each barplot. Because lack of IL-2 signaling may affect cellular viability, values were compared with the sample treated with daclizumab at 10 µg/ml. **** P < 0.0001, *** P < 0.001, ** P < 0.01, ** P < 0.05 on one-sided t-test.



Supplementary Figure 15. *Proliferation of cell lines.* Proliferation of PBMC stimulated with IL-2 (red) or PHA (grey), as shown in Figure S13A and S14A, respectively, are depicted together with the proliferation of cancer cell lines Jurkat (orange), MT-4 (blue) and HeLa (green). Cells were treated for 72 h with the indicated drugs, and proliferation was measured by the MTT assay. Values were normalized as percentage of cells compared with cells left in the absence of drug. Values were compared with the vehicle control (DMSO 0.1%). Mean ± SD of 2-3 independent experiments are shown, as illustrated by the dots in each barplot. **** P < 0.0001, *** P < 0.001, ** P < 0.01, * P < 0.05 on one-sided t-test.



Supplementary Figure 16. Phosphorylation of STAT5. (A) Pooled results of STAT5 phosphorylation quantification by flow cytometry. Pre-stimulated PBMC were pretreated for 1 h with the indicated doses of compound, and cells were stimulated with IL-2 (0.5 µg/mL) for 20 min. Cells were fixed, and STAT5 phosphorylation was measured by flow cytometry. Values were normalized comparing with cells stimulated in the absence of drug. Mean ± SD of 2-3 independent experiments are shown. **** P < 0.0001, *** P < 0.001, ** P < 0.01, * P < 0.05 on one-sided t-test. (B) Representative flow cytometry results for selected compounds, showing PBMC proliferation (after 72 hours) and STAT5 phosphorylation (after 20 minutes) following IL-2 stimulation. We plot the percentage of cells that have proliferated and the mean fluorescence values corresponding to the phospho-STAT5 staining. A representative experiment, out of 3 replicates, is shown. (C) Representative results for compounds that showed autofluorescence in flow cytometry analysis of STAT5 phosphorylation. Cell proliferation was quantified by flow cytometry and STAT5 phosphorylation was measured by western blot. Representative blots out of 3 independent experiments are shown. Compounds are listed in Supplementary Table 1.



Supplementary Figure 17. Comparison between traditional and modern cell line panels. (A) Shape of the signature I matrices for the current D2 space (NCI60, in CC and hypothetical space built from PharmacoDB blue). а [http://pharmacodb.pmgenomics.ca] (in red); in this case, the input features matrix is composed of drug-gene correlations as provided by PharmacoDB. (B) Similar molecules in D2-NCI60 share MoA (ROC curve). (C) Analogous analysis performed with D2-PharmacoDB. (D) Nearest-neighbours search performed with signatures after dropping out a certain percentage of NCI60 cell lines. Cell line removal was done randomly (white dots) and by prioritizing cell lines based on a random-forest feature importance analysis (filled dots). Performance is measured with an AUC-ROC over an array of similar/dissimilar pairs (10 nearest-neighbours; sampling of negatives was repeated 100 times). (E) Equivalent analysis to A, using CTRP instead of NCI60. (F) An orthogonal case, where similar molecules in D2 are checked for their sharing MoA. Cell lines are iteratively dropped from the signature either randomly or based on the prioritization scheme of A. (G) Equivalent analysis to C, using CTRP instead of NCI60. (H-K) An alternative view on the same analysis, this time focusing on the number of cell lines considered (top 25), rather than the percentage of cell lines dropped.

Supplementary Table 1. *Summary of CC spaces and their sources.* A brief description of the type of data in each CC coordinate is given. Only major data sources are listed. For more details, URLs and references, please see the Methods.

Space	Name	Description	Source(s)
A1	2D fingerprints	Binary representation of the 2D structure of a molecule. The neighborhood of every atom is encoded using circular topology hashing.	RDKIT
A2	3D fingerprints	Similar to A1, the 3D structures of the three best conformers after energy minimization are hashed into a binary representation without the need for structural alignment.	E3FP
A3	Scaffolds	Largest molecular scaffold (usually a ring system) remaining after applying Murcko's pruning rules. In addition, we keep the corresponding framework, i.e. a version of the scaffold where all atoms are carbons and all bonds are single. The scaffold and the framework are encoded with path-based fingerprints, suitable for capturing substructures in similarity searches.	RDKIT
A4	Structural keys	166 functional groups and substructures widely accepted by medicinal chemists (MACCS keys).	RDKIT
A5	Physicochemical parameters	Physicochemical parameters such as molecular weight, logP and refractivity. Number of hydrogen-bond donors and acceptors, rings, etc. Drug-likeness measurements e.g. number of structural alerts, Lipinski's rule-of-5 violations or chemical beauty (QED).	RDKIT and Silico-IT
B1	Mechanisms of action	Drug targets with known pharmacological action and modes (agonist, antagonist, etc.).	DrugBank and ChEMBL
B2	Metabolic genes	Drug-metabolizing enzymes, transporters and carriers.	DrugBank and ChEMBL
В3	Crystals	Small molecules co-crystallized with protein chains. Data are organized on the basis of the structural families of the protein chains.	PDB and ECOD
Β4	Binding	Compound-protein binding data available in major public chemogenomics databases. Data come mainly from academic publications and patents. Binding affinities below a class-specific threshold are favored (kinases \leq 30 nM, GPCRs \leq 100 nM, nuclear receptors \leq 100 nM, ion channels \leq 10 μ M and others \leq 1 μ M), and activities at most one order of magnitude higher are kept (capped at 10 μ M).	ChEMBL and BindingDB
В5	HTS bioassays	Hits from screening campaigns against protein targets (mainly confirmatory functional assays below 10 μM).	PubChem Bioassays (from ChEMBL)
C1	Small-molecule roles	Ontology terms associated with small molecules that have recognized biological roles, such as known drugs, metabolites and other natural products.	ChEBI
C2	Small molecule pathways	Curated reconstruction of human metabolism, containing metabolites and reactions. Data are represented as a network where nodes are metabolites and edges connect substrates and products of reactions.	Recon

СЗ	Signaling pathways	Canonical pathways related to known receptors of compounds (as recorded in B4). Pathways are assigned via a guilt-by-association approach, i.e. a molecule is related to a pathway when at least one of the targets is a member of it.	Reactome
C4	Biological processes	Similar to C3, biological processes from the gene ontology are associated with compounds via a guilt-by- association approach from B4 data. All parent terms are kept, from the 'leaves' of the ontology to its 'root'.	Gene Ontology
C5	Interactome	Neighborhoods of B4 targets are collected by inspecting several protein-protein interaction networks. A random- walk algorithm is used to obtain a robust measure of 'proximity' in the network.	STRING, InWeb and Pathway Commons, among others
D1	Gene expression	Transcriptional response of cell lines upon exposure to small molecules. A reference collection of gene expression profiles is used to map all compound profiles using a two-sided gene set enrichment analysis.	L1000 Connectivity Map (Touchstone reference)
D2	Cancer cell lines	Small-molecule sensitivity data (GI50) of a panel of 60 cancer cell lines.	NCI-60
D3	Chemical genetics	Growth inhibition profiles in a panel of ~300 yeast mutants. Data are combined with yeast genetic interaction data, so that compounds can be assimilated to genetic alterations when they have similar profiles.	MOSAIC
D4	Morphology	Changes in U-2 OS cell morphology measured after compound treatment using a multiplexed-cytological 'cell painting' assay. 812 morphology features are recorded via automated microscopy and image analysis.	LINCS Portal
D5	Cell bioassays	Small-molecule cell bioassays reported in ChEMBL. Mainly, growth and proliferation measurements found in the literature.	ChEMBL
E1	Therapeutic areas	Anatomical Therapeutic Chemical (ATC) classification codes of drugs. All ATC levels are considered.	DrugBank and KEGG
E2	Indications	Indications of approved drugs and drugs in clinical trials. A controlled medical vocabulary is used.	DrugBank and ChEMBL
E3	Side effects	Side effects extracted from drug package inserts via text- mining techniques.	SIDER
E4	Diseases and toxicology	Manually curated relationships between chemicals and diseases. Chemicals include drug molecules and environmental substances, among others.	СТD
E5	Drug-drug interactions	Changes in the effect of a drug when is co-administered with a second drug. Data are related to pharmacokinetic issues and/or adverse events.	DrugBank

Supplementary Table 2. *Compounds used in the signature reversion predictions.* Drugs were purchased from the indicated suppliers or kindly provided by The Broad Institute or the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (US). CD25/IL-2 R alpha antibody (daclizumab) was purchased from Novus Biologicals. APE1 inhibitor III is an analog of the APE1 inhibitor initially identified in the computational screening.

Compound InChlKey	Name or formula	Supplier	Assay
AKNNEGZIBPJZJG-ZWKOTPCHSA-N	Noscapine (17)	Tocris	fAD cells
AHJRHEGDXFFMBM-UHFFFAOYSA-N	Palbociclib (18)	Selleckchem	fAD cells
HKHOVJYOELRGMV-XYOKQWHBSA-N	AG-494 (19)	Abcam	fAD cells
DCVZSHVZGVWQKV-UHFFFAOYSA-N	H-9 dihydrochloride	Tocris	IL-2 proliferation
GIFYKGIODMXYNT-UHFFFAOYSA-N	ML133	Sigma-Aldrich	IL-2 proliferation
IMLGWNWCXQYTLB-UHFFFAOYSA-N	STOCK1S-29444	InterBioScreen	IL-2 proliferation
IQNTXIMXKCURDC-YBEGLDIGSA-N	JAK3 inhibitor VI	Merck	IL-2 proliferation
KXDRZYJTTYXIEO-DHZHZOJOSA-N	STOCK5S-29655	InterBioScreen	IL-2 proliferation
MORBXZMIXGYQDB-UHFFFAOYSA-N	ML202	NCATS	IL-2 proliferation
PIXJURSCCVBKRF-UHFFFAOYSA-N	АТРА	Tocris	IL-2 proliferation
PMDAVRLCSZSLRH-UHFFFAOYSA-N	BRD-K06217810-001-01-7 (20)	NCATS	IL-2 proliferation
POKKSIROKFEGGD-UHFFFAOYSA-N	ML077	Aobious	IL-2 proliferation
RWQKHEORZBHNRI-BMIGLBTASA-N	Ochratoxin A	Tocris	IL-2 proliferation
UWYZHKAOTLEWKK-UHFFFAOYSA-N	1,2,3,4- Tetrahydroisoquinoline	Sigma-Aldrich	IL-2 proliferation
VNYMEPQAFKNDRQ-UHFFFAOYSA-N	Z55175877 (21)	EnamineStore	IL-2 proliferation
WOXKDUGGOYFFRN-IIBYNOLFSA-N	Tadalafil	Tocris	IL-2 proliferation
WWUZIQQURGPMPG-KRWOKUGFSA-N	Sphingosine	Sigma-Aldrich	IL-2 proliferation
XOLMRFUGOINFDQ-YBEGLDIGSA-N	RO-3306	Sigma-Aldrich	IL-2 proliferation
XPLJEFSRINKZLC-ATVHPVEESA-N	SU11652 (22)	Enzo	IL-2 proliferation
YRYMOQJJPIZJFQ-UHFFFAOYSA-N	MSL002264496	NCATS	IL-2 proliferation
	CID 25243782		
YVCXQRVVNQMZEI-UHFFFAOYSA-N	WHI-P97	MedChemExpress	IL-2 proliferation
AXRCEOKUDYDWLF-UHFFFAOYSA-N	WAY-262611 (23)	MedChemExpress	IL-2 proliferation

RPQZTTQVRYEKCR-WCTZXXKLSA-N	Zebularine	Sigma-Aldrich	IL-12 IFNG stimulation
WJRWSLORVIHRNX-UHFFFAOYSA-N	VU0240551	Tocris	IL-12 IFNG stimulation
VJICXSGFDHBEPP-UHFFFAOYSA-N	CID 44460125	NCATS	IL-12 IFNG stimulation
QAGFDTOVQKCZRK-UHFFFAOYSA-N	CID 44142086	NCATS	IL-12 IFNG stimulation
IYRMWMYZSQPJKC-UHFFFAOYSA-N	Kaempferol (24)	Sigma-Aldrich	IL-12 IFNG stimulation
VGCKEWCZFDSELD-UHFFFAOYSA-N	CID 350929	Sigma-Aldrich	IL-12 IFNG stimulation
FTSUPYGMFAPCFZ-UHFFFAOYSA-N	Quinpirole dihydrochloride	Tocris	IL-12 IFNG stimulation
FLOSMHQXBMRNHR-UHFFFAOYSA-N	Methazolamide	Sigma-Aldrich	IL-12 IFNG stimulation
BZGQMGJZXODGJM-UHFFFAOYSA-N	CID 1998940	NCATS	IL-12 IFNG stimulation
KQHKZXBKJWOHHU-UHFFFAOYSA-N	CID 3647906	NCATS	IL-12 IFNG stimulation
RPXVIAFEQBNEAX-UHFFFAOYSA-N	CNQX	Tocris	IL-12 IFNG stimulation
JMSPCTGDYFVMJZ-UHFFFAOYSA-N	APE1 inhibitor III (25)	Merck	EGFR downregulation

Supplementary Table 3. Sequences for CRISPR/Cas9 gene editing and analysis.

Primer	Sequence (5'-3')
APP V/17F SgRNA_Forward	
APP V717F sgRNA_Reverse	AAACAGGTGATGACGATCACTGTC
ssODN APP V717F template	TTCTTAATTTGTTTTCAAGGTGTTCTTTGCAGAAGATGTGGGTTCAAACAAA
PSEN1 M146V sgRNA_Forward	CACCGTGTTGTCATGACTATCCTCC
PSEN1 M146V sgRNA_Reverse	AAACGGAGGATAGTCATGACAACAC
ssODN PSEN1 M146V template	AGAATCTATACCCCATTCACAGAAGATACCGAGACTGTGGGCCAGAGAGCCCTGCACTCA ATTCTGAATGCTGCCATCATGATCAGTGTCATTGTTGTGGTAACCATCCTCCTGGTGGTT CTGTATAAATACAGGTGCTATAAGGTGAGCATGAGACACAGATCTTTGCTTTCCACCCTG TTCTTCTTATGGTTGGGTAT
gDNA PSEN1_Forward	ACTCTGCAGATGAGAGGCAC
gDNA PSEN1_Reverse	TGTTCCACAGTGAGGAGGAAG
gDNA APP_Forward	AAGCGCTATCTTCCCACCAC
gDNA APP_Reverse	ACCCAAGCATCATGGAAGCA
cDNA PSEN1_Forward	GCGGCGGGAAGCGTATACC
cDNA PSEN1_Reverse	CGTGACTCAGGTGTAGAGCG
cDNA APP_Forward	CGAAGTTGAGCCTGTTGATGC
cDNA APP_Reverse	AGGTTGGATTTTCGTAGCCGT
qPCR IFNG_Forward	TCGGTAACTGACTTGAATGTCCA
qPCR IFNG_Reverse	TCGCTTCCCTGTTTTAGCTGC
qPCR GAPDH_Forward	TTGAGGTCAATGAAGGGGTC
qPCR GAPDH_Reverse	GAAGGTGAAGGTCGGAGTCA