Electronic Supplementary Material (ESI) for ChemComm.

Electronic Supporting Information

Correspondence should be addressed to G.J.L.B.: gbernardes@medicina.ulisboa.pt, gb453@cam.ac.uk

Table of contents

1. Methods

1.1 Machine learning

SPiDER. Target prediction was carried out on the publicly available web server (www.cadd.ethz.ch/software/spider.html) as previously reported.[1-5] In short, celastrol was projected onto two self-organizing maps together with reference compounds from the COBRA database.[6] Chemical structures are processed with the "wash" function of the Molecular Operating Environment (MOE, Chemical Computing Group, Montreal, Canada), prior to description with the CATS2[7] and MOE2D descriptors. Predictions are carried out by calculating the Euclidean distances of the molecules to the reference compounds in COBRA. The output comprises target families at a confidence level of $p < 5\%$. The distances are converted to p values, according to a pre-calculated background distribution of distances between molecules annotated to bind different targets. The arithmetic average of these p values serves as confidence score for the target prediction. As such, each prediction can be associated with another p value that indicates the statistical significance of the prediction.[2]

DEcRyPT 2.0. Ligand and bioactivity data was collected from ChEMBL22 and filtered as previously described.[8] Ligands were normalized with the "wash" function in MOE 2015.10 and bioaffinity data (K_i, K_D, IC_{50} or EC_{50}) transformed to the respective antilog value (p*Affinity*). Only molecules previously tested against single proteins of *Homo sapiens* were considered. Regression random forest models were built for each individual target using CATS2 descriptors, provided that >50 ligands were reported per target.[9] The models (trees = 500; max features = squared root; min_samples_split = 2) were subjected to stratified 10-fold cross validation and mean absolute errors calculated. To calculate the background p*Affinity* distribution for each of the 1026 targets, we collected 139,352 molecules previously identified as dark chemical matter (DCM).[10] A Score value [Score = $(X_{Pred}–\mu_{DCM})/(\sigma_{DCM}/n^{0.5})$; X_{Pred}: Predicted value for celastrol; μ_{DCM}: Average of predicted values for DCM; σ_{DCM}: Standard deviation of predictions for DCM; $n^{0.5}$: rescaling factor, where n equals number of targets] was calculated to provide further assistance in the prioritization of assays.

Projection of chemical space. Ligand data for cannabinoid and progesterone receptors was collected and pre-processed as described for DEcRyPT 2.0. The CATS2 descriptors were calculated and the descriptor space projected to the plane using the *t*-distributed stochastic neighborhood embedding (*t*-SNE) algorithm (learning rate = 600; iterations = 1000).

1.2 Biochemistry

Functional and radioligand displacement assays were performed at Cerep (France) through well-established and validated methods[11-15] and on a fee-for-service basis, as outlined in Tables S1-2.

Table S1. Summary of functional assay conditions.

Assay	Source	Stimulus / Ligand	Incubation	Measured component	Detection
CB1 (agonist)[13]	CHO cells	None	20 min, 37 °C	cAMP	HTRF
CB1 (antagonist)[13]	CHO cells	CP55940 (0.3 nM)	20 min, 37 °C	cAMP	HTRF
CB2 (agonist)[13]	CHO cells	None	10 min, 37 °C	cAMP	HTRF
CB2 (antagonist)[13]	CHO cells	WIN 55212-2 (3 nM)	10 min, 37 °C	cAMP	HTRF
PR (agonist)[14]	Human recombinant	None	RT	Coactivator recruitment	AlphaScreen
PR (antagonist)[14]	Human recombinant	Progesterone (100 nM)	RT	Coactivator recruitment	AlphaScreen
VDR (agonist)[15]	Human recombinant	None	30 min, RT	Coactivator recruitment	AlphaScreen
VDR (antagonist)[15]	Human recombinant	Calcitriol (100 nM)	30 min, RT	Coactivator recruitment	AlphaScreen

CB1 – cannabinoid receptor-1; CB2 – cannabinoid receptor-2; PR – progesterone receptor; VDR – Vitamin D receptor. Controls: CB1 agonist, EC_{50}(CP55940) = 0.029 nM. CB1 antagonist, IC_{50}(AM281) = 9.3 nM. CB2 agonist, EC_{50}(WIN55212-2) = 0.14 nM. CB2 antagonist, IC_{50}(AM630) = 650 nM. PR agonist, IC_{50}(progesterone) = 16 nM. PR antagonist, IC_{50}(mifepristone) = 12 nM. VDR agonist, EC_{50}(calcitriol) = 3.9 nM.

Table S2. Summary of binding (radioligand displacement) assay conditions.

Assay	Source	Ligand	Non-specific	Incubation	Detection
CB1 (agonist radioligand)[11]	CHO cells	[^{3}H]CP55940 (0.5 nM)	WIN55212-2 (10 µM)	120 min, 37 °C	Scintillation counting
CB2 (agonist radioligand)[12]	CHO cells	[^{3}H]WIN55212-2 (0.8 nM)	WIN55212-2 (5 µM)	120 min, 37 °C	Scintillation counting

CB1 – cannabinoid receptor-1; CB2 – cannabinoid receptor-2. Controls: CB1 agonist radioligand, K_i(CP55940) = 0.94 nM (nHill = 0.8); CB2 agonist radioligand, K_i(WIN55212-2) = 1.4 nM (nHill = 0.9).

1.3 Dynamic Light Scattering

Dynamic light scattering (Zetasizer Nano S, Malvern, UK) was used to determine compound colloidal aggregation potential. The particle sizes were measured at 25 °C. A 100 mM stock solution of celastrol was prepared in DMSO, following dilution to deionized water to obtain an analyte solution of 100 μM (0.1% DMSO). Colloidal aggregation was measured through sequential dilutions.

1.4 Bioinformatics

Analysis of the NCI-60 panel data. Normalized gene expression (averaged intensity of expression values combined from five microarray platforms) and drug activity (*z*-transformed - $\log GI_{50}$ values; higher values correspond to higher sensitivity to celastrol) data for NCI-60 cancer cell lines were downloaded from the CellMiner database (https://discover.nci.nih.gov/cellminer/loadDownload.do). Spearman's correlation, run with the cor.test R function (method = ''spearman''),[16] was used to test the interdependence between *CB1* or *CB2* expression levels and celastrol activity in 59 cancer cell lines (those for which both types of data were available). The Kruskal-Wallis rank sum test for differences between distributions was implemented using the kruskal.test R function.[16]

TCGA. Publicly available RNAseqV2 read counts, quantified with RSEM,[17] and clinical data for 9,721 tumour and 725 matched-normal samples from The Cancer Genome Atlas (TCGA; https://cancergenome.nih.gov/) were downloaded from Firebrowse (http://firebrowse.org/). Read counts were further quantile-normalized using voom.[18] Wilcoxon rank-sum tests for differences between distributions were performed using the wilcox.test R function.[16]

2. Supplementary data

2.1 Cross-validation metrics

Table S3. Summary of obtained metric values.

	DEcRyPT 2.0		Y-randomization test[b]	
Target	MAE	MSE	MAE	MSE
Mineralocorticoid	0.491	0.423	0.778	0.969
PTP[a]	0.384-0.642	0.276-0.971	0.818-1-767	0.970-5.551
Progesterone	0.513	0.469	1.048	1.691
Glutamate ionotropic[a]	0.564-0.695	0.690-1.214	0.980-1.362	1.550-3.190
Liver X[a]	0.489-0.503	0.401-0.450	0.894-0.999	1.226-1.586
TRP channel	0.556	0.527	1.074	1.794
Phospholipase[a]	0.347-0.588	0.285-0.543	0.732-1.158	0.844-2.083
Smoothened	0.407	0.288	0.772	0.929
GPCR19	0.641	0.683	1.093	1.834
Polymerase	0.554	0.536	0.749	1.165
Farnesoid X	0.452	0.408	0.853	1.127
Vitamin D	0.410	0.464	1.777	4.598
Estrogen[a]	0.533-0.588	0.560-0.600	1.147-1.299	1.997-2.651
Microtubules	n.a.	n.a.	n.a.	n.a.
Androgen	0.498	0.451	1.024	1.632
Glucocorticoid	0.459	0.400	0.891	1.287
Aromatase	0.608	0.671	1.175	2.201
Integrins	0.512	0.468	1.409	3.088
Cannabinoid[a]	0.543-0.572	0.499-0.600	1.079-1.117	1.773-1894
Acid glycoprotein	n.a.	n.a.	n.a.	n.a.

[a] Considers several proteins of the same family. [b] p*Affinity* values were shuffled and a random forest model was built and subjected to stratified 10-fold cross-validation using the same hyperparameters as in DEcRyPT 2.0. MAE: Mean absolute error. MSE: Mean squared error.

2.2 Radioligand displacement assays

Table S4. Radioligand displacement data for celastrol against the CB1 and CB2 receptors.

		% Inhibition of control specific binding	
Receptor	**Concentration / μM**	**1st measurement**	**2nd measurement**
CB1	0.1	-21.0	-21.0
	1.0	-43.4	-20.8
	10	9.7	22.4
	100	85.4	102.8
CB2	0.1	10.6	12.3
	1.0	-16.1	-14.0
	10	-15.6	-31.8
	100	43.8	66.3

Control: CB1, CP55940 K_i = 0.94 nM (nHill = 0.8); CB2, WIN 555212-2 K_i = 1.4 nM (nHill = 0.9).

2.3 Dynamic Light Scattering

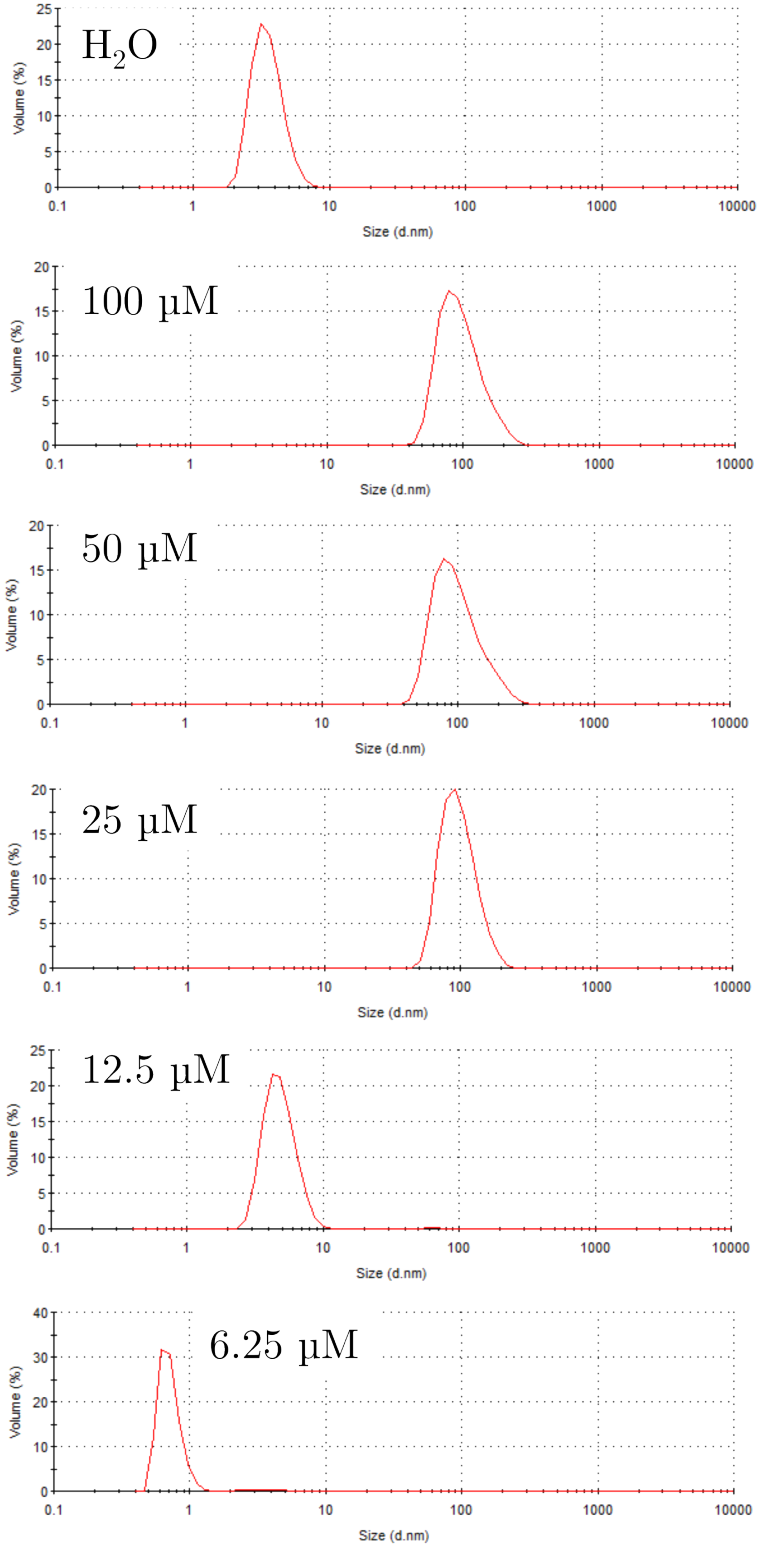

Figure S1. Particle volume measured by dynamic light scattering at different concentrations of celastrol (H_2O + 0.1% DMSO).

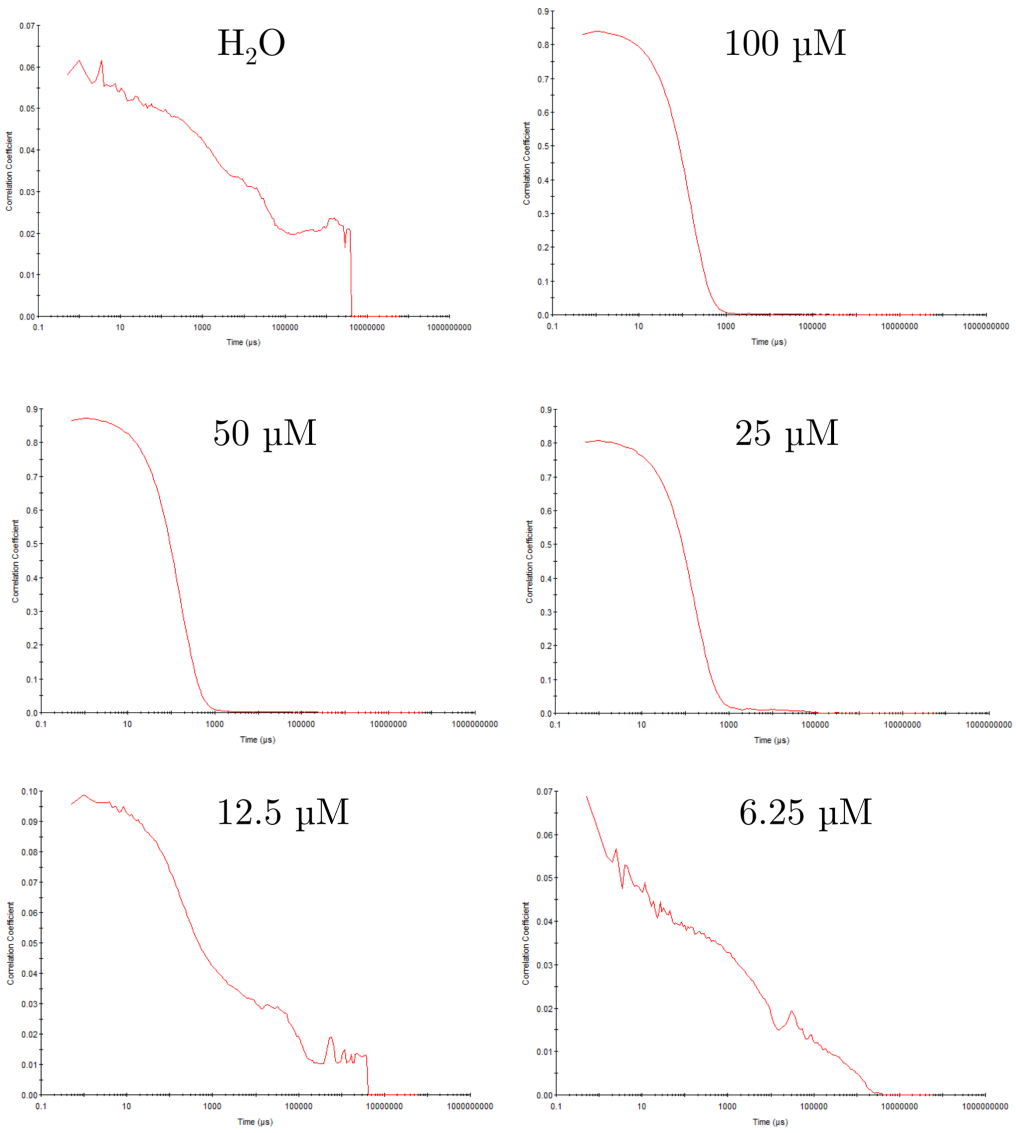

Figure S2. Autocorrelation curves obtained for celastrol at different concentrations.

2.4 Bioinformatics

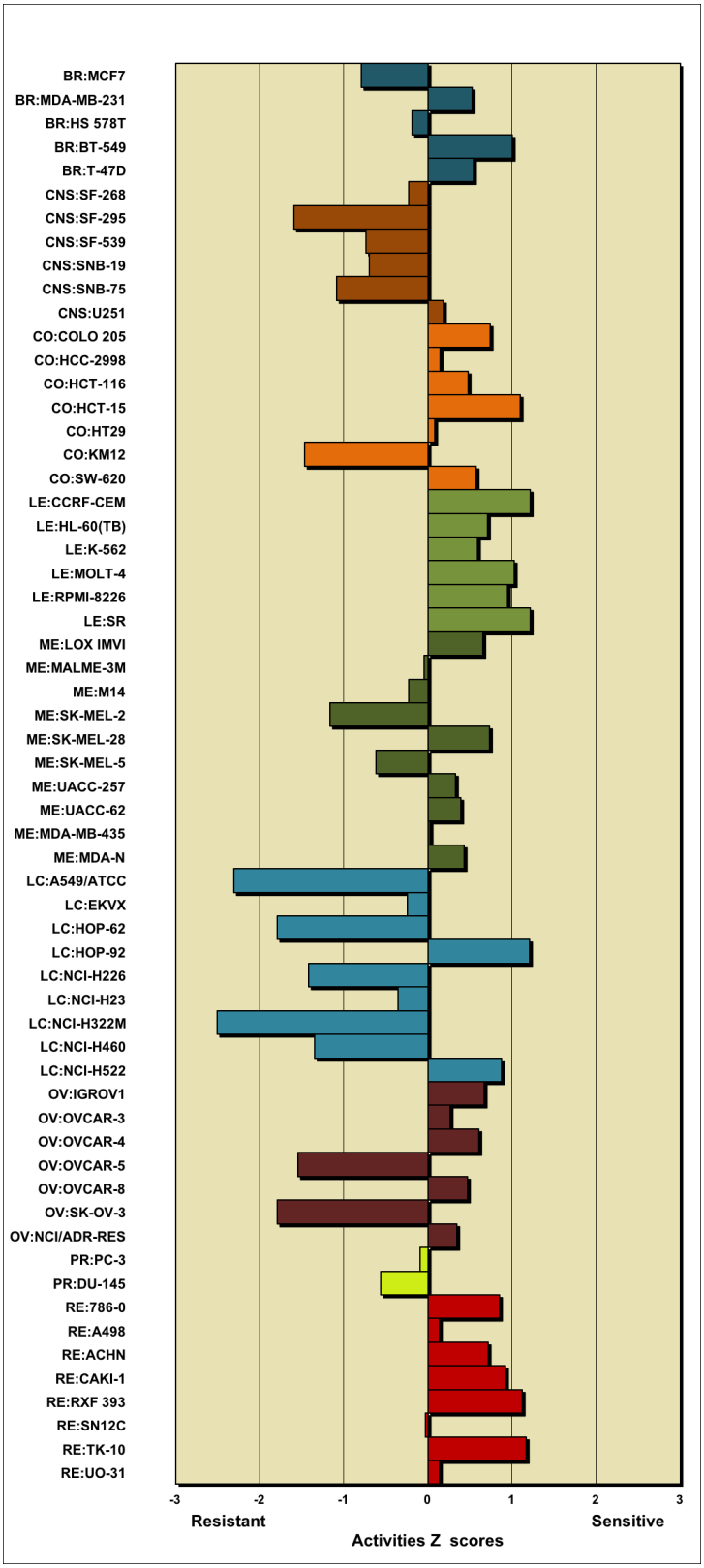

Figure S3. Celastrol activity on NCI-60 cancer cell lines. Bar plots of celastrol activity (-logGI_{50}) z-scores) against NCI-60 cancer cell lines. Cell lines are grouped by tissue of origin. Plot generated at https://discover.nci.nih.gov/cellminer/.

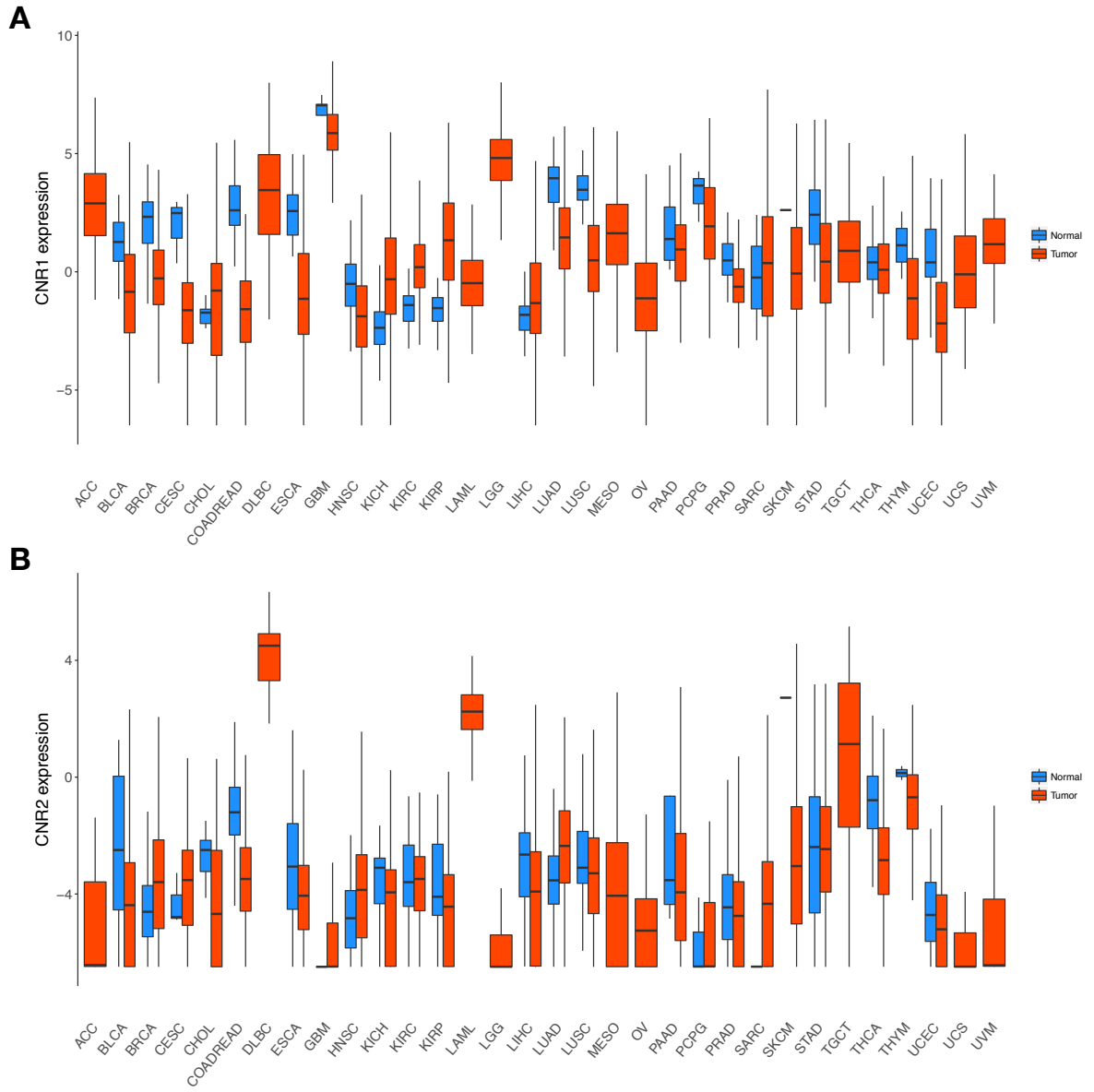

Figure S4. Cannabinoid receptor-1 (*CNR1*) and 2 (*CNR2*) expression in human cancer tissues. Box plots of (A) *CNR1* and (B) *CNR2* expression across TCGA cancer types, coloured by sample type (tumour and matched-normal samples are represented in red and blue, respectively).

ACC: adrenocortical carcinoma; BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; CESC: cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL: cholangiocarcinoma; COADREAD: colon and rectum adenocarcinoma; DLBC: lymphoid neoplasm diffuse large B-cell lymphoma; ESCA: oesophageal carcinoma; GBM: glioblastoma multiforme; HNSC: head and neck squamous cell carcinoma; KICH: kidney chromophobe; KIRC: kidney renal clear cell carcinoma; KIRP: kidney renal papillary cell carcinoma; LAML: acute myeloid leukemia; LGG: low-grade glioma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; MESO: mesothelioma; OV: ovarian serous cystadenocarcinoma; PAAD: pancreatic adenocarcinoma; PCPG: pheochromocytoma and paraganglioma; PRAD: prostate adenocarcinoma; SARC: sarcoma; SKCM: skin cutaneous melanoma; STAD: stomach adenocarcinoma; TGCT: testicular germ cell tumours; THCA: thyroid carcinoma; THYM: thymoma; UCEC: uterine corpus endometrial carcinoma; UCS: uterine carcinosarcoma; UVM: uveal melanoma.

Table S5. Results of cannabinoid receptor-1 differential expression analyses between tumour and matched-normal TCGA samples. Related to Figure S3A.

	Normal samples	Tumor samples	Mann-Whitney p value	FDR	Mean normal samples	Mean tumor samples	Log2FC
BRCA	112	1100	1.15E-33	1.73E-32	2.109	-0.255	-2.364
COADREAD	51	382	1.96E-29	1.47E-28	2.785	-1.778	-4.563
LUSC	51	501	1.11E-20	5.53E-20	3.591	0.503	-3.089
KIRC	72	534	2.64E-20	9.91E-20	-1.440	0.323	1.763
LUAD	59	517	4.10E-20	1.23E-19	3.743	1.372	-2.372
PRAD	52	498	4.01E-11	1.00E-10	0.509	-0.603	-1.111
KIRP	32	291	6.78E-11	1.45E-10	-1.609	1.142	2.752
STAD	35	415	4.99E-07	9.35E-07	2.589	0.399	-2.190
UCEC	24	177	3.70E-06	6.16E-06	0.630	-1.817	-2.448
HNSC	44	522	7.27E-06	1.09E-05	-0.628	-1.918	-1.290
ESCA	11	185	2.08E-05	2.65E-05	2.475	-0.966	-3.441
BLCA	19	408	2.12E-05	2.65E-05	1.274	-0.896	-2.170
KICH	25	66	2.66E-05	3.07E-05	-2.265	-0.024	2.241
LIHC	50	373	2.36E-02	2.53E-02	-1.865	-1.148	0.717
THCA	59	509	8.88E-02	8.88E-02	0.446	0.169	-0.278

Table S6. Results of cannabinoid receptor-2 differential expression analyses between tumour and matched-normal TCGA samples. Related to Supplementary Figure S3B.

Cohort	Normal samples	Tumor samples	Mann-Whitney pvalue	FDR	Mean normal samples	Mean tumor samples	Log2FC
COADREAD	51	382	2.50E-16	3.76E-15	-1.24	-3.60	-2.35
THCA	59	509	7.74E-16	5.80E-15	-0.76	-2.92	-2.16
BRCA	112	1100	1.92E-06	9.61E-06	-4.52	-3.60	0.91
LUAD	59	517	1.47E-05	5.52E-05	-3.43	-2.46	0.97
LIHC	50	373	8.44E-05	2.53E-04	-3.13	-4.23	-1.09
BLCA	19	408	2.12E-03	5.31E-03	-2.31	-4.26	-1.95
HNSC	44	522	8.28E-03	1.78E-02	-4.63	-3.91	0.72
LUSC	51	501	3.40E-02	6.38E-02	-2.65	-3.38	-0.72
KIRP	32	291	5.06E-02	8.44E-02	-3.81	-4.56	-0.74
KICH	25	66	6.00E-02	9.01E-02	-3.41	-4.23	-0.82
ESCA	11	185	1.32E-01	1.81E-01	-2.76	-3.84	-1.08
UCEC	24	177	1.45E-01	1.81E-01	-4.60	-5.08	-0.48
PRAD	52	498	2.35E-01	2.72E-01	-4.39	-4.64	-0.25
KIRC	72	534	3.75E-01	4.02E-01	-3.55	-3.70	-0.15
STAD	35	415	6.63E-01	6.63E-01	-2.58	-2.37	0.20

Table S7. Correlation between targets predicted by SPiDER / DEcRyPT 2.0 and the antiproliferative activity of celastrol.

SPiDER	Gene name	Gene symbol	Spearman_coef	Spearman_pvalue
Cannabinoid	Cannabinoid Receptor 1	CNR1	0,290	0,026
Mineralocorticoid	Mineralocorticoid receptor	NR3C2	-0,282	0,030
Integrins	Integrin Subunit Alpha 1	ITGA1	-0,259	0,048
Smoothened	Smoothened	SMO	-0,248	0,058
Phospholipase	Phospholipase A2 Group IVA	PLA2G4A	-0,240	0,067
Glucocorticoid	Glucocorticoid Receptor	NR3C1	-0,143	0,281
Androgen	Androgen Receptor	AR	-0,131	0,323
Polymerase	DNA Polymerase Alpha 1, Catalytic Subunit	POLA1	0,112	0,400
Aromatase	Aromatase	CYP19A1	-0,099	0,455
Estrogen	Estrogen receptor 1	ESR1	-0,093	0,484
Vitamin D	Vitamin D Receptor	VDR	-0,092	0,489
Acid glycoprotein	Alpha-1-Acid Glycoprotein 1	ORM1	0,069	0,605
Glutamate ionotropic	glutamate ionotropic receptor NMDA type subunit 1	GRIN1	-0,060	0,651
Liver X	Liver X Receptor Alpha	NR1H3	-0,052	0,694
Cannabinoid	Cannabinoid Receptor 2	CNR2	-0,047	0,722
Progesterone	Progesterone receptor	PGR	0,040	0,762
PTPb	Protein Tyrosine Phosphatase, Receptor Type B	PTPRB	0,028	0,832
TRP channel	TRPV1	TRPV1	-0,020	0,878
Microtubules	Tubulin Alpha 1a	TUBA1A	0,017	0,900
Farnesoid X	Farnesoid X Receptor	NR1H4	-0,011	0,937
GPCR19	GPCR19	GPBAR1	NA	NA

Table S8. Correlation between targets reported in the literature for celastrol and its antiproliferative activity.

Gene name	Gene symbol	Reference	Spearman_coef	Spearman_pvalue
Proteasome 26S subunit, ATPase 1	PSMC1	https://www.ncbi.nlm.nih.gov/pubmed/16651429	-0,358	0,005
Hypoxia-inducible factor-1Œ±	HIF1A	https://www.spandidos-publications.com/10.3892/ijmm.2011.600	-0,224	0,089
Heat Shock Protein 90 Alpha Family Class B Member 1	HSP90AB1	https://www.spandidos-publications.com/10.3892/ijmm.2011.600	0,064	0,630
NFKB Inhibitor Alpha	NFKBIA	https://www.spandidos-publications.com/etm/14/1/819	-0,046	0,731
Fanconi Anemia Complementation Group D2	FANCD2	https://onlinelibrary.wiley.com/doi/full/10.1111/cas.12679	-0,030	0,823
Vascular Endothelial Growth Factor A	VEGFA	https://www.spandidos-publications.com/10.3892/ijmm.2011.600	0,006	0,965

3. References

1. D. Reker, A. M. Perna, T. Rodrigues, P. Schneider, M. Reutlinger, B. Monch, A. Koeberle, C. Lamers, M. Gabler, H. Steinmetz, R. Muller, M. Schubert-Zsilavecz, O. Werz and G. Schneider, Nat. Chem., 2014, **6**, 1072-1078.
2. D. Reker, T. Rodrigues, P. Schneider and G. Schneider, Proc. Natl. Acad. Sci. U.S.A., 2014, **111**, 4067-4072.
3. T. Rodrigues, D. Reker, J. Kunze, P. Schneider and G. Schneider, Angew. Chem. Int. Ed., 2015, **54**, 10516-10520.
4. T. Rodrigues, D. Reker, P. Schneider and G. Schneider, Nat. Chem., 2016, **8**, 531-541.
5. G. Schneider, D. Reker, T. Chen, K. Hauenstein, P. Schneider and K. H. Altmann, Angew. Chem. Int. Ed., 2016, **55**, 12408-12411.
6. P. Schneider and G. Schneider, QSAR Comb. Sci., 2003, **22**, 713-718.
7. M. Reutlinger, C. P. Koch, D. Reker, N. Todoroff, P. Schneider, T. Rodrigues and G. Schneider, Mol. Inf., 2013, **32**, 133-138.
8. T. Rodrigues, N. Hauser, D. Reker, M. Reutlinger, T. Wunderlin, J. Hamon, G. Koch and G. Schneider, Angew. Chem. Int. Ed., 2015, **54**, 1551-1555.
9. T. Rodrigues, M. Werner, J. Roth, E. H. G. da Cruz, M. C. Marques, P. Akkapeddi, S. A. Lobo, A. Koeberle, F. Corzana, E. N. da Silva Junior, O. Werz and G. J. L. Bernardes, Chem. Sci., 2018, **9**, 6899-6903.
10. A. M. Wassermann, E. Lounkine, D. Hoepfner, G. Le Goff, F. J. King, C. Studer, J. M. Peltier, M. L. Grippo, V. Prindle, J. Tao, A. Schuffenhauer, I. M. Wallace, S. Chen, P. Krastel, A. Cobos-Correa, C. N. Parker, J. W. Davies and M. Glick, Nat. Chem. Biol., 2015, **11**, 958-966.
11. R. S. Martin, L. A. Luong, N. J. Welsh, R. M. Eglen, G. R. Martin and S. J. MacLennan, Br. J. Pharmacol., 2000, **129**, 1707-1715.
12. S. Munro, K. L. Thomas and M. Abu-Shaar, Nature, 1993, **365**, 61-65.
13. C. C. Felder, K. E. Joyce, E. M. Briley, J. Mansouri, K. Mackie, O. Blond, Y. Lai, A. L. Ma and R. L. Mitchell, Mol. Pharmacol., 1995, **48**, 443-450.
14. S. A. Onate, V. Boonyaratanakornkit, T. E. Spencer, S. Y. Tsai, M. J. Tsai, D. P. Edwards and B. W. O'Malley, J. Biol. Chem., 1998, **273**, 12101-12108.
15. C. X. Yuan, M. Ito, J. D. Fondell, Z. Y. Fu and R. G. Roeder, Proc. Natl. Acad. Sci. U.S.A., 1998, **95**, 7939-7944.
16. R. C. Team, 2017.
17. B. Li and C. N. Dewey, BMC bioinformatics, 2011, **12**, 323.
18. C. W. Law, Y. Chen, W. Shi and G. K. Smyth, Genome Biol., 2014, **15**, R29.