Predicting Putative Chemical Targets using Transcriptomics

EU-ToxRisk
Oct 31, 2018

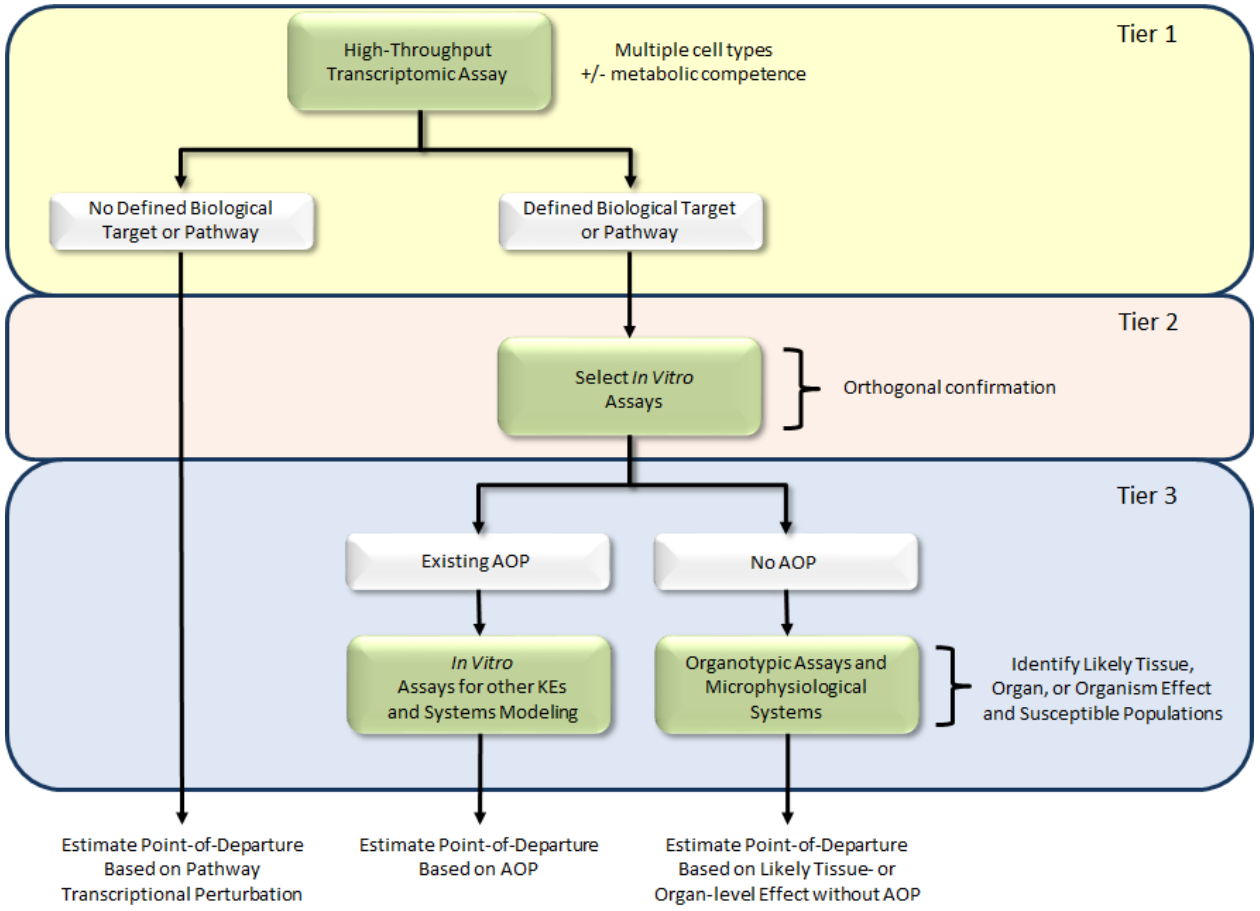
Imran Shah
NCCT/ORD/EPA

Objectives

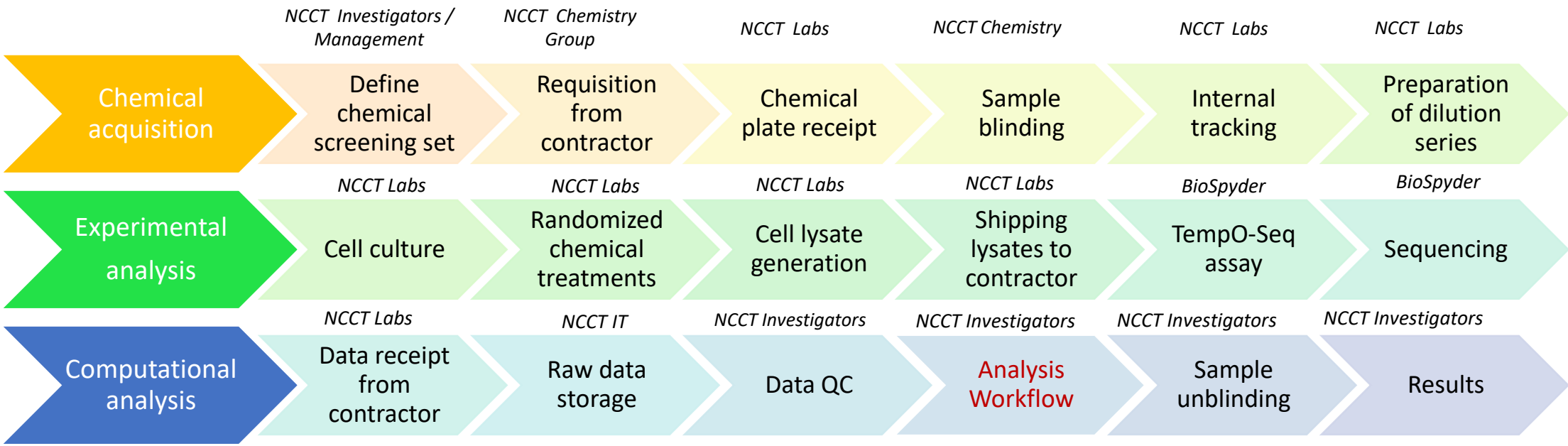
- A flexible, portable and cost efficient platform to comprehensively evaluate the potential biological pathways and processes impacted by chemical exposure
  → High-throughput transcriptomics (HTTr)

- Identify the concentration at which biological pathways/processes begin to be impacted
- Predict biological targets for chemicals with specific modes-of-action

A strategic vision and operational road map for computational toxicology at the U.S. Environmental Protection Agency [DRAFT]



HTTr Workflow



NCCT HTTr Project Team

National Center for Computational Toxicology

Joshua **Harrill**
Toxicologist

Clinton **Willis**
NSSC (JH)

Imran **Shah**
Computational Systems Biologist

R. Woodrow **Setzer**
Mathematical Statistician

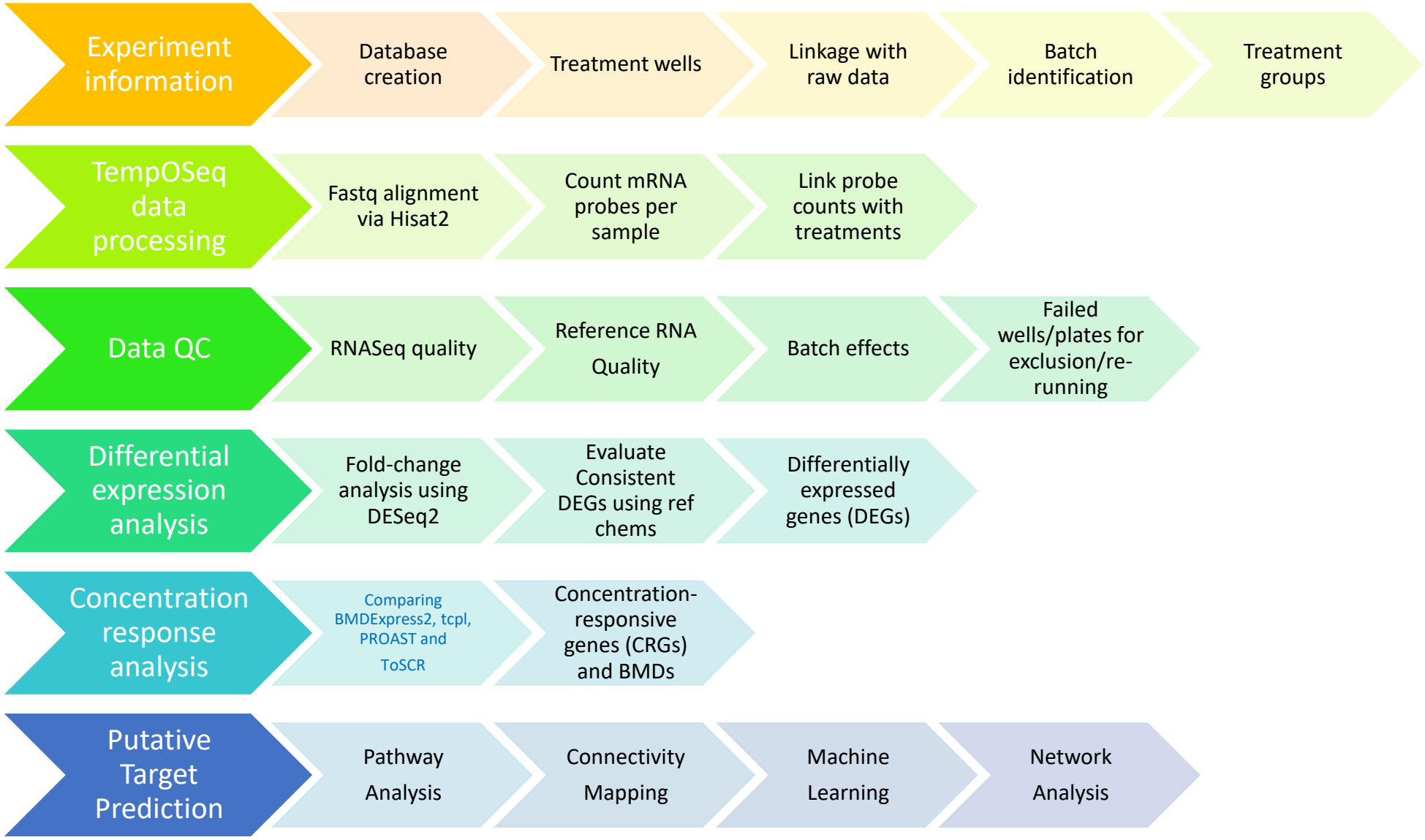
Derik **Haggard**
ORISE Fellow

Thomas **Sheffield**
ORISE Fellow

Richard **Judson**
Bioinformatician

Russell **Thomas**
Director

HTTr Analysis Pipeline (Oct 2018)



HTTr Phase I Screen

- Chemicals
  - EPA Sample ID = **2,112**
  - Unique dsstox_sid = 2,049 (63 duplicates)
  - CMap v 2.0 overlap: 222/2,049 (~10%) (7/222 are duplicates)
- Study (Batches)
  - Batch level 1: Study conducted in four “blocks”
    - block_id ∈ {1,2,3,5}
  - Batch level 2: 48 groups of replicate plates (“plate groups”)
    - plate_group_id ∈ {1,2,3,…,48}
  - Batch level 3: 144 plates
    - plate_id ∈ {'TC00284655', 'TC00284656', 'TC00284657’,….,}
- HTTr profiles
  - 54,432 samples
  - 21,111 probes (12,330 seen)
  - 19,287 genes (~10,000 seen)

- Chemicals and Batches
  - Each plate group has 44 chemicals
  - Batch 1 has 396 chemicals, 2/3 have 704 chemicals each and 5 has 308 chemicals
- Treatment groups (17,040)
  - 2112 test chemicals, 8 concs (8x2112) = 16,896
  - 3 Reference chemicals, 1 concs (3x48) = 144

block_id	Chems
1	396
2	704
3	704
5	308

http://bitbucket.zn.epa.gov/projects/HTTR

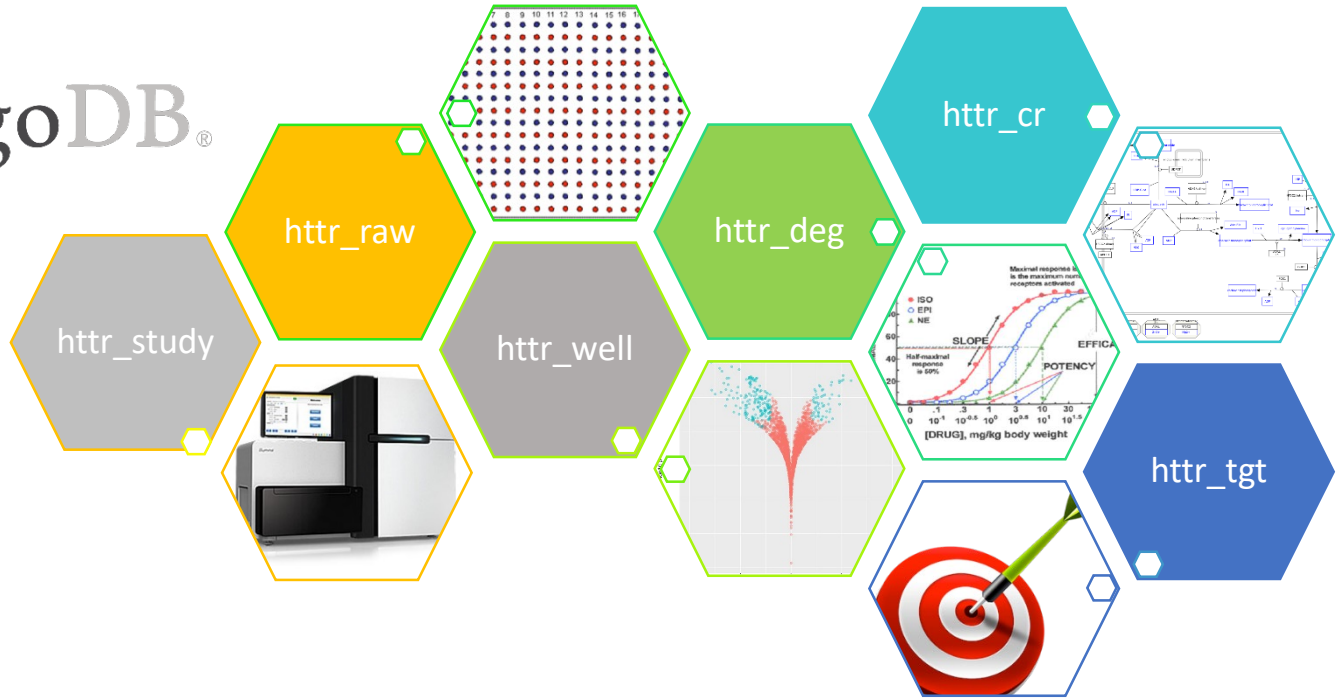
HTTr Computational Framework

Python & R analysis pipeline

jupyter

TempOSeq data processing → Differential expression analysis → Concentration response analysis → Target/ Pathway Prediction

mongoDB



REST API

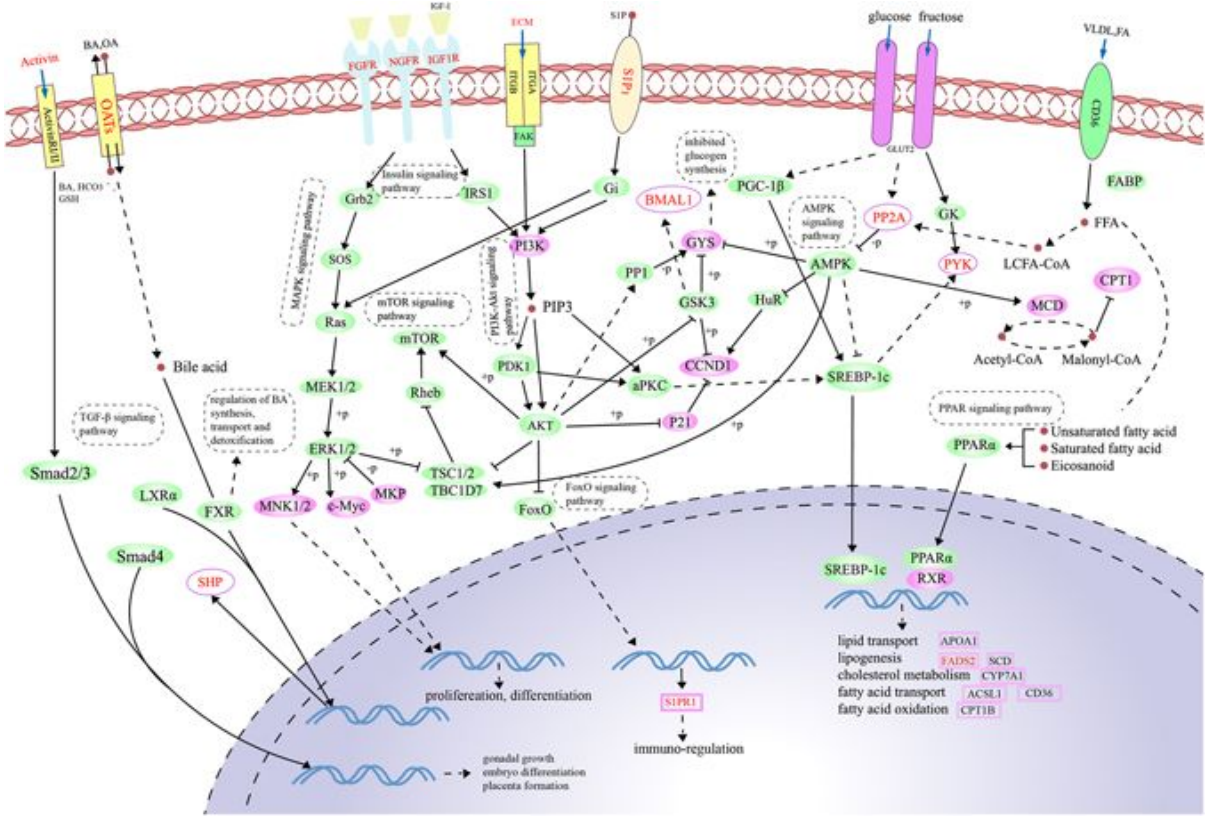
Flask

```
http://httr-dev.epa.gov/api/httr/v1/

searchChem
getChemPlates
getPlateInfo
getPlateGroups
getChemProbeCounts
getChemDEG

getChemCRG
getChemTargets
```

Predicting Chemical Targets is Challenging

How can we use gene-expression data to predict putative targets?

- Data-driven: chemicals that produce similar mRNA profiles have similar targets:
  - Connectivity mapping: infer targets based on KNN using diverse metrics
  - Pathway analysis: Does not always predict target but can be helpful in some cases
  - Signature analysis: Build target-specific mRNA signature
- Infer targets using mRNA and network analysis



https://www.nature.com/articles/s41598-017-06634-w

Predicting Putative Chemical Targets

Connectivity mapping	"Pathway" analysis	Signature creation	Network analysis
Compare entire transcriptomic profile to reference database to find target using kNN	Compare entire transcriptomic profile to pathways (bags of genes) to find pathway 'hits' using different scoring schemes	Using transcriptomic profiles for each target to create classifiers ("signature"); predict targets using httr profile of test chemicals	Use transcriptomic profile alone with genetic-regulatory and signaling data to infer putative targets

<u>Key questions</u>

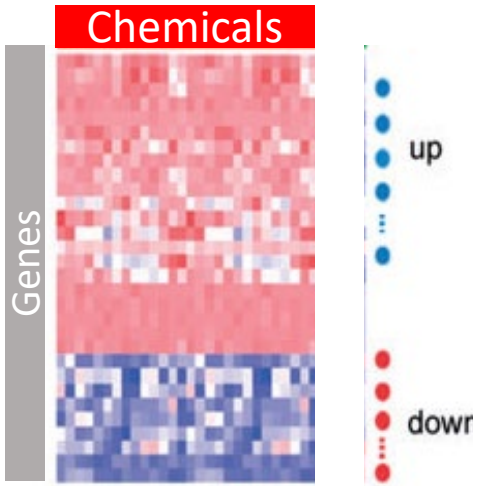
- Accuracy: can the approach correctly predict targets? Are the reference chemicals correctly characterised?
- Data: are there sufficient data for the approach to work?
- Annotations: does the approach rely on manual curation?
- Scalability: can the approach effectively scale to thousands of chemicals and potential targets?

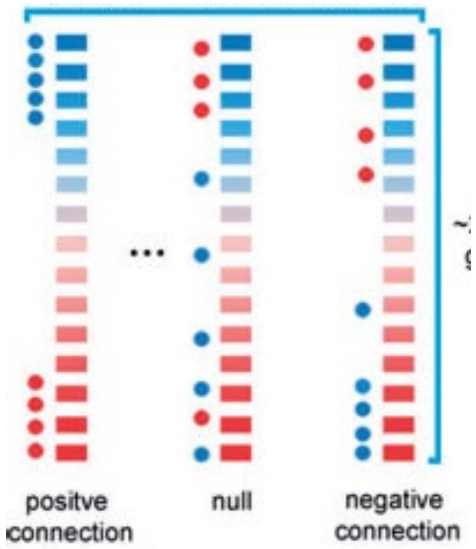
Target Prediction by Connectivity Mapping

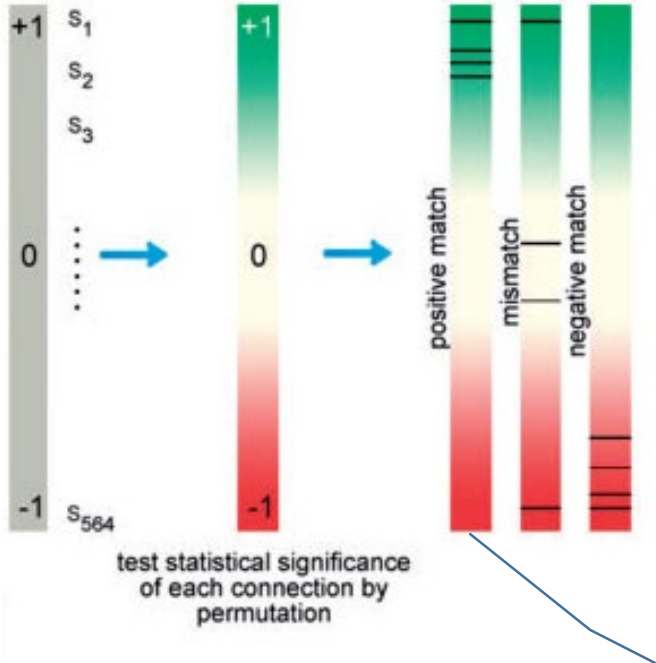
Input DEGs or CRGs

Query Signature DB CMap or BSP

Find best positive matches







BioSpyder HTTr (BSP)

Infer Targets by best match

Lamb *et al* (2006)
Musa *el al* (2017)

cMap vs cMap

Database signature creation

- Use CMap v2 database: Affymetrix data on 1176 chemicals, 5 cell lines
- Translate FC profiles in up/down profiles (signatures)
  - Convert L2FC data to standardized Z vector
  - For z0=1,2,3 create discrete Z where value = 1 if Z>z0 and -1 where Z<z0
- Store signatures in database for rapid searching

Target	pos	neg	pos_annot	BA	Sn	Sp	th0
ADRA	52	402	57	0.59	0.56	0.63	0.12
SCN	48	397	52	0.55	0.48	0.61	0.09
DRD	42	394	47	0.65	0.64	0.65	0.17
HRH	42	387	45	0.60	0.58	0.63	0.18
HTR	37	398	43	0.65	0.65	0.65	0.15
COX2	33	353	34	0.57	0.52	0.62	0.07
ADRB	31	354	36	0.56	0.51	0.61	0.09
KCN	30	380	34	0.50	0.39	0.60	0.11
CHRM	28	365	35	0.62	0.61	0.62	0.12
COX1	27	355	27	0.55	0.50	0.61	0.07
GR	23	355	24	0.61	0.61	0.62	0.09
CACN	22	358	22	0.63	0.64	0.62	0.11
ER	16	279	17	0.64	0.62	0.65	0.22
PDE	13	289	14	0.75	0.88	0.61	0.07
NAT	13	278	17	0.70	0.76	0.63	0.22
HTT	13	271	16	0.68	0.73	0.63	0.21
MAO	11	236	13	0.70	0.78	0.62	0.07
PR	11	288	11	0.65	0.68	0.62	0.12
PPAR	10	247	11	0.61	0.61	0.61	0.08

BSP (Q) vs CMap (DB)

gene	media	timeh	BA	Sn	Sp	neg	pos
AR	DMEM	6h	0.39	0.43	0.35	73	4
ESR	DMEM	6h	0.59	0.67	0.51	77	3
HMGCR	DMEM	6h			0.40	57	2
PDE	DMEM	6h		0.00	0.62	59	2
PPAR	DMEM	6h	0.56	0.56	0.57	148	6

Results are shown for Pilot study
Results for PhI screen worse
Could match trichostatin A but not genistein and sirolimus profiles!

Predicting Putative Chemical Targets

Connectivity mapping

Compare entire transcriptomic profile to reference database to find target using kNN

Pros:
- Need just one profile / target

Cons:
- Sensitive but not specific within platform
- Low cross-platform accuracy
- Requires pre-anntoted examples (labour)

Pathway analysis

Compare entire transcriptomic profile to pathways (bags of genes) to find pathway 'hits' using different scoring schemes

Signatures/classifiers

Using transcriptomic profiles for each target to create classifiers ("biomarker panel"); predict targets for using httr profile of test chemicals

Network analysis

Use transcriptomic profile alone with genetic-regulatory and signaling data to infer putative targets

Target Elucidation by Pathway Analysis

- “Pathways” means many things:
  - Canonical pathways
  - Just a bag of genes could be “signature” of biological activity
- Two relevant approaches:
  - Over-representational analysis (ORA). Evaluate pathway membership for subset of highest DEGs (shown to have limited accuracy)
  - Gene set analysis (GSA). Compare entire transcriptomic profile to pathways (bags of genes) to find pathway ‘hits’ using different scoring schemes (more promising)

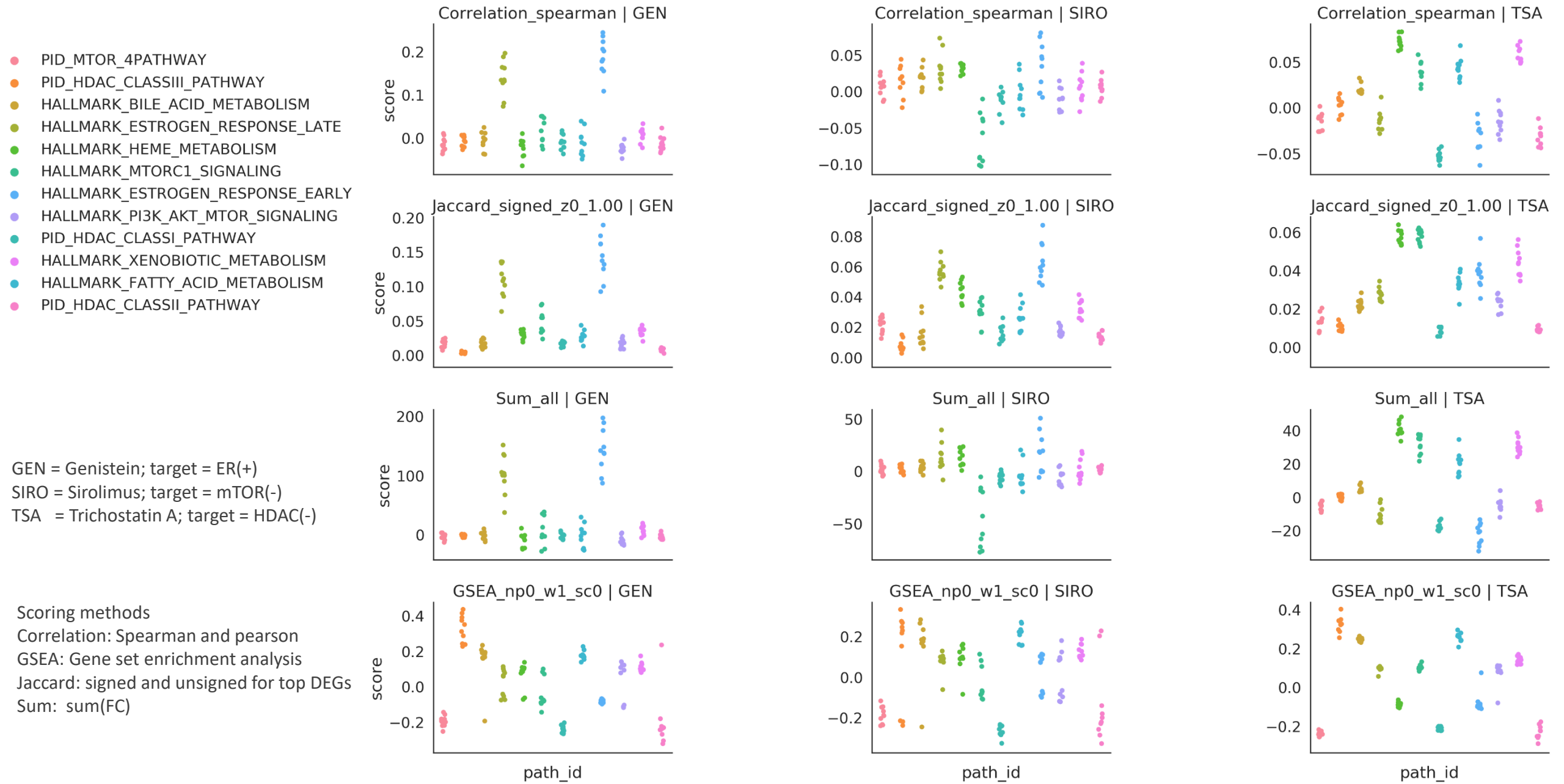
- GSA Approach
  - Input: DEGs as L2FC or Z-scores, pathway gene sets
  - Output: Score
- Use entire HTTr-PhI study
- Scoring methods
  - Correlation (Pearson, Spearman)
  - Gene set enrichment analysis GSEA
  - Jaccard (signed and unsigned)
  - XSum
  - XCos
- Pathways: MSigDB v6
  - Hallmark
  - Reactome
  - NCI Pathways
  - Others

Highlighted Pathways

- Based on three reference chemicals
  - GEN = Genistein; target = ER(+)
  - SIRO = Sirolimus; target = mTOR(-)
  - TSA = Trichostatin A; target = HDAC(-)

- MSigDB v6 Pathways :
  - 50 Hallmark pathways
  - 252 NCI pathways
  - 674 Reactome pathways
- Results shown for 12 pathways in table include
  - Metabolism pathways
  - ER-responsive genes
  - HDAC pathways
  - mTOR pathways

standard_name	Cat	contributor_org	n
HALLMARK_BILE_ACID_METABOLISM	H	Broad Institute	112
HALLMARK_ESTROGEN_RESPONSE_EARLY	H		200
HALLMARK_ESTROGEN_RESPONSE_LATE	H		200
HALLMARK_FATTY_ACID_METABOLISM	H		158
HALLMARK_HEME_METABOLISM	H		200
HALLMARK_MTORC1_SIGNALING	H		200
HALLMARK_PI3K_AKT_MTOR_SIGNALING	H		105
HALLMARK_XENOBIOTIC_METABOLISM	H		200
PID_HDAC_CLASSIII_PATHWAY	C2	National Cancer Institute and Nature Publishing Group	27
PID_HDAC_CLASSII_PATHWAY	C2		34
PID_HDAC_CLASSI_PATHWAY	C2		66
PID_MTOR_4PATHWAY	C2		69



GEN = Genistein; target = ER(+)
SIRO = Sirolimus; target = mTOR(-)
TSA = Trichostatin A; target = HDAC(-)

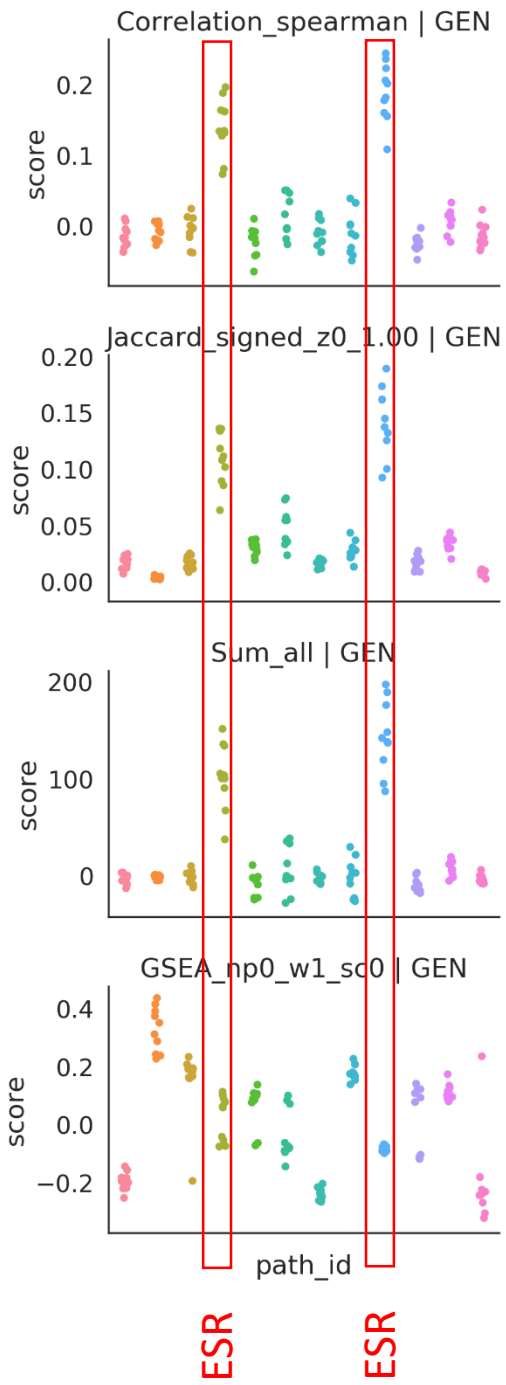
Scoring methods
Correlation: Spearman and pearson
GSEA: Gene set enrichment analysis
Jaccard: signed and unsigned for top DEGs
Sum: sum(FC)

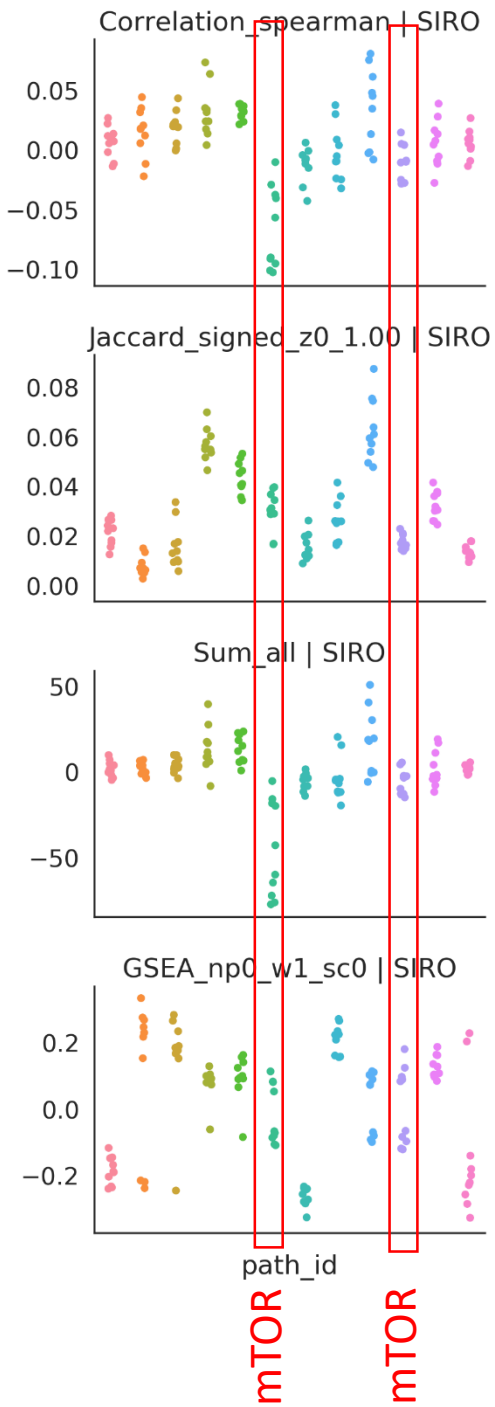
- PID_MTOR_4PATHWAY
- PID_HDAC_CLASSIII_PATHWAY
- HALLMARK_BILE_ACID_METABOLISM
- HALLMARK_ESTROGEN_RESPONSE_LATE
- HALLMARK_HEME_METABOLISM
- HALLMARK_MTORC1_SIGNALING
- HALLMARK_ESTROGEN_RESPONSE_EARLY
- HALLMARK_PI3K_AKT_MTOR_SIGNALING
- PID_HDAC_CLASSI_PATHWAY
- HALLMARK_XENOBIOTIC_METABOLISM
- HALLMARK_FATTY_ACID_METABOLISM
- PID_HDAC_CLASSII_PATHWAY

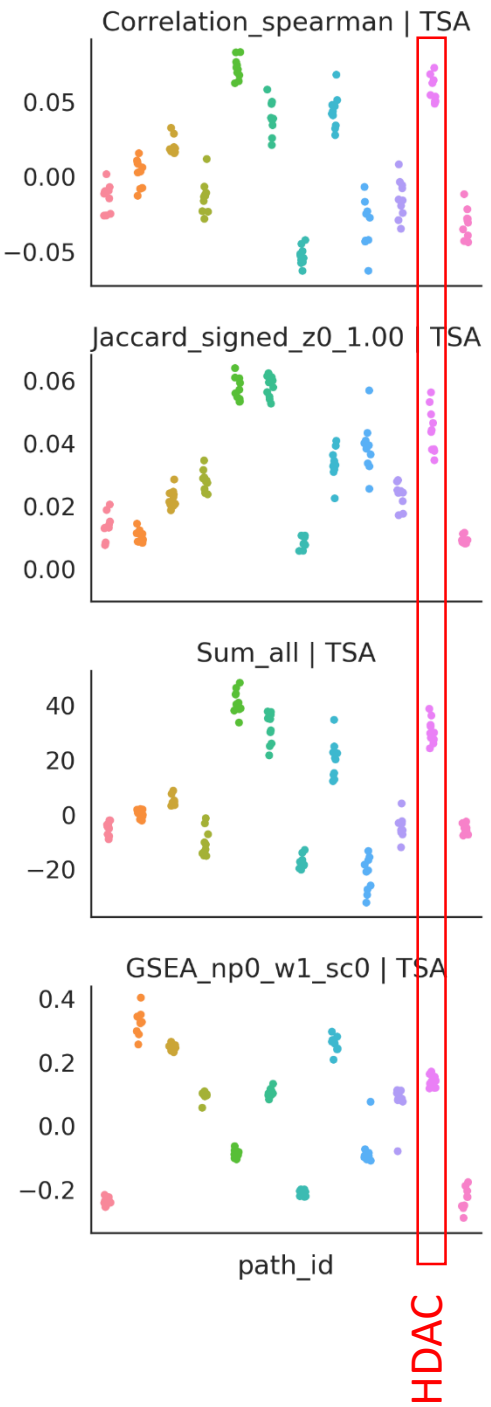
GEN = Genistein; target = ER(+)
SIRO = Sirolimus; target = mTOR(-)
TSA = Trichostatin A; target = HDAC(-)

Scoring methods
Correlation: Spearman and pearson
GSEA: Gene set enrichment analysis
Jaccard: signed and unsigned for top DEGs
Sum: sum(FC)

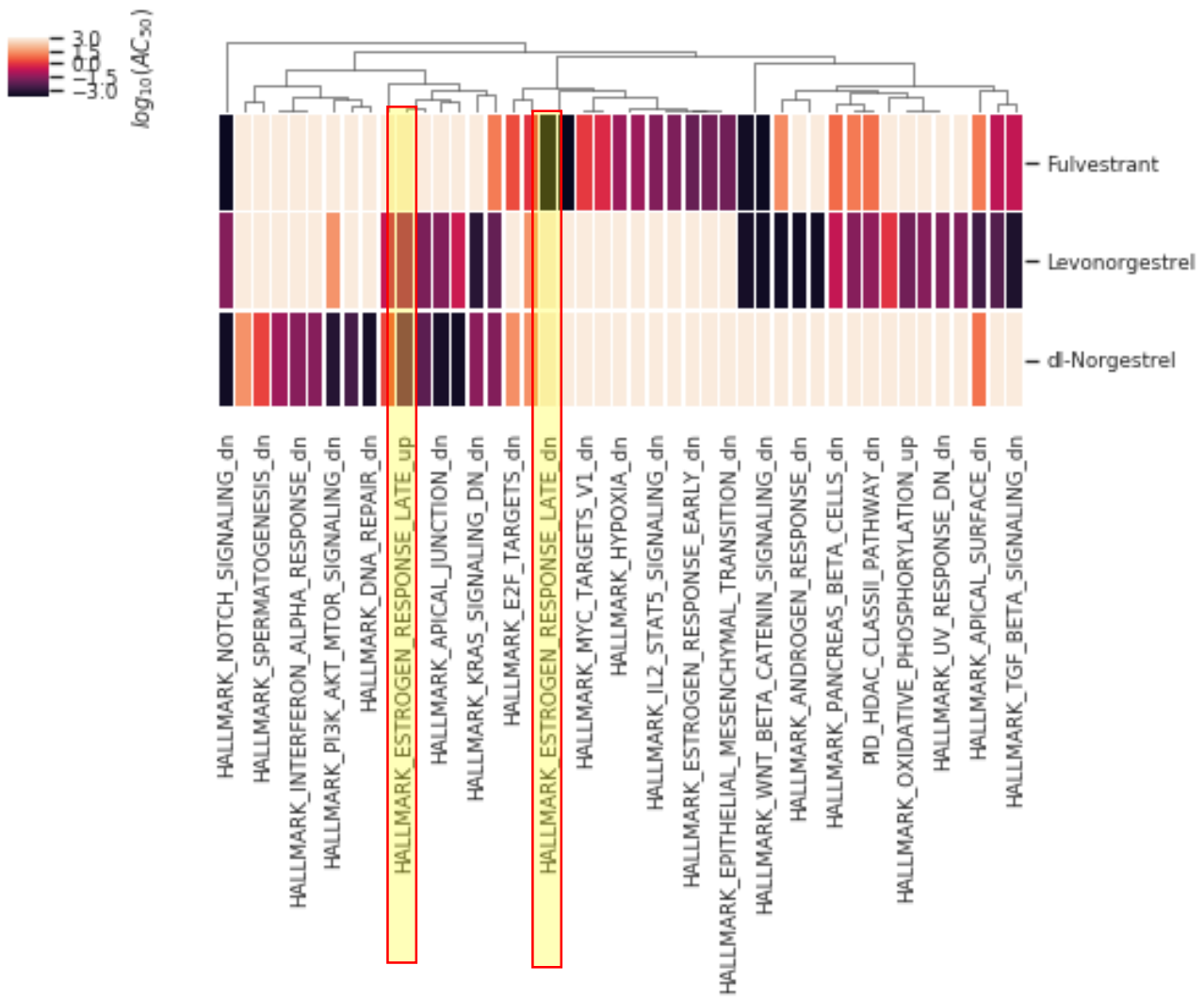
Most scoring schemes identify relevant target pathways

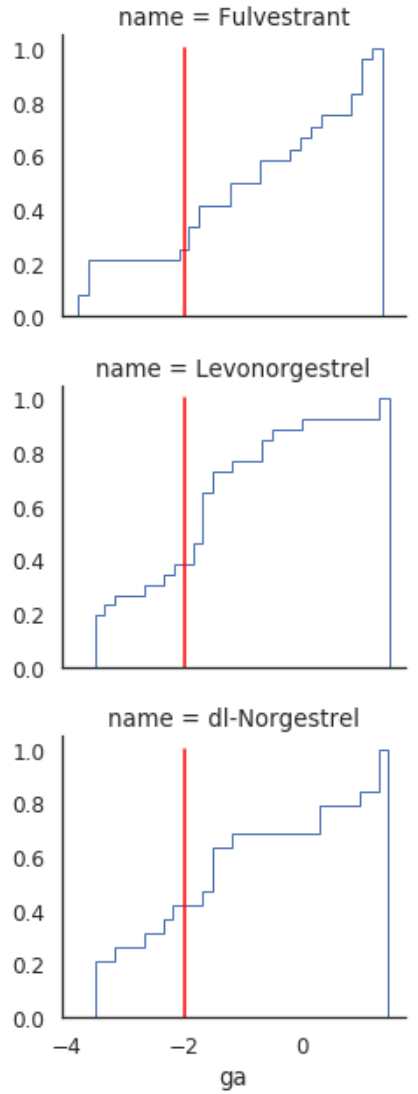






Analyzing chemicals without curated targets





Predicting Putative Chemical Targets

Connectivity mapping

Compare entire transcriptomic profile to reference database to find target using kNN

Pros:
- Need just one profile / target

Cons:
- Sensitive but not specific within platform
- Low cross-platform accuracy
- Requires chemical annot.

Pathway analysis

Compare entire transcriptomic profile to pathways (bags of genes) to find pathway 'hits' using different scoring schemes

Pros:
- More accurate (specificity)
- Assumes pathway→target
- Derive conc-response

Cons:
- Needs curated pathways
- No ideal pathway collection
- Multiple scoring schemes
- Pathways ≠ Targets
- Requires chemical annot.

*Tom Sheffield

Signatures/classifiers

Using profiles for target to create classifiers / signatures ("biomarkers"); search using entire profiles of test chemicals

Network analysis

Use transcriptomic profile alone with genetic-regulatory and signaling data to infer putative targets

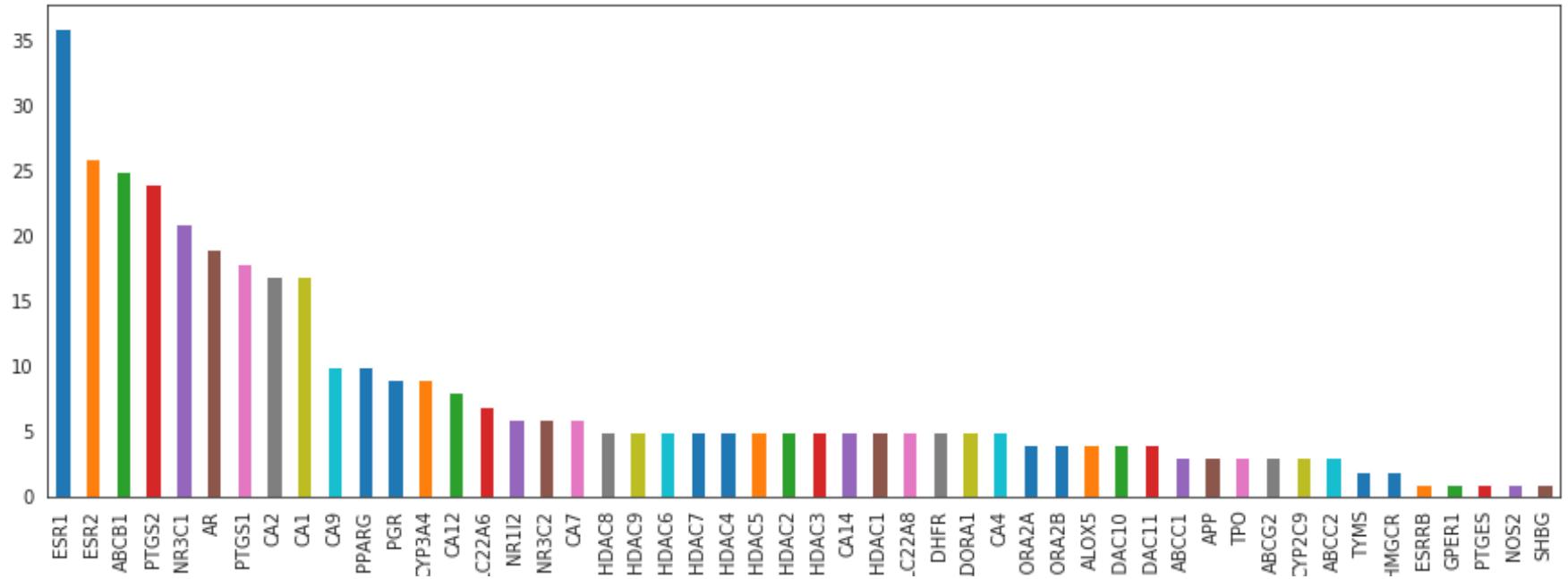
Developing signatures to predict targets

- Two main steps:
  - Signature identification: Use chemical-target annotations to find consistent DEGs patterns
  - Search and score test chemical profiles against signature using different algorithms
- Can be framed as a machine learning problem

- Initial approach
  - Signature identification
    - Use CMap v2
    - Find top n DEGs by abs(L2FC)
    - n ~ [100,1000,100]
  - Scoring methods
    - Jaccard (signed and unsigned)
    - Correlation (Pearson, Spearman)
    - GSEA
    - XSum
  - Analyse HTTr profiles from large screen

Distribution of targets based on RefChem v2

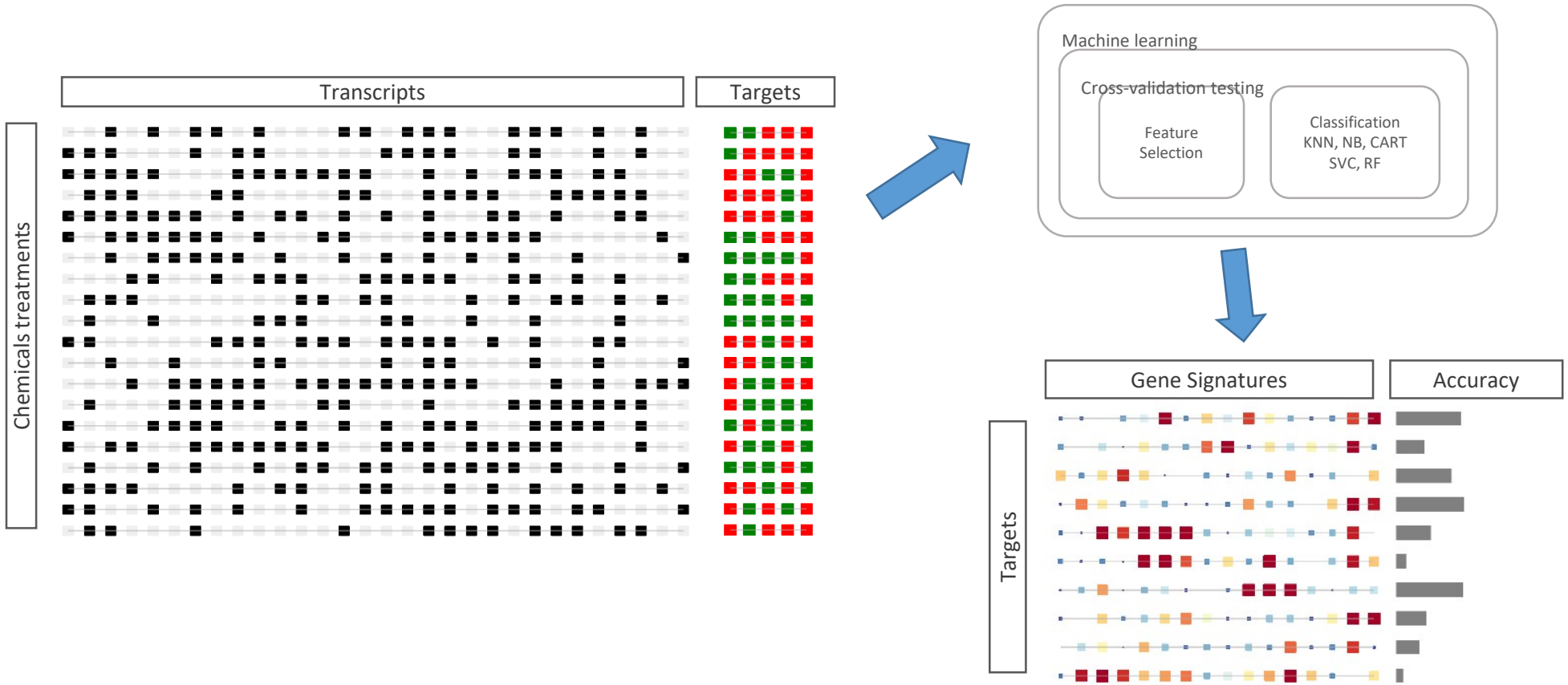
Select all targets that
Are annotated in
MCF7-Ph-I Screen



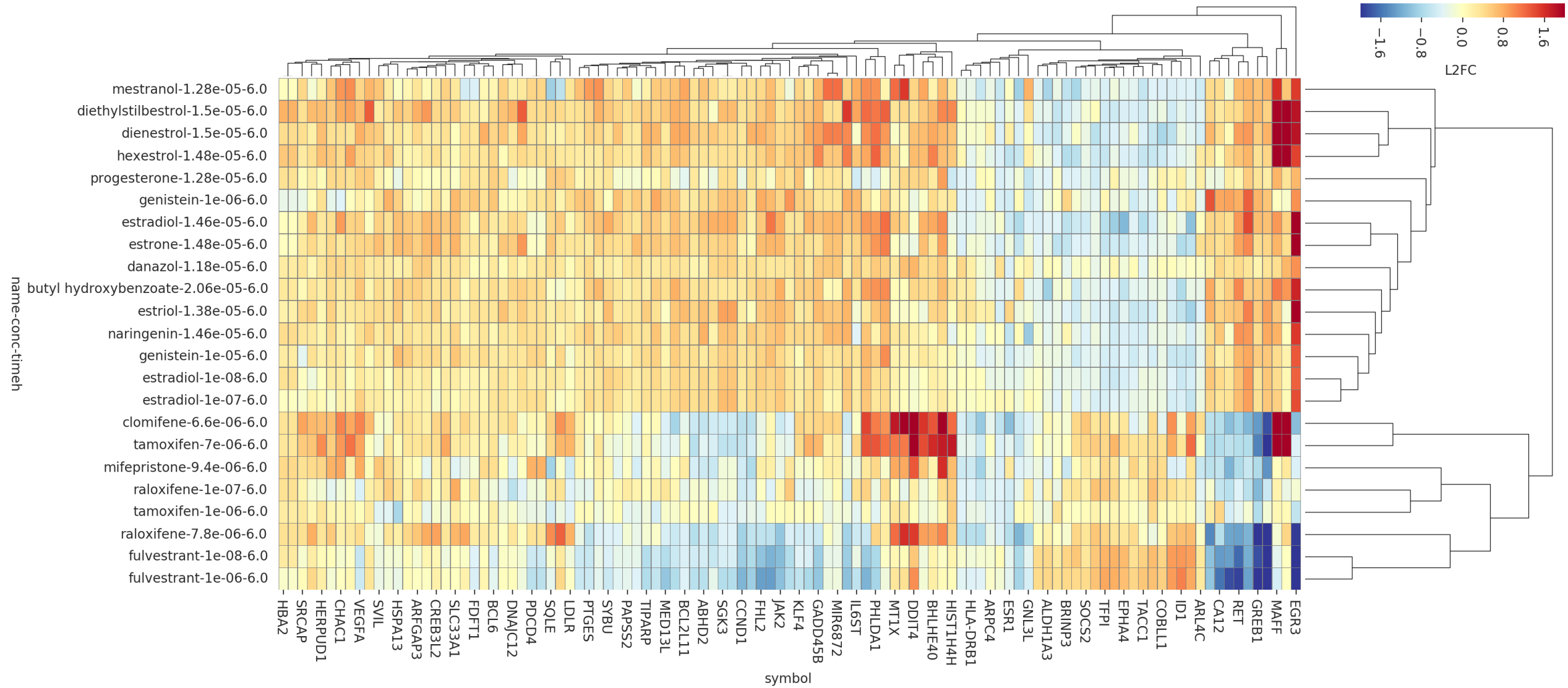
Nuclear receptors
Cytochrome P450s
ABC Transporters
Carbonic anhydrases

Histone deacetylases
Adenosine receptors

Signatures via Machine Learning



ER (any mode) Signature



Searching cMap v2 using ER Signature

dsstox_sid	name	target	cell	conc	jaccard	olap	timeh
Other	equilin	Other	MCF7	0.000015	0.176	110	6
Other	equilin	Other	MCF7	0.000015	0.166929	106	6
Other	lynestrenol	Other	MCF7	0.000014	0.163303	89	6
Other	prasterone	Other	MCF7	1.22E-05	0.162055	82	6
DTXSID3020465	diethylstilbestrol	ESRRG|ES	MCF7	0.000015	0.153846	82	6
DTXSID3020465	diethylstilbestrol	ESRRG|ES	MCF7	0.000015	0.144531	74	6
Other	trifluoperazine	Other	MCF7	0.00001	0.140777	87	6
Other	epitiostanol	Other	MCF7	0.000013	0.137876	87	6
Other	equilin	Other	MCF7	0.000015	0.1376	86	6
Other	etynodiol	Other	MCF7	1.04E-05	0.135314	82	6
DTXSID0020814	mestranol	ESR1	MCF7	1.28E-05	0.134052	87	6
DTXSID8022371	testosterone	AR|ESR2	MCF7	1.16E-05	0.131068	81	6
DTXSID6023656	thioridazine	DRD2|KCN	MCF7	0.00001	0.127303	76	6
DTXSID6023656	thioridazine	DRD2|KCN	MCF7	0.00001	0.126761	72	6
DTXSID4022369	fulvestrant	ESR2|ESR1	MCF7	1E-08	0.12623	77	6
Other	suloctidil	Other	MCF7	1.18E-05	0.124756	64	6
Other	prochlorperazine	Other	MCF7	0.00001	0.124214	79	6
DTXSID5022308	genistein	ESR2|ESR1	MCF7	0.00001	0.123664	81	6
Other	suloctidil	Other	MCF7	1.18E-05	0.120553	61	6
Other	mometasone	Other	MCF7	7.6E-06	0.119869	73	6
DTXSID2022880	danazol	AR|ESR1	MCF7	1.18E-05	0.119449	78	6
Other	trifluoperazine	Other	MCF7	0.00001	0.11936	82	6
Other	fluphenazine	Other	MCF7	0.00001	0.119163	74	6
DTXSID9020110	astemizole	KCNH2|HI	MCF7	8.8E-06	0.119005	67	6
Other	butyl hydroxyben	Other	MCF7	2.06E-05	0.118902	78	6
Other	ciclosporin	Other	MCF7	3.4E-06	0.118699	73	6
Other	ivermectin	Other	MCF7	4.6E-06	0.118098	77	6

- “Horse” estrogen
- Synthetic progesterone
- Synthetic progesterone
- Pro-androgen/estrogen
- Dopamine antagonist / Antipsychotic
  gynecomastia in males

Most of the top hits are ER-related

Searching MSigDB using ER Signature

category_code	jaccard	olap	organism	standard_name
H	0.127329193	41	human	HALLMARK_ESTROGEN_RESPONSE_EARLY
C2	0.094972067	34	human	DUTERTRE_ESTRADIOL_RESPONSE_6HR_UP
H	0.093373494	31	human	HALLMARK_ESTROGEN_RESPONSE_LATE
C2	0.087378641	18	human	MASSARWEH_RESPONSE_TO_ESTRADIOL
C2	0.083067093	26	human	NAGASHIMA_NRG1_SIGNALING_UP
C2	0.07826087	18	human	STEIN_ESR1_TARGETS
H	0.077151335	26	human	HALLMARK_TNFA_SIGNALING_VIA_NFKB
C2	0.076666667	23	human	PHONG_TNF_RESPONSE_VIA_P38_PARTIAL
C6	0.075301205	25	human	RAF_UP.V1_DN
C2	0.074712644	26	human	BHAT_ESR1_TARGETS_NOT_VIA_AKT1_UP
C2	0.07266436	21	human	PODAR_RESPONSE_TO_ADAPHOSTIN_UP
C2	0.069686411	20	human	BROCKE_APOPTOSIS_REVERSED_BY_IL6
C6	0.068452381	23	human	LTE2_UP.V1_DN
C2	0.065822785	26	human	MASSARWEH_TAMOXIFEN_RESISTANCE_DN
C2	0.065759637	29	human	CREIGHTON_ENDOCRINE_THERAPY_RESISTANCE_4
C2	0.064748201	27	human	BHAT_ESR1_TARGETS_VIA_AKT1_UP
C2	0.063829787	12	human	FRASOR_RESPONSE_TO_ESTRADIOL_UP
C2	0.0625	13	human	NAGASHIMA_EGF_SIGNALING_UP
C2	0.06031746	19	human	ELVIDGE_HYPOXIA_UP
C2	0.059259259	16	human	KAN_RESPONSE_TO_ARSENIC_TRIOXIDE

Searching ER Signature against MCF7-PhI

Most genistein (reference chemical) Profiles match the signature

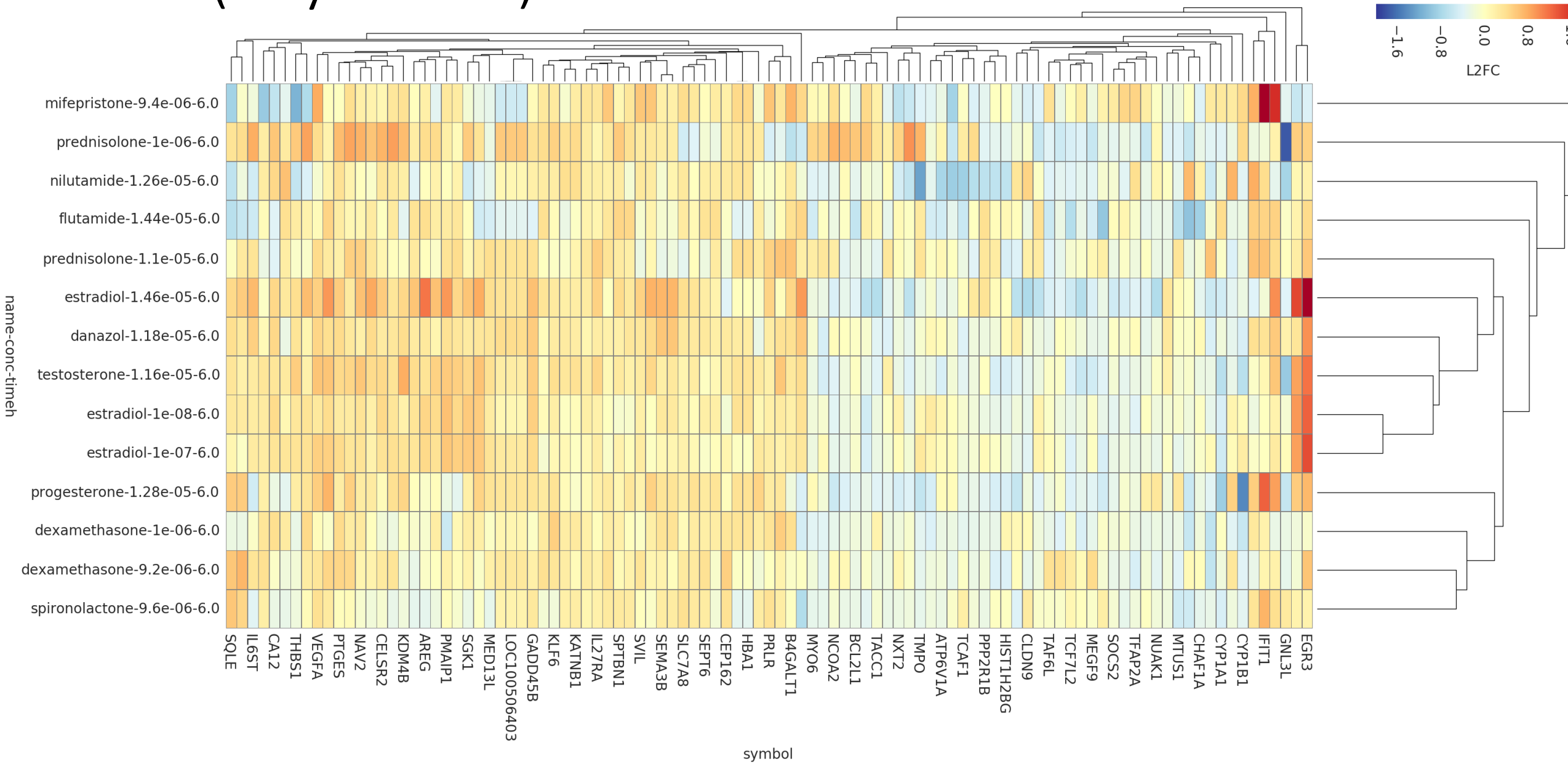
Many hits with chemicals that have Not been curated

target	chem	conc	jaccard	n_d	n_u	ola
ABCB1|ESR2|CYP3A4|ESR1	Tamoxifen	50	0.0582386	300	300	41
Other	Dehydroepiandrosterone	25.42373	0.056338	300	300	40
Other	HMR1171 trifluoroacetate (1:1)	35	0.0548523	300	300	39
Other	SSR146977	25	0.0546218	300	300	39
ESR2|ESR1	Fulvestrant	0.014967	0.0527778	300	300	38
KCNH2|HRH1	Astemizole	25.42373	0.0518934	300	300	37
Other	Estradiol cypionate	0.299461	0.0516039	300	300	37
ESR2|ESR1	Fulvestrant	0.04997	0.0516039	300	300	37
NR3C2|NR3C1|NR1I2|ABCB1	Dexamethasone	25.42373	0.0511757	300	300	37
Other	Niclosamide	12.71186	0.0504909	300	300	36
ESR2|ESR1	Fulvestrant	1.473477	0.0499307	300	300	36
Other	dl-Norgestrel	9.433962	0.0499307	300	300	36
ALOX15|ALOX5|ALOX12	Nordihydroguaiaretic acid	25.42373	0.0498615	300	300	36
Other	SSR 240612	11.44068	0.0497238	300	300	36
Other	Diethylstilbestrol dipropionate	25.42373	0.0495868	300	300	36
Other	SSR69071	25	0.0492264	300	300	35
Other	SSR 240612	45	0.0486787	300	300	35
ESR2|ESR1	Fulvestrant	12.71186	0.0486787	300	300	35
ESR2|ESR1	Fulvestrant	4.716981	0.0486111	300	300	35
Other	Disulfiram	25.42373	0.0476858	300	300	34
ESR1	Raloxifene hydrochloride	25.42373	0.0469613	300	300	34
Other	Cyclosporin A	9.433962	0.0469613	300	300	34
ESR2|ESR1	Fulvestrant	0.149731	0.0469613	300	300	34
Other	Emamectin benzoate	25.42373	0.0464135	300	300	33
Other	Fluorometholone	0.299461	0.0460251	300	300	33
Other	Fluorometholone	2.946955	0.0458971	300	300	33
AHR	Benzo(k)fluoranthene	9.433962	0.0458333	300	300	33
Other	Flurandrenolide	25.42373	0.0457064	300	300	33
Other	Hydramethylnon	25.42373	0.0453297	300	300	33
Other	Methyl Violet	25.42373	0.0450704	300	300	32
Other	Antimony trichloride	25.42373	0.0448808	300	300	32
Other	Rhodamine 6G	2.946955	0.0448179	300	300	32
Other	Sodium (2-pyridylthio)-N-oxide	25.42373	0.0448179	300	300	32
Other	Tributyltin chloride	2.946955	0.0446927	300	300	32
Other	1-Chloro-2,4-dinitrobenzene	25.42373	0.0445682	300	300	32
NR3C1	Methylprednisolone	0.09994	0.0443828	300	300	32
Other	Dibenz(a,h)anthracene	8.018868	0.0443213	300	300	32
Other	Cyclosporin A	25.42373	0.0443213	300	300	32
ESRRG|ESRRB|ESR2|ESR1	Diethylstilbestrol	0.029935	0.04426	300	300	32
ESR2|ESR1	Fulvestrant	50	0.0441989	300	300	32

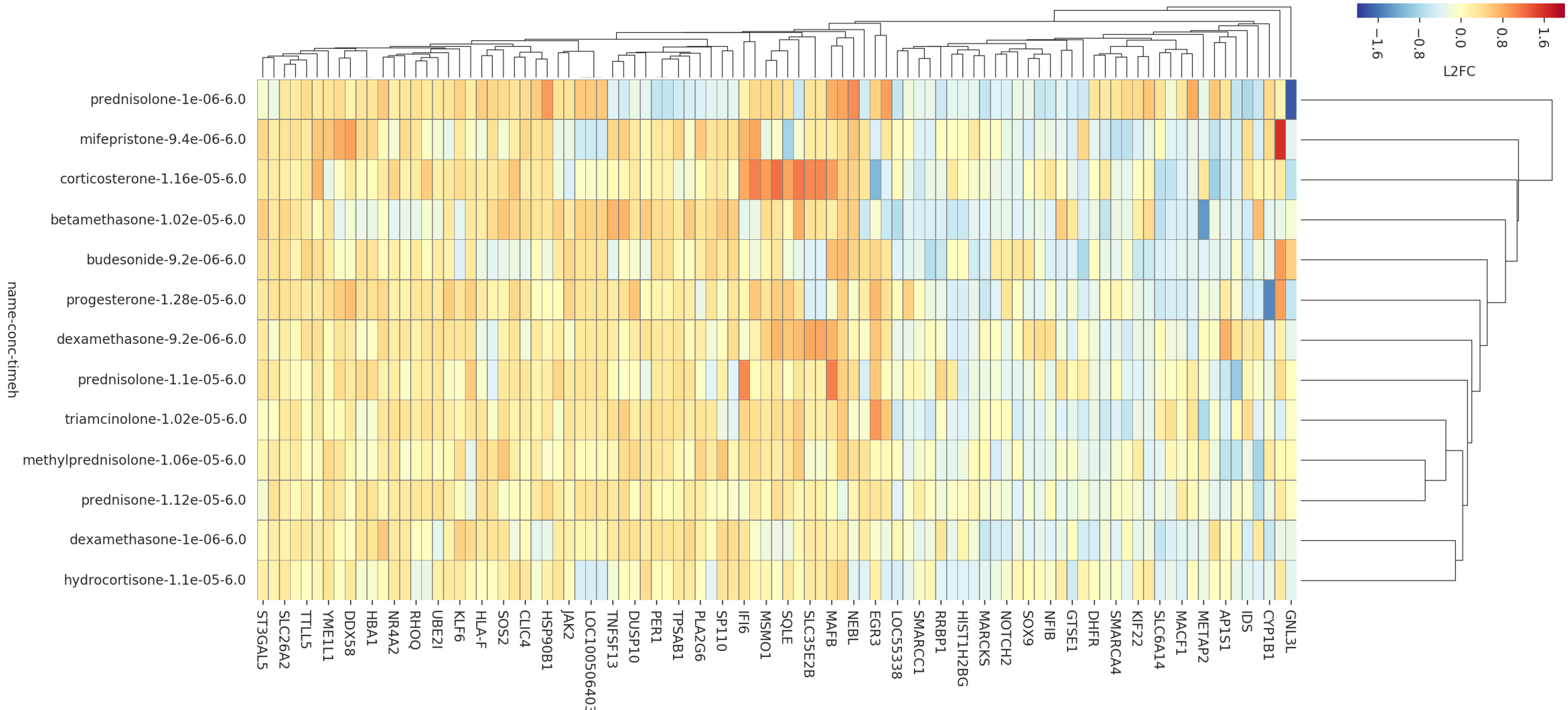
Searching ER Signature against plate level data

trt_id	target	chem_name	conc	jaccard	n_dn	n_reads	n_up	olap
lauryl-gallate-mnc10-TC00503946_D22-025.42uM	Other	Lauryl gallate	25.4237288	0.029925	130	4211100	120	12
genistein-mnc10-TC00503946_O21-010.00uM	ESR2|ESR1	Genistein	10	0.026132	14	6328481	412	15
fluorometholone-mnc10-TC00503946_L23-000.03uM	Other	Fluorometholone	0.0299347	0.024752	2	7507254	249	10
fluorometholone-mnc10-TC00503946_C07-009.43uM	Other	Fluorometholone	9.4339623	0.02449	16	5146073	72	6
thimerosal-mnc10-TC00503946_N18-025.42uM	Other	Thimerosal	25.4237288	0.024313	744	4376114	1031	46
fluorometholone-mnc10-TC00503946_G20-000.99uM	Other	Fluorometholone	0.9940358	0.022581	10	5792117	144	7
lauryl-gallate-mnc10-TC00503946_L04-009.43uM	Other	Lauryl gallate	9.4339623	0.022222	29	5159184	130	7
tyrphostin-mnc10-TC00503946_I02-009.43uM	EGFR	Tyrphostin	9.4339623	0.02079	324	2671704	4	10
trichostatin-a-mnc10-TC00503946_E08-001.00uM	HDAC1|HDAC3|HDAC2|	Trichostatin A	1	0.019567	439	4053104	388	19
arotinoid-acid-mnc10-TC00503946_N02-002.95uM	RARA|RARB|RARG	Arotinoid acid	2.9469548	0.019169	6	6674192	150	6
1-phenoxy-2-propanol-mnc10-TC00503946_N24-009.4	Other	1-Phenoxy-2-propano	9.4339623	0.018707	52	6336136	384	11
dipyrone-monohydrate-mnc10-TC00503946_F09-000.	Other	Dipyrone monohydrat	0.09994	0.018315	94	4194783	21	5
thiourea-mnc10-TC00503946_F17-000.10uM	Other	Thiourea	0.09994	0.017897	277	2817184	15	8
naptalam-mnc10-TC00503946_K07-009.43uM	Other	Naptalam	9.4339623	0.017467	30	5260324	40	4
econazole-nitrate-mnc10-TC00503946_P11-002.95uM	Other	Econazole nitrate	2.9469548	0.016736	323	2585137	0	8
fluorometholone-mnc10-TC00503946_C23-000.30uM	Other	Fluorometholone	0.299461	0.016616	14	7639715	1169	22
d-glucitol-mnc10-TC00503946_E02-000.10uM	Other	D-Glucitol	0.09994	0.016598	74	3642760	8	4
2-ethylhexyl-diphenyl-phosphate-mnc10-TC0050394	Other	2-Ethylhexyl diphenyl	9.4339623	0.016447	454	2346415	1	10
2-propyl-1-heptanol-mnc10-TC00503946_P13-000.03u	Other	2-Propyl-1-heptanol	0.0299347	0.016425	885	2233580	4	17
atropine-sulfate-monohydrate-mnc10-TC00503946_H	Other	Atropine sulfate monc	0.299461	0.01626	206	2561668	6	6
cadmium-chloride-mnc10-TC00503946_P03-009.43uM	Other	Cadmium chloride	9.4339623	0.016129	6	6026597	272	7
betaine-mnc10-TC00503946_H06-000.99uM	Other	Betaine	0.9940358	0.016053	908	2584567	5	17

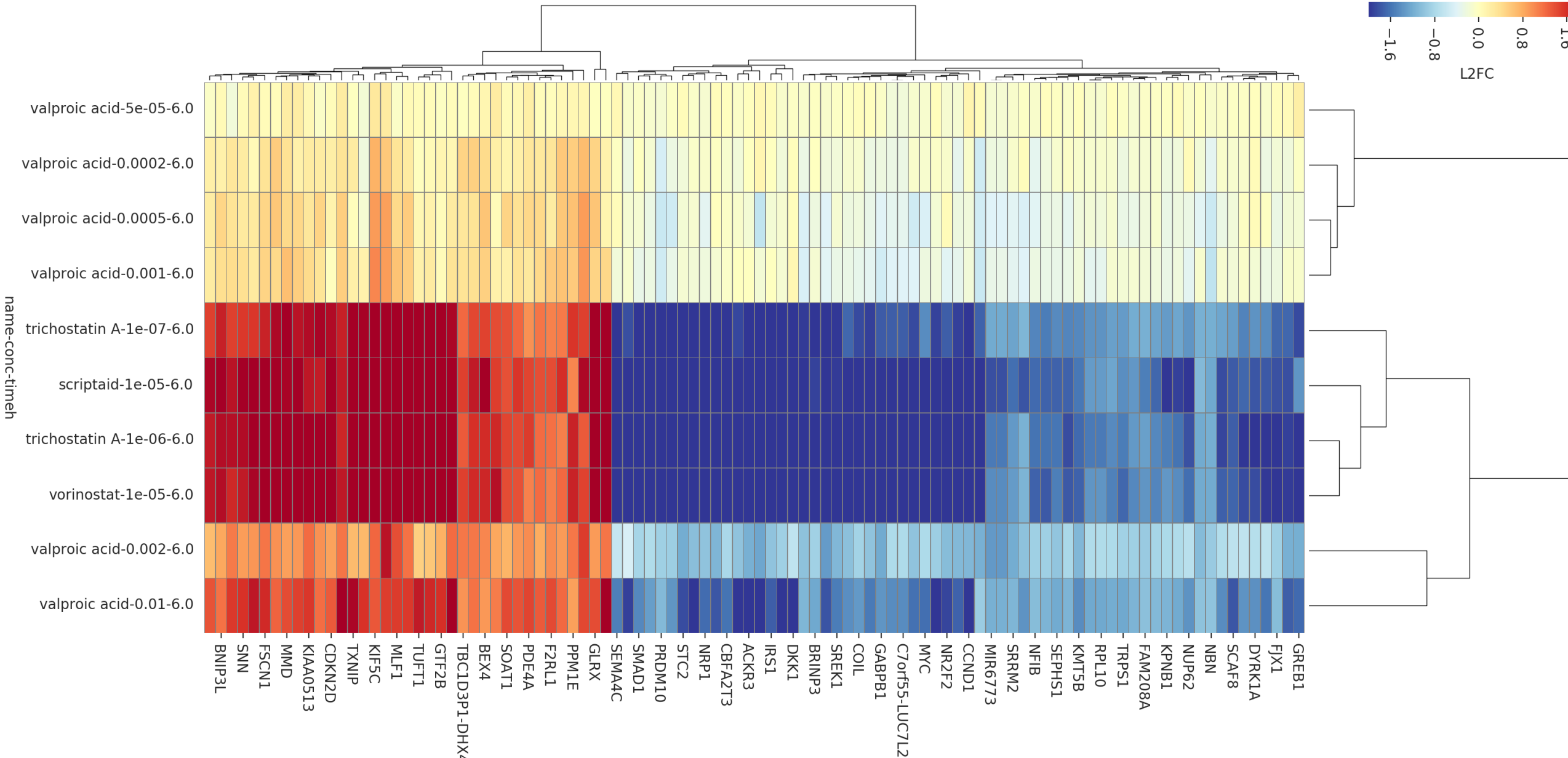
AR (any mode)


L2FC
−1.6
−0.8
0.0
0.8
name-conc-timeh
mifepristone-9.4e-06-6.0
prednisolone-1e-06-6.0
nilutamide-1.26e-05-6.0
flutamide-1.44e-05-6.0
prednisolone-1.1e-05-6.0
estradiol-1.46e-05-6.0
danazol-1.18e-05-6.0
testosterone-1.16e-05-6.0
estradiol-1e-08-6.0
estradiol-1e-07-6.0
progesterone-1.28e-05-6.0
dexamethasone-1e-06-6.0
dexamethasone-9.2e-06-6.0
spironolactone-9.6e-06-6.0
SQLE
IL6ST
CA12
THBS1
VEGFA
PTGES
NAV2
CELSR2
KDM4B
AREG
PMAIP1
SGK1
MED13L
LOC100506403
GADD45B
KLF6
KATNB1
IL27RA
SPTBN1
SVIL
SEMA3B
SLC7A8
SEPT6
CEP162
HBA1
PRLR
B4GALT1
MYO6
NCOA2
BCL2L1
TACC1
NXT2
TMPO
ATP6V1A
TCAF1
PPP2R1B
HIST1H2BG
CLDN9
TAF6L
TCF7L2
MEGF9
SOCS2
TFAP2A
NUAK1
MTUS1
CHAF1A
CYP1A1
CYP1B1
IFIT1
GNL3L
EGR3
symbol

NR3C1 (Glucocorticoid Recep.)



HDAC inhibitors



Performance of CMap v2 Affymetrx-derived signatures for predicting targets using BioSpyder HTTr data

Curation of hits necessary to determine specificity

	CMap v2 / Affymetrix	BioSpyder HTTr-Phase I			
Target	Signature size	PPV	Positives	Positive Chemicals found	Top 5 Prediction (Uncurated)
CYP2C9	131	1	1	Fluconazole	Emodin, Phenazopyridine hydrochloride, Lactofen, Hexachlorophene, 2-Amino-5-azotoluene
ESR1	257	1	11	o,p'-DDT, Genistein, 4-Nonylphenol, 4-Hydroxytamoxifen, Diethylstilbestrol, Raloxifene hydrochloride, Bisphenol A, 17beta-Estradiol, 5alpha-Dihydrotestosterone, Mifepristone, 4-(1,1,3,3-Tetramethylbutyl)phenol	dl-Norgestrel, SSR504734, Haloperidol, Cyclosporin A, Astemizole
HDAC1	124	1	2	Trichostatin A, Valproic acid	2-(Thiocyanomethylthio)benzothiazole, Azinphos-methyl, Sodium (2-pyridylthio)-N-oxide, 3,3'-Dichlorobenzidine dihydrochloride
DHFR	215	1	2	Pyrimethamine, Methotrexate	Adriamycin hydrochloride, PharmaGSID_48505, Etoposide, Resveratrol, Nisoldipine
NR1I2	139	1	2	17beta-Estradiol, Bisphenol A	dl-Norgestrel, Endosulfan, Isodrin, Genistein, 17alpha-Estradiol
PGR	115	1	1	Mifepristone	Flurandrenolide, Fluorometholone, Dexamethasone, Melengestrol acetate, Betamethasone
HMGCR	236	1	1	Lovastatin	Resveratrol, dl-Norgestrel, o,p'-DDT, Tamoxifen, Chlorhexidine
ABCC2	357	1	1	Methotrexate	4-Nitrosodiphenylamine, Resveratrol, Adriamycin hydrochloride, Nisoldipine, 8-Hydroxyquinoline sulfate
TYMS	329	1	1	Methotrexate	Etoposide, Resveratrol, 4-Nitrosodiphenylamine, Cytarabine hydrochloride, PharmaGSID_48505
ESR2	281	0.85714286	7	Genistein, Diethylstilbestrol, 4-Nonylphenol, Bisphenol A, 4-Hydroxytamoxifen, 17beta-Estradiol	dl-Norgestrel, 17alpha-Estradiol, Haloperidol, Cyclosporin A, Isodrin
AR	261	0.77777778	9	o,p'-DDT, 17beta-Estradiol, 5alpha-Dihydrotestosterone, Flutamide, Bisphenol A, Mifepristone, 17-Methyltestosterone	dl-Norgestrel, Melengestrol acetate, Dehydroepiandrosterone, 8-Hydroxyquinoline, Genistein
NR3C2	352	0.5	2	Mifepristone	Fluocinolone acetonide, Bexarotene, 1-Naphthol, Dexamethasone, dl-Norgestrel
ABCB1	117	0.5	2	Reserpine	Fabesetron hydrochloride, Abamectin, SAR115740, SSR69071, Chlorobenzilate
NR3C1	148	0.5	4	Triamcinolone, Mifepristone	Medroxyprogesterone acetate, Fluorometholone, Melengestrol acetate, Dexamethasone, Prednisolone
CA1	176	0.5	4	Phenol, Sodium nitrite	Triclopyr, Triclopyr butotyl, p-Bromodiphenyl ether, 2-Fluoroacetamide, 1-Ethyl-2-methylbenzene
CA2	341	0.5	4	Celecoxib, Phenol	PharmaGSID_48509, Acenaphthylene, CP-105696, Aloe-emodin, 2-Fluoroacetamide
PTGS1	307	0.25	4	Indomethacin	SSR69071, 17alpha-Estradiol, Chlordane, Cetylpyridinium bromide, Zoxamide

Predicting Putative Chemical Targets

Connectivity mapping

Compare entire transcriptomic profile to reference database to find target using kNN

Pros:
- Need just one profile / target

Cons:
- Sensitive but not specific within platform
- Low cross-platform accuracy
- Requires chemical annot.

Pathway analysis

Compare entire transcriptomic profile to pathways (bags of genes) to find pathway 'hits' using different scoring schemes

Pros:
- More accurate (specificity)
- Derive conc-response

Cons:
- Needs curated pathways
- No ideal pathway collection
- Multiple scoring schemes
- Pathways ≠ Targets
- Requires chemical annot.

Signatures/classifiers

Using profiles for target to create classifiers / signatures ("biomarkers"); search using entire profiles of test chemicals

Pros:
- More accurate
- Derive conc-response
- Target/mode-specific

Cons:
- Requires chemical annot.
- Require LARGE profile db

Network analysis

Use transcriptomic profile alone with genetic-regulatory and signaling data to infer putative targets

Next steps

- Systematically evaluate performance for signatures, pathways, connectivity-mapping
  - Further curate reference chemicals for known targets (E.g. nuclear receptors)
  - Use external chemical, gene knock-out and over-expression data from GEO for further validation
  - Evaluate utility of LINCS
- Potency estimates based on signatures, pathways, modules → BER
- Explore integrative (chemistry, network analysis and external data) approaches for predicting putative targets
- Continue making data available via (development) webservices

Summary

- Technology: Targeted RNA-Seq based HTTr is a promising platform for comprehensive and cost-effective evaluation of chemically induced disruption of biological processes/pathways
- Workflow: We have developed a standardized, scalable, and portable workflow to generate large-scale HTTr data for thousands of chemicals.
- Target Prediction: Stepwise evaluation of different approaches:
  - Connectivity mapping: Can be used for any profile, sensitive but not specific
  - Pathway/Signature analysis (GSA): Sensitive and more specific but does not identify all targets
  - Machine learning: Most accurate but not possible to evaluate all chemicals due to insufficient annotation
  - Use a joint approach to evaluate all 17,040 profiles
- Future: Network-based *de novo* target prediction