EvE Bio Methods - Data Release 1

Table of contents

Compound Library 2
- Library Creation 2
- Identity Confirmation 2
- Source Plate Preparation 2

Nuclear Receptor Assay 4
- Description 4
- Protocol 5
- Data Acquisition 8
  - Outlier Identification 9
  - Data Normalization 9

Quantitative Screening Process 10
- Assay Validation 10
- Screening Phase 10
- Compound Progression 11
- Profiling Phase 13
  - Outlier Identification 13
  - Curve Fitting 13
- Summary Result Determination 14

Compound Library

Library Creation

According to both the ZINC database and our own examination of the Drugs@FDA database, there are just over 1,600 FDA-approved small molecule drug active ingredients, excluding compounds that would be impractical or inappropriate to screen, such as inhalational anesthetics, topical drugs, simple salts and nutrients, antibodies, and radioisotopes. Commercial vendors of formatted compound libraries that include substantial coverage of FDA small-molecule drugs are numerous. Other vendors also make a wide range of pharmaceutical compounds, their major active metabolites, and tool compounds available either as individual off-the-shelf products or by custom synthesis.

Our preliminary observations indicated significant but not complete overlap among many of these commercial FDA-focused library products, resulting in EvE Bio assembling its own library from multiple vendors. The final selection and purchase of compounds that comprise the compound library was impacted by availability, cost, and licensing requirements. In total, the initial compound library screened comprised 1,397 compounds. Identifiers for compounds in the library can be found in the data browser compound page or in the downloadable metadata tables (which also include suppliers for each compound).

Identity Confirmation

Following the purchase of the compound library, a sample of each compound was sent for mass spectrometry (MS) analysis to confirm its identity. This analysis was performed by Momentum Biotechnologies, a provider of MS-based native detection technologies. Momentum employed RapidFire-MS to determine whether the correct molecular formula was observed for each sample submitted, using both positive and negative modes. All compounds were tested at a final concentration of 10 μM.

Based on experience shared by Momentum, we expected to identify 95% of the compounds in the compound library. Several reasons may explain why a compound was not identified by MS. For example, compounds with low molecular weight (<100 daltons) give poor signal, while compounds without nitrogen and metals or those with poor ionization may be difficult to identify. Following the initial analysis and follow-up, we were able to correctly identify 96% of the samples, providing high confidence in the integrity of our screening library.

Source Plate Preparation

The compound source plates for the screening phase were generated in-house. Two independent copies for each source plate were created. Compounds comprising the library were individually pipetted from tubes containing 10 mM stock in dimethylsulfoxide (DMSO) and were added to Echo-compatible 384-well source plates. Compounds were subsequently diluted to 1 mM and serially diluted 1:4 utilizing a Cybio FeliX (Analytik Jena) in 100% DMSO to generate 3 concentrations for each compound. DMSO was added to the column used to

define minimum response. The column for maximum response was left empty. The plate layout is shown in Figure 1.

The compound source plates for the profiling phase were generated in-house. Compounds were individually picked from the original diluted compound tubes (used in the screening phase) containing 10 mM stocks in DMSO and were added to 384-well low volume serial dilution source plates (SPT Labtech). Compounds were subsequently diluted to 1 mM and serially diluted 1:4 utilizing a Cybio FeliX (Analytik Jena) in 100% DMSO to generate 11 concentrations for each compound. DMSO was added to the column used to define minimum response. The column for maximum response was left empty. The plate layout is shown in Figure 2.

For screening plates, nanoliter copies in assay ready plates (ARPs) were made at Frontier Scientific, using an Echo Acoustic Liquid Handler (Beckman Coulter). For profiling plates, nanoliter copies in ARPs were made in-house using a Mosquito (SPT Labtech).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	Cmpd1	Cmpd1	Cmpd1	Cmpd17	Cmpd17	Cmpd17	min	Cmpd33	Cmpd33	Cmpd33	Cmpd49	Cmpd49	Cmpd49	max	Cmpd65	Cmpd65	Cmpd65	Cmpd81	Cmpd81	Cmpd81	Cmpd97	Cmpd97	Cmpd97	empty
B	Cmpd2	Cmpd2	Cmpd2	Cmpd18	Cmpd18	Cmpd18	min	Cmpd34	Cmpd34	Cmpd34	Cmpd50	Cmpd50	Cmpd50	max	Cmpd66	Cmpd66	Cmpd66	Cmpd82	Cmpd82	Cmpd82	Cmpd98	Cmpd98	Cmpd98	empty
C	Cmpd3	Cmpd3	Cmpd3	Cmpd19	Cmpd19	Cmpd19	min	Cmpd35	Cmpd35	Cmpd35	Cmpd51	Cmpd51	Cmpd51	max	Cmpd67	Cmpd67	Cmpd67	Cmpd83	Cmpd83	Cmpd83	Cmpd99	Cmpd99	Cmpd99	empty
D	Cmpd4	Cmpd4	Cmpd4	Cmpd20	Cmpd20	Cmpd20	min	Cmpd36	Cmpd36	Cmpd36	Cmpd52	Cmpd52	Cmpd52	max	Cmpd68	Cmpd68	Cmpd68	Cmpd84	Cmpd84	Cmpd84	Cmpd100	Cmpd100	Cmpd100	empty
E	Cmpd5	Cmpd5	Cmpd5	Cmpd21	Cmpd21	Cmpd21	min	Cmpd37	Cmpd37	Cmpd37	Cmpd53	Cmpd53	Cmpd53	max	Cmpd69	Cmpd69	Cmpd69	Cmpd85	Cmpd85	Cmpd85	Cmpd101	Cmpd101	Cmpd101	empty
F	Cmpd6	Cmpd6	Cmpd6	Cmpd22	Cmpd22	Cmpd22	min	Cmpd38	Cmpd38	Cmpd38	Cmpd54	Cmpd54	Cmpd54	max	Cmpd70	Cmpd70	Cmpd70	Cmpd86	Cmpd86	Cmpd86	Cmpd102	Cmpd102	Cmpd102	empty
G	Cmpd7	Cmpd7	Cmpd7	Cmpd23	Cmpd23	Cmpd23	min	Cmpd39	Cmpd39	Cmpd39	Cmpd55	Cmpd55	Cmpd55	max	Cmpd71	Cmpd71	Cmpd71	Cmpd87	Cmpd87	Cmpd87	Cmpd103	Cmpd103	Cmpd103	empty
H	Cmpd8	Cmpd8	Cmpd8	Cmpd24	Cmpd24	Cmpd24	min	Cmpd40	Cmpd40	Cmpd40	Cmpd56	Cmpd56	Cmpd56	max	Cmpd72	Cmpd72	Cmpd72	Cmpd88	Cmpd88	Cmpd88	Cmpd104	Cmpd104	Cmpd104	empty
I	Cmpd9	Cmpd9	Cmpd9	Cmpd25	Cmpd25	Cmpd25	min	Cmpd41	Cmpd41	Cmpd41	Cmpd57	Cmpd57	Cmpd57	max	Cmpd73	Cmpd73	Cmpd73	Cmpd89	Cmpd89	Cmpd89	Cmpd105	Cmpd105	Cmpd105	empty
J	Cmpd10	Cmpd10	Cmpd10	Cmpd26	Cmpd26	Cmpd26	min	Cmpd42	Cmpd42	Cmpd42	Cmpd58	Cmpd58	Cmpd58	max	Cmpd74	Cmpd74	Cmpd74	Cmpd90	Cmpd90	Cmpd90	Cmpd106	Cmpd106	Cmpd106	empty
K	Cmpd11	Cmpd11	Cmpd11	Cmpd27	Cmpd27	Cmpd27	min	Cmpd43	Cmpd43	Cmpd43	Cmpd59	Cmpd59	Cmpd59	max	Cmpd75	Cmpd75	Cmpd75	Cmpd91	Cmpd91	Cmpd91	Cmpd107	Cmpd107	Cmpd107	empty
L	Cmpd12	Cmpd12	Cmpd12	Cmpd28	Cmpd28	Cmpd28	min	Cmpd44	Cmpd44	Cmpd44	Cmpd60	Cmpd60	Cmpd60	max	Cmpd76	Cmpd76	Cmpd76	Cmpd92	Cmpd92	Cmpd92	Cmpd108	Cmpd108	Cmpd108	empty
M	Cmpd13	Cmpd13	Cmpd13	Cmpd29	Cmpd29	Cmpd29	min	Cmpd45	Cmpd45	Cmpd45	Cmpd61	Cmpd61	Cmpd61	max	Cmpd77	Cmpd77	Cmpd77	Cmpd93	Cmpd93	Cmpd93	Cmpd109	Cmpd109	Cmpd109	empty
N	Cmpd14	Cmpd14	Cmpd14	Cmpd30	Cmpd30	Cmpd30	min	Cmpd46	Cmpd46	Cmpd46	Cmpd62	Cmpd62	Cmpd62	max	Cmpd78	Cmpd78	Cmpd78	Cmpd94	Cmpd94	Cmpd94	Cmpd110	Cmpd110	Cmpd110	empty
O	Cmpd15	Cmpd15	Cmpd15	Cmpd31	Cmpd31	Cmpd31	min	Cmpd47	Cmpd47	Cmpd47	Cmpd63	Cmpd63	Cmpd63	max	Cmpd79	Cmpd79	Cmpd79	Cmpd95	Cmpd95	Cmpd95	Cmpd111	Cmpd111	Cmpd111	empty
P	Cmpd16	Cmpd16	Cmpd16	Cmpd32	Cmpd32	Cmpd32	min	Cmpd48	Cmpd48	Cmpd48	Cmpd64	Cmpd64	Cmpd64	max	Cmpd80	Cmpd80	Cmpd80	Cmpd96	Cmpd96	Cmpd96	Cmpd112	Cmpd112	Cmpd112	empty

Figure 1: Screening Plate Layout

Screening plates contain up to 112 compounds per source plate, with the highest concentration of each compound on the left and the lowest on the right. Minimum and maximum controls each occupy one column, where all wells within the column are replicates.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	Cmpd1	Cmpd1	Cmpd1	Cmpd1	Cmpd1	min	Cmpd1	Cmpd1	Cmpd1	Cmpd1	Cmpd1	Cmpd1	Cmpd17	Cmpd17	Cmpd17	Cmpd17	Cmpd17	max	Cmpd17	Cmpd17	Cmpd17	Cmpd17	Cmpd17	Cmpd17
B	Cmpd2	Cmpd2	Cmpd2	Cmpd2	Cmpd2	min	Cmpd2	Cmpd2	Cmpd2	Cmpd2	Cmpd2	Cmpd2	Cmpd18	Cmpd18	Cmpd18	Cmpd18	Cmpd18	max	Cmpd18	Cmpd18	Cmpd18	Cmpd18	Cmpd18	Cmpd18
C	Cmpd3	Cmpd3	Cmpd3	Cmpd3	Cmpd3	min	Cmpd3	Cmpd3	Cmpd3	Cmpd3	Cmpd3	Cmpd3	Cmpd19	Cmpd19	Cmpd19	Cmpd19	Cmpd19	max	Cmpd19	Cmpd19	Cmpd19	Cmpd19	Cmpd19	Cmpd19
D	Cmpd4	Cmpd4	Cmpd4	Cmpd4	Cmpd4	min	Cmpd4	Cmpd4	Cmpd4	Cmpd4	Cmpd4	Cmpd4	Cmpd20	Cmpd20	Cmpd20	Cmpd20	Cmpd20	max	Cmpd20	Cmpd20	Cmpd20	Cmpd20	Cmpd20	Cmpd20
E	Cmpd5	Cmpd5	Cmpd5	Cmpd5	Cmpd5	min	Cmpd5	Cmpd5	Cmpd5	Cmpd5	Cmpd5	Cmpd5	Cmpd21	Cmpd21	Cmpd21	Cmpd21	Cmpd21	max	Cmpd21	Cmpd21	Cmpd21	Cmpd21	Cmpd21	Cmpd21
F	Cmpd6	Cmpd6	Cmpd6	Cmpd6	Cmpd6	min	Cmpd6	Cmpd6	Cmpd6	Cmpd6	Cmpd6	Cmpd6	Cmpd22	Cmpd22	Cmpd22	Cmpd22	Cmpd22	max	Cmpd22	Cmpd22	Cmpd22	Cmpd22	Cmpd22	Cmpd22
G	Cmpd7	Cmpd7	Cmpd7	Cmpd7	Cmpd7	min	Cmpd7	Cmpd7	Cmpd7	Cmpd7	Cmpd7	Cmpd7	Cmpd23	Cmpd23	Cmpd23	Cmpd23	Cmpd23	max	Cmpd23	Cmpd23	Cmpd23	Cmpd23	Cmpd23	Cmpd23
H	Cmpd8	Cmpd8	Cmpd8	Cmpd8	Cmpd8	min	Cmpd8	Cmpd8	Cmpd8	Cmpd8	Cmpd8	Cmpd8	Cmpd24	Cmpd24	Cmpd24	Cmpd24	Cmpd24	max	Cmpd24	Cmpd24	Cmpd24	Cmpd24	Cmpd24	Cmpd24
I	Cmpd9	Cmpd9	Cmpd9	Cmpd9	Cmpd9	min	Cmpd9	Cmpd9	Cmpd9	Cmpd9	Cmpd9	Cmpd9	Cmpd25	Cmpd25	Cmpd25	Cmpd25	Cmpd25	max	Cmpd25	Cmpd25	Cmpd25	Cmpd25	Cmpd25	Cmpd25
J	Cmpd10	Cmpd10	Cmpd10	Cmpd10	Cmpd10	min	Cmpd10	Cmpd10	Cmpd10	Cmpd10	Cmpd10	Cmpd10	Cmpd26	Cmpd26	Cmpd26	Cmpd26	Cmpd26	max	Cmpd26	Cmpd26	Cmpd26	Cmpd26	Cmpd26	Cmpd26
K	Cmpd11	Cmpd11	Cmpd11	Cmpd11	Cmpd11	min	Cmpd11	Cmpd11	Cmpd11	Cmpd11	Cmpd11	Cmpd11	Cmpd27	Cmpd27	Cmpd27	Cmpd27	Cmpd27	max	Cmpd27	Cmpd27	Cmpd27	Cmpd27	Cmpd27	Cmpd27
L	Cmpd12	Cmpd12	Cmpd12	Cmpd12	Cmpd12	min	Cmpd12	Cmpd12	Cmpd12	Cmpd12	Cmpd12	Cmpd12	Cmpd28	Cmpd28	Cmpd28	Cmpd28	Cmpd28	max	Cmpd28	Cmpd28	Cmpd28	Cmpd28	Cmpd28	Cmpd28
M	Cmpd13	Cmpd13	Cmpd13	Cmpd13	Cmpd13	min	Cmpd13	Cmpd13	Cmpd13	Cmpd13	Cmpd13	Cmpd13	Cmpd29	Cmpd29	Cmpd29	Cmpd29	Cmpd29	max	Cmpd29	Cmpd29	Cmpd29	Cmpd29	Cmpd29	Cmpd29
N	Cmpd14	Cmpd14	Cmpd14	Cmpd14	Cmpd14	min	Cmpd14	Cmpd14	Cmpd14	Cmpd14	Cmpd14	Cmpd14	Cmpd30	Cmpd30	Cmpd30	Cmpd30	Cmpd30	max	Cmpd30	Cmpd30	Cmpd30	Cmpd30	Cmpd30	Cmpd30
O	Cmpd15	Cmpd15	Cmpd15	Cmpd15	Cmpd15	min	Cmpd15	Cmpd15	Cmpd15	Cmpd15	Cmpd15	Cmpd15	Cmpd31	Cmpd31	Cmpd31	Cmpd31	Cmpd31	max	Cmpd31	Cmpd31	Cmpd31	Cmpd31	Cmpd31	Cmpd31
P	Cmpd16	Cmpd16	Cmpd16	Cmpd16	Cmpd16	min	Cmpd16	Cmpd16	Cmpd16	Cmpd16	Cmpd16	Cmpd16	Cmpd32	Cmpd32	Cmpd32	Cmpd32	Cmpd32	max	Cmpd32	Cmpd32	Cmpd32	Cmpd32	Cmpd32	Cmpd32

Figure 2: Profiling Plate Layout

Profiling plates contain up to 32 compounds per source plate, with the highest concentration of each compound on the left and the lowest on the right. Minimum and maximum controls each occupy one column, where all wells within the column are replicates.

Nuclear Receptor Assay

Description

Biochemical co-activator recruitment assays for nuclear receptors (NR) are an established and robust method for the screening of potential ligands in high-throughput screening campaigns. Ligand binding to a NR of interest can mediate a conformational change within the protein that allows for recruitment of co-activator proteins to the AF2 domain. Peptide motifs within these co-activator proteins, such as the LXXLL motif, have been demonstrated to be the key amino acids sresponsible for the ligand-induced binding of these co-activator proteins to nuclear receptors. These types of assays can be configured to detect agonist, antagonist, and inverse agonist molecules.

For the co-activator recruitment assay, an epitope tagged NR ligand binding domain (LBD) and biotinylated co-factor peptide are mixed with appropriate donor and acceptor molecules and added to ligands to generate a time-resolved fluorescence resonance energy transfer (TR-FRET) signal. This signal is modulated by a ligand-induced conformational change, thus implying ligand binding, as shown in Figure 3.

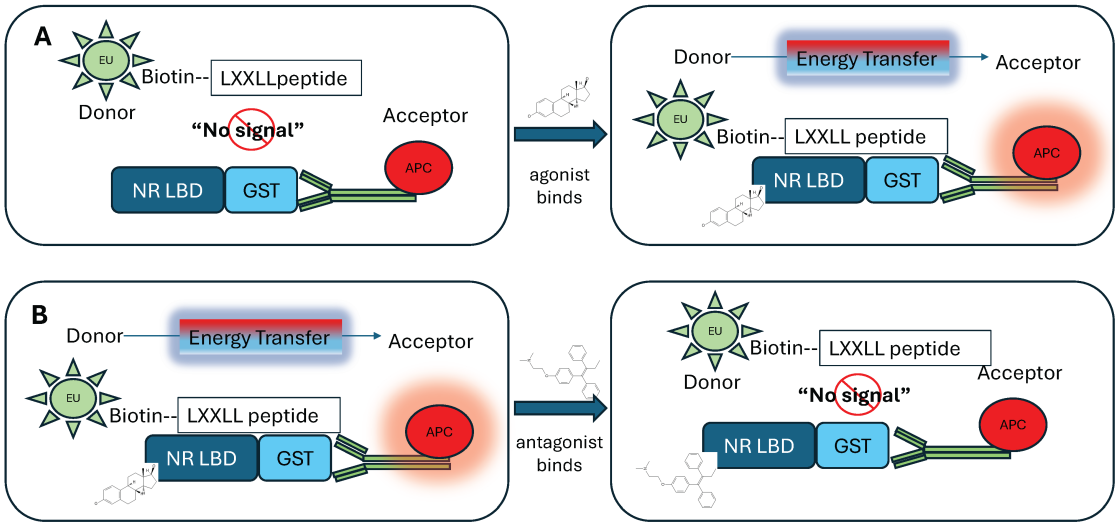

Figure 3: Nuclear Receptor TR-FRET Assay, Agonist and Antagonist Modes
A) Agonist Mode: In absence of agonist, co-activator does not associate with NR resulting in minimal fluorescence. Agonist binding induces a conformational change that shifts the AF2 domain to allow the LXXLL peptide to bind to the receptor, thus bringing the donor and acceptor molecules in proximity resulting in increased fluorescence. *B) Antagonist Mode:* Antagonist displaces agonist resulting in dissociation of complex and reduction of fluorescence.

Protocol

A 2x solution of purified recombinant epitope-tagged NR LBD protein and a separate solution of 2x concentration of biotinylated co-factor peptide, europium-labeled streptavidin (donor), and allophycocyanin (APC)-labeled anti-tag antibody (acceptor) were prepared in assay buffer (50 mM HEPES, pH 7.5, 50 mM KCl, 0.15 mM CHAPS, 0.1 mg/mL bovine serum albumin, 1 mM dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine (TCEP)). A 5 μL volume of the NR LBD solution was added to black small volume 384-well plates (Greiner Bio-One) that contained test compounds diluted in DMSO and was equilibrated with compound for 10 minutes prior to the addition of 5 μL of the co-factor peptide/donor/acceptor solution. The antagonist configuration of the assay is the same as described above but includes the control agonist spiked at a concentration 2x the EC80 into the 2x co-factor peptide/donor/acceptor solution. The generated signal was collected for 665 nm to 620 nm. The ratio of the 665 nm to the 620 nm was used to normalize the data using a DMSO control and the standard agonist control.

Reagent information and concentrations for each assay are included in Table 1, Table 2, Table 3, Table 4, and Table 5. Target identifiers can be found in the data browser or the downloadable metadata tables, as can additional assay data.

Table 1: Nuclear Receptor TR-FRET Assay Reagents

Tagged Nuclear Receptor Ligand-Binding Domain			Biotin Peptide		APC-labeled Antibody		Europium-labeled Streptavidin	Reducing Agent
Target	**Tag**	**[nM]**	**Peptide**	**[nM]**	**Antibody**	**[nM]**	**[nM]**	**[1 mM]**
CAR	GST	3	SRC-1	16	Anti-GST	7.5	2	TCEP
ERa	GST	10	SRC-1	16	Anti-GST	5	2	DTT
ERb	GST	5	SRC-1	16	Anti-GST	5	2	DTT
FXR	GST	5	SRC-1	16	Anti-GST	7.5	2	DTT
LXRa	GST	5	SRC-1	16	Anti-GST	7.5	2	TCEP
LXRb	GST	5	SRC-1	16	Anti-GST	7.5	2	DTT
PPARa	GST	5	SRC-1	16	Anti-GST	7.5	2	DTT
PPARd	GST	5	SRC-1	16	Anti-GST	7.5	2	DTT
PPARg	GST	5	SRC-1	16	Anti-GST	7.5	2	DTT
PR	GST	30	SRC-1	16	Anti-GST	30/40[a]	2	TCEP
PXR	GST	3	SRC-1	16	Anti-GST	7.5	2	TCEP
RARa	GST	5	SRC-1	16	Anti-GST	15	2	TCEP
RARb	GST	5	SRC-1	16	Anti-GST	15	2	TCEP
RARg	GST	5	SRC1	16	Anti-GST	15	2	TCEP
RXRa	GST	2.5	SRC-1	16	Anti-GST	7.5	2	TCEP
RXRb	GST	2.5	SRC-1	16	Anti-GST	10	1.5	TCEP
RXRg	GST	3	SRC-1	16	Anti-GST	10	1.5	TCEP
TRa	GST	1	SRC-1	16	Anti-GST	7.5	2	TCEP
TRb	GST	1	SRC-1	16	Anti-GST	7.5	2	TCEP

[a] PR: 30 nM anti-GST antibody was used in the agonist format; 40 nM anti-GST antibody was used in the antagonist format.

APC, allophycocyanin; **DTT,** dithiothreitol; **GST,** glutathione S-transferase;

TCEP, tris(2-carboxyethyl)phosphine; **SRC-1,** steroid receptor co-activator 1; **[nM]**, final concentration.

Table 2: Nuclear Receptor TR-FRET Assay Agonist and Antagonist Normalizing Controls and Spike

Target	Agonist (AG) Configuration		Antagonist (ANT) Configuration			
	AG Control	[µM]	AG spike (EC80)	[µM]	ANT control	[µM]
CAR	CITCO[a]	10	non-spiked	N/A	CINPA	100
ERa	Estradiol	1	Estradiol	0.0025	Raloxifene	10
ERb	Estradiol	1	Estradiol	0.005	Raloxifene	10
FXR	6-Ethyl CDCA	10	6-Ethyl CDCA	0.5	DY268	10
LXRa	T0901317	10	GW3965	1	GSK2033	10
LXRb	GW3965	10	GW3965	1	GSK2033	10
PPARa	GW7647	10	GW7647	0.02	GW6471	10
PPARd	GW501516	10	GW501516	0.01	GSK3787	10
PPARg	GW1929	10	GW1929	0.05	GW9662	10
PR	Progesterone	10	Progesterone	0.005	RU486	10
PXR	T0901317	10	T0901317	0.1	non-spiked[b]	N/A
RARa	Palovarotene	5	Palovarotene	0.12	BMS493	1
RARb	Palovarotene	5	Palovarotene	0.04	BMS493	1
RARg	Palovarotene	5	Palovarotene	0.12	BMS493	1
RXRa	LG100268	1	LG100268	0.025	UVI3003	10
RXRb	LG100268	1	LG100268	0.025	UVI3003	10
RXRg	LG100268	1	LG100268	0.03	UVI3003	10
TRa	L-thyroxine	1	L-thyroxine	0.005	TR antagonist 1	50
TRb	L-thyroxine	1	L-thyroxine	0.02	TR antagonist 1	50

[a] Due to high basal activity, the CAR antagonist S07662 was spiked into the agonist format at 10 µM to increase the signal window, and no spike was used in the antagonist format.

[b] Due to the lack of available PXR antagonists, a non-spiked control was used for normalization.

Table 3: Nuclear Receptor TR-FRET Assay Proteins

Name	Source	Catalog #
CAR	Thermo Fisher Scientific	PV4838
ERa	Thermo Fisher Scientific	A15677
ERb	Thermo Fisher Scientific	A15665
FXR	Thermo Fisher Scientific	PV4835
LXRa	Thermo Fisher Scientific	PV4657
LXRb	Thermo Fisher Scientific	PV4660
PPARa	Thermo Fisher Scientific	PV4692
PPARd	Thermo Fisher Scientific	PV4694
PPARg	Thermo Fisher Scientific	PV4546
PR	Thermo Fisher Scientific	A15672
PXR	Thermo Fisher Scientific	PV4841
RARa	AB Vector	N1B1
RARb	AB Vector	N1B2
RARg	AB Vector	N1B3
RXRa	Thermo Fisher Scientific	PV4799
RXRb	AB Vector	N2B2
RXRg	AB Vector	N2B3
TRa	Thermo Fisher Scientific	PV4762
TRb	Thermo Fisher Scientific	PV4764

Table 4: Nuclear Receptor TR-FRET Assay Peptides

Name	Source	Sequence	Catalog #
SRC-1	AnaSpec	Biotin-CPSSHSSLTERHKILHRLLQEGSPS-OH	AS-62152

Table 5: Nuclear Receptor TR-FRET Assay Donor and Acceptor Molecules

Name	Source	Catalog #
Anti-GST Antibody, APC-tagged	Columbia BioSciences	D3-1310
Europium-labeled streptavidin	Columbia BioSciences	D17-221

APC, allophycocyanin; **GST,** glutathione-S-transferase

Data Acquisition

Following the final equilibration of the assay plates as described in the protocol, compound activity was quantified using a PHERAstar plate reader (BMG Labtech) equipped with a homogeneous time-resolved fluorescence (HTRF®) TR-FRET Module. Assay plates were read under the following instrument settings:

- HTRF® optical module
- Excitation: 337 nm
- Integration time= 60 μs

- Emission: 620 nm/665 nm
- Integration time = 400 μs
- Number of flashes = 200

Outlier Identification

Outlying points in the minimum and maximum control values were removed from subsequent transformation steps if they were outside the median +/- 10 median absolute deviations (MAD) range for the corresponding set of control values for all plates in the same replicate. The MAD is a robust version of standard deviation.

Data Normalization

Raw data were converted to a percent (%) response scale ranging from 0 (minimal activity) to 100 (maximal activity). This normalization is based on the minimum and maximum control data on the same plate.

$$Activity = 100*(Sample - MeanMin)/(MeanMax - MeanMin)$$

Quantitative Screening Process

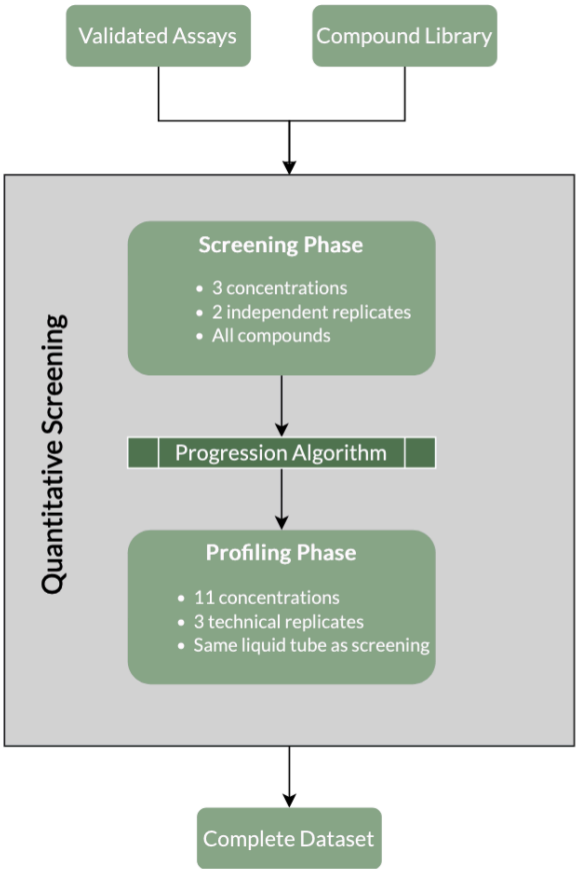

Figure 4: Quantitative Screening Process Overview

Assay Validation

For each new target tested, an assay was developed and optimized for Z prime and target pharmacology. The assay validation included confirmation of expected results using known pharmacological standard molecules with reported activity at the target. This validation was repeated with the standard plate upon transfer of the assay into the quantitative screening process before assays advanced to primary screens.

Screening Phase

The primary screen consisted of ARPs generated from two independently created sets of source plates containing all compounds from the compound library. Compounds were screened at 3 concentrations with a maximum concentration of 10 μM. The maximum effect compound was added using a Mosquito to each ARP at the time of the assay. Each primary screen included a standard plate for each copy of the library screened.

For each screening assay, a comparison was made between the values obtained in that screen and historical values for the molecules in the standard plate. If potency values were within 0.5 log units, the assay passed a quality assurance (QA) check at the assay level. This was combined with calculating a Z prime for each plate which was used to pass/fail each plate.

Compound Progression

When the screening phase was complete, a rules-based algorithm (developed based on pilot studies) was applied to the screening results to determine which compounds should progress to the profiling phase.

As an input to the progression process, “bundles” of targets were identified, roughly based on target sub-families. The same set of compounds progressed for all targets in a bundle to better characterize selectivity and specificity. Bundles are specified in the downloadable metadata tables.

Compounds were advanced for 1 of 3 reasons:

1. It was likely that a pXC50 (-log XC50) would be quantifiable in the profiling phase
2. The compound was likely quantifiable for other targets in the same bundle
3. Allowing the compound to progress was logistically convenient

The algorithm assigns one of seven labels to each combination (target/assay) based on the screening phase data. These form the basis of progression decisions, along with bundling and convenience criteria, as shown in Figure 5.

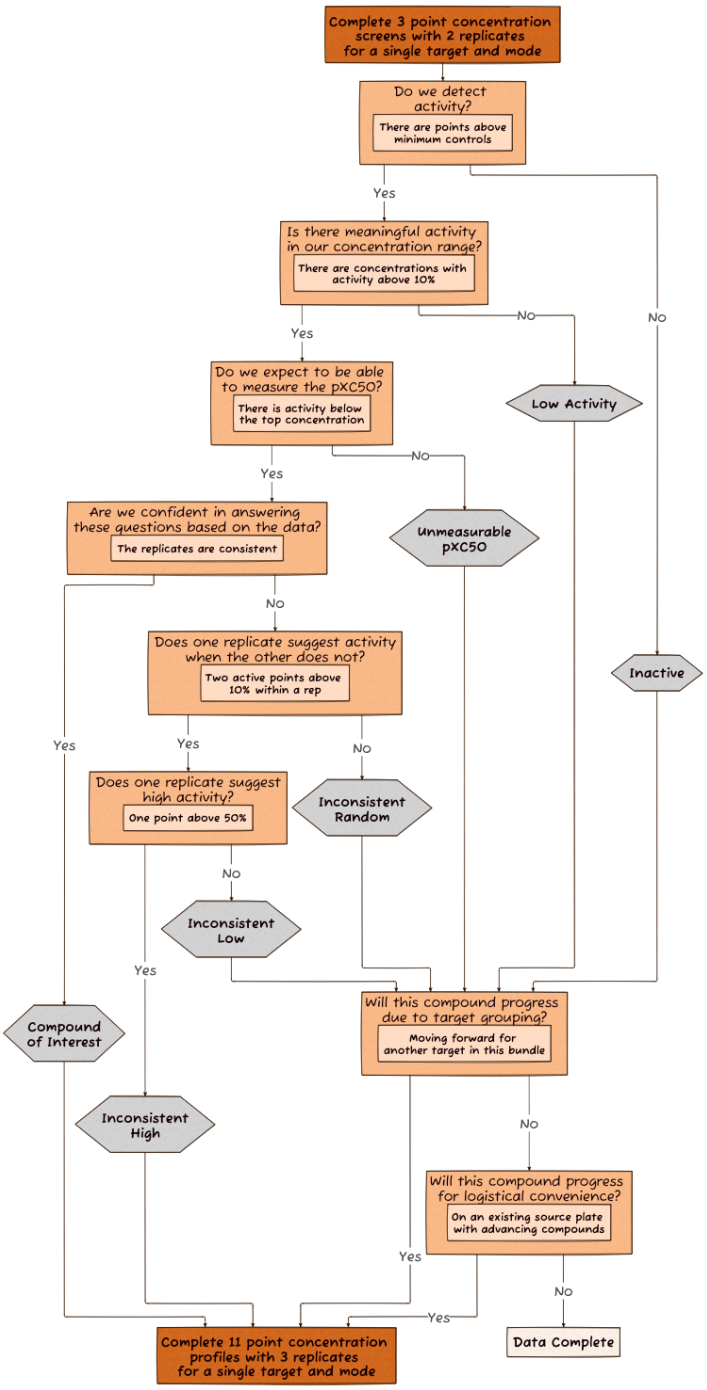

Figure 5: Compound Progression Algorithm Flow Chart

1) *Active:* Any point is above the threshold defined as 3 standard deviations of the minimum controls, 2) *Meaningful:* Both replicates above 10% activity for 1+ concentrations, 3) *Measurable:* Both replicates at the middle or lowest concentration are active and meaningful, 4) *Consistent:* Both replicates either do or do not have 2+ points that are active and meaningful

Profiling Phase

Compounds progressed from the screening phase were further profiled in an 11-point concentration response curve. Compounds were screened with a maximum concentration of 10 μM. The maximum effect compound was added using a Mosquito to each ARP at the time of the assay. Each assay utilized a standard plate containing the same compounds used for the screening phase standard plate.

For each profiling assay, a comparison was made between the values obtained in the assay and historical values for the molecules in the standard plate. If potency values were within 0.5 log units, the assay passed a QA check at the assay level. This was combined with calculating a Z prime for each plate which was used to pass/fail each plate. Curves were profiled in 3 ARPs generated from a single source plate (i.e., technical replicates).

Outlier Identification

Profiling data was subject to expert review and potential outlier marking. Outliers were marked based on the following principles:

- Err on the side of non-removal of data points.
- Points should not be removed to "improve" fit parameters or increase agreement with historical data.
- Removal of outliers may be appropriate where it will enable a reasonable fit when a fit is not possible based on all data points.
- Bell-shaped concentration curves may justify removal of the highest concentration(s).
- Bi-phasic responses may justify removal of the highest concentration(s) when the first phase is known to be biologically relevant.
- Removal of the highest concentration(s) may be justified in instances where maximum responses are unfeasible but do not trigger QA failure.
- If there was significant variation with the observed data points which create a reasonable doubt as to whether a point should be removed and/or removal of different points results in calculated pXC50's that differ by more than 0.5 log units, reject the curve and retest.
- For the 11-point concentration response curves, a maximum of 4 points may be removed, although this should be the exception.

Curve Fitting

A 4-parameter logistic regression model was fit to the normalized profiling (11-point) data. The fit was performed using the R package drc (version 3.0-1), with the drm function and LL.4 model specification. The curve fit provided point estimates and 95% confidence intervals for the asymptotic minimum, asymptotic maximum, slope, and inflection point (XC50). Additionally, the pXC50 was calculated (pXC50 = -log(XC50)).

The following validation criteria were applied to determine whether and how the curve fit results should be reported:

- If the mean response at all concentrations was less than 3 standard deviations of the minimum control on the same plates, the compound was considered inactive, and no fit parameters were reported.
- The fit was considered invalid if any of the following criteria were met, and no fit parameters were reported:
  - The fitting function could not estimate the XC50
  - The fitting function could not estimate the upper confidence interval for the XC50
  - The R-squared value was less than 0.7
  - The slope estimate was not between 0 and 5
  - Data was not available for at least 8 of the 11 concentrations
- If the estimated pXC50 was less than 5, the pXC50 was reported as <5, and no other fit parameters were reported.

Primary reporting of 11-point curve results was based on a single curve fit to data from all three profiling replicates. Results from separate curve fits to each replicate were also provided in the downloadable data.

Summary Result Determination

The quantitative screening process was designed to generate a result for every compound against every target assay (the "Result" column in the summary table). The overall summary result was derived from both the screening and profiling phases, and intended to represent the best single description based on all the data collected. When both profiling phase and screening phase data were available, the summary result was based on the profiling phase. Otherwise, it was based on the screening phase.

Each combination (compound / assay) had a result in 1 of 4 categories:

- **Inactive:** All datapoints at all concentrations were below the inactive threshold for the assay in that phase, which is 3 standard deviations of the minimum controls. Note that this threshold can vary substantially by assay.
  - **Inactive (3pt)** is for combinations not progressing to the profiling stage, in which case 6 datapoints (two replicates at three concentrations) were evaluated.
  - **Inactive (3pt)** is for combinations progressing to the profiling stage, in which case 33 datapoints (three replicates at 11 concentrations) were evaluated. These combinations may or may not have been inactive in the screening phase.
- **Likely Inactive:** At least one datapoint caused the combination not to meet the criteria for Inactive. However, the pattern of points above the inactive threshold is one that was found in pilot studies to be associated with inactivity upon profiling follow-up. For example, a single point may be above the threshold.
  - **Likely Inactive (3pt)** is the only subcategory, as the "Likely Inactive" label is only applied based on data from the screening phase.

- **Active - Unquantified:** The data indicates activity, but a quantified pXC50 was not obtained. There are many reasons to explain why pXC50 was not obtained, depending on whether the combination reached the profiling phase.
  - **Active - Unquantified (3pt)** can only apply to combinations that did not progress to the profiling phase. These combinations either have activity for both points at one concentration that are above the inactive threshold but below 10% activity or the highest concentration has both points above 10% activity but no other concentrations do. In the latter case, the combination does not progress to the profiling phase because there is no expectation that additional lower concentrations would result in the ability to quantify the pXC50.
  - **Active - Unquantified (11pt)** applies in one of two cases for combinations that progressed to the profiling phase. In the first case, the 4-parameter logistic fit did not meet validation criteria (see curve fitting section above). If the fit was valid, either the estimated pXC50 was below 5, or the upper confidence limit for the pXC50 was undeterminable.
  - **Active - Quantified:** The combination progressed to the profiling phase, where a 4-parameter logistic curve was fit that met curve fitting validation criteria, and a pXC50 value can be reported.
  - **Active - Quantified (11pt)** is the only subcategory, as combinations must have reached the profiling phase to have curves fit.