Supplementary materials

Nuisance compounds, PAINS filters and dark chemical matter in the GSK HTS collection

Subhas J. Chakravorty *et al.*

Table 1S – Description of most noisy marketed drugs (IFI greater than 10 %) in the GSK collection

Drug name	IFI	IFIwk	Mode of action*	Indication*
Daunorubicin	28.1	26.9	Topoisomerase inhibitor (cytotoxic)	Cancer
Bithionol	23.1	23.5	Redox interference	Worm infections
Doxorubicin	21.2	20.3	Topoisomerase inhibitor (cytotoxic)	Cancer
Deslanoside	19.3	18.2	Na-K-ATPase membrane pump blocker	Congestive heart failure
Metildigoxin	14.0	13.3	Na-K-ATPase membrane pump blocker	Congestive heart failure
Sertraline	14.0	13.0	Selective serotonin uptake inhibitor	Depression
Digitoxin	13.9	12.7	Na-K-ATPase membrane pump blocker	Congestive heart failure
Sorafenib	13.7	14.8	Protein kinase inhibitor (CRAF, BRAF, VEGFR…)	Cancer
Masoprocol	13.4	12.3	5-lipoxygenase inhibitor	Actinic keratose
Sertindole	13.0	11.9	Antagonist of dopamine D2, serotonin 5-HT2A and 5-HT2C, and alpha1-adrenoreceptors	Schizophrenia
Apomorphine	12.8	12.6	Agonist of dopamine D2, D3, D4 and D5 receptors	Parkinson's disease
Niclosamide	12.6	11.1	Uncoupler of electron transport chain to ATP synthase	Worm infections
Levodopa	11.6	11.7	Prodrug of dopamine	Parkinson's disease
Clomipramine	10.4	9.8	Serotonin uptake inhibitor	Depression
Broxyquinolone	10.3	9.0	Unknown	Protozoan infections

*From PubChem (https://www.ncbi.nlm.nih.gov/pccompound) accessed on Aug 29, 2017

Figure 1S: Flow-chart describing the data driven approach used to define new nuisance filters

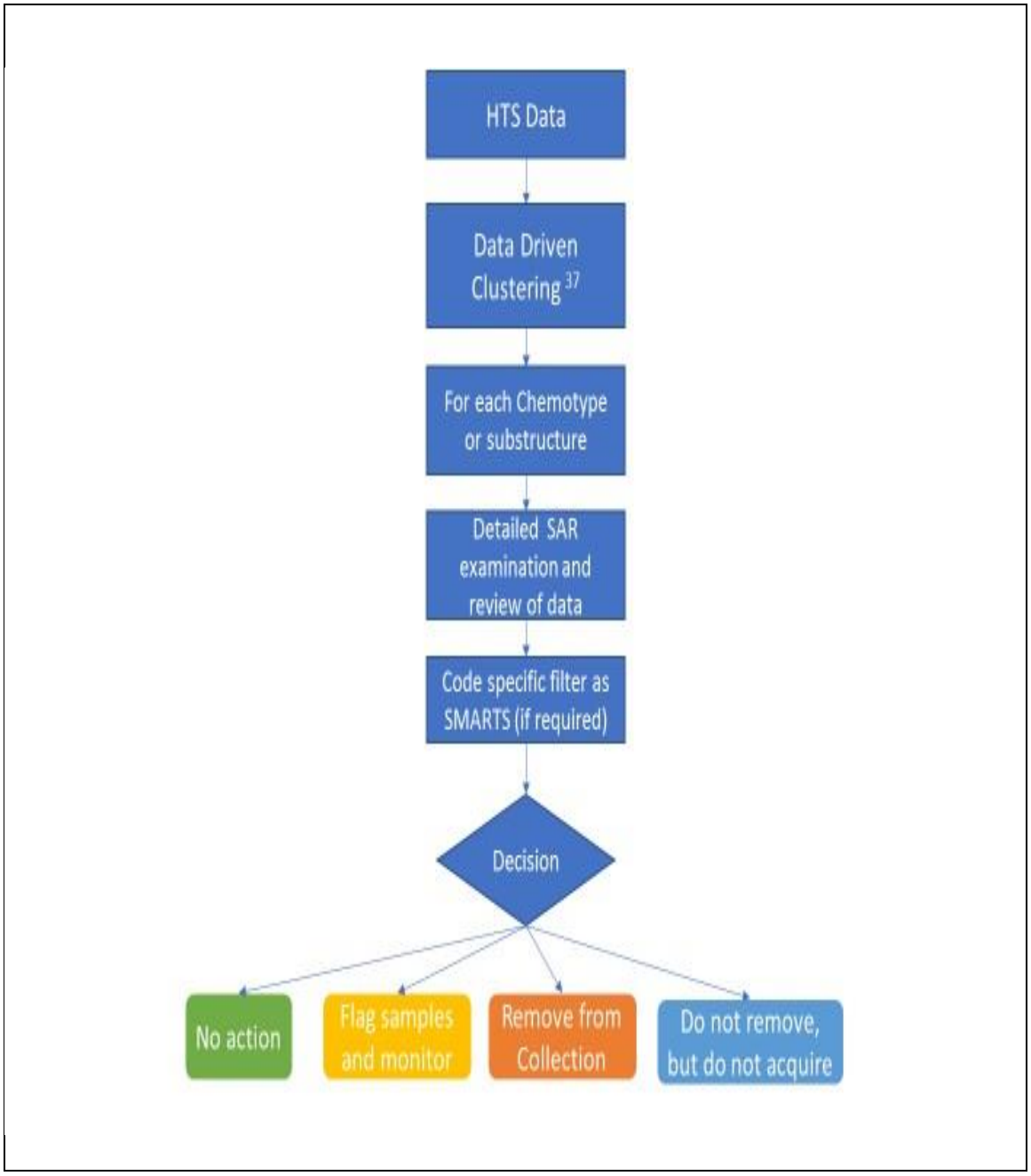


Table 4S: Filters showing target class bias

Source	Filter	SMARTS	Target Class Bias
NIH	saponin	O1CCCCC1OC2CCC3CCCCC3C2	TRANSPORTERS
GSK_CTC	Dihydroisoquinoline	[$([#6][C;!H0]=[C;!H0]C1NCCc2ccccc12),$([#6][CH]=[CH]C1=NCCc2ccccc12)]	TRANSPORTERS
GSK_CTC	O-Sulphonamide_benzamide	[$(O=C(Nc1ccccc1)c1ccccc1NS(=O)(=O)c1ccccc1)	TRANSPORTERS
NIH;BMS	aryl_phosphonate	P(=O)-[O;!R]-a	KINASES
Lilly	biotin	[SD2r5]1-C-[C;R2]2-N-C(=O)-N-C2-C1	KINASES
GSK_CTC	Pyrimidinesulfone	[#6]S(=O)(=O)c1ncc([Cl,Br])c(n1)C([#7])=O	OTHER_TARGETS
PainsB	cyano_imine_B	N#CC(C#N)=N[NH]c1ccccc1	7TM
GSK_CTC	Thiadiazolesulfinyl	O=Cc1nnsc1S(=O)[#6]	7TM & TRANSPORTERS

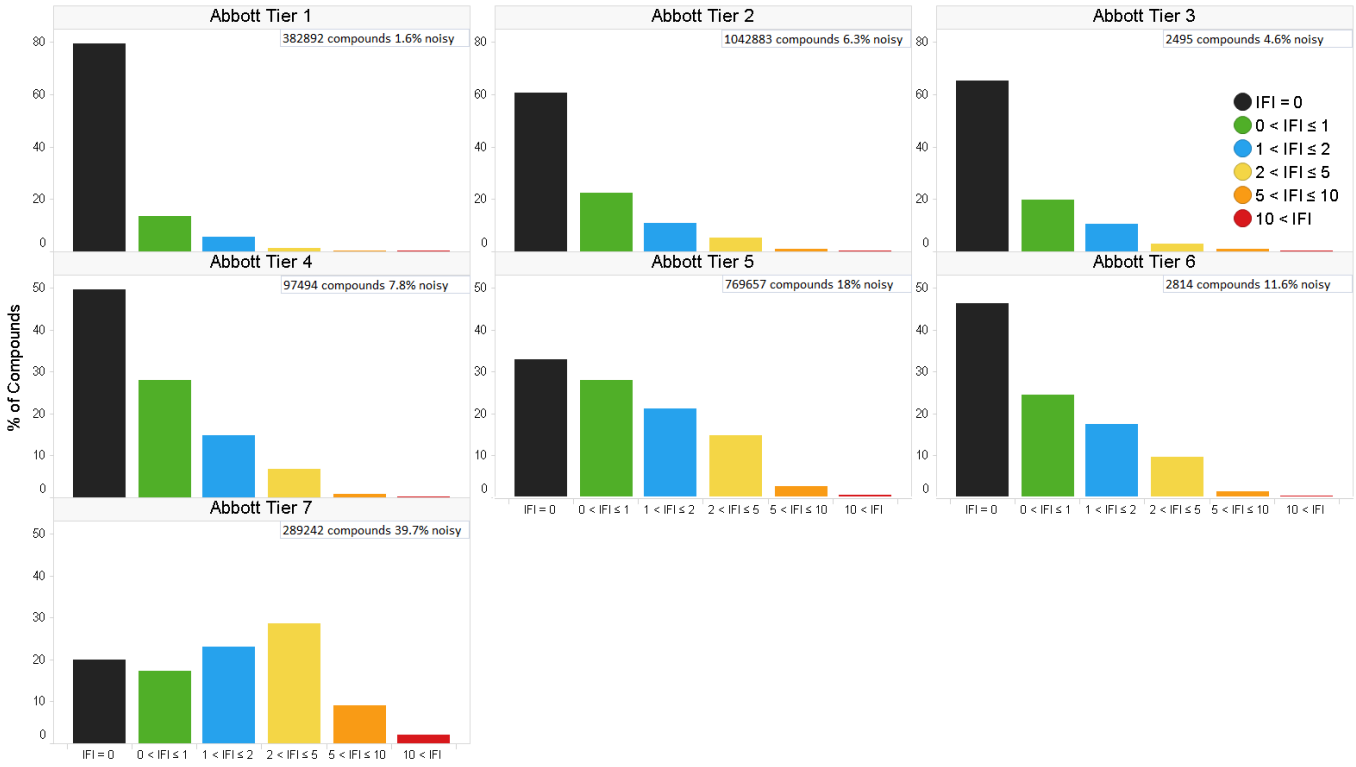
Table 5S: Filters showing assay technology bias

Source	Filter	SMARTS	Assay Technology Bias
ChEMBL-fixes	Filter82_pyridinium-fixed	[c,n]1[c,n][c,n][c,n][c,n][c,n][n+](C)1	ABSORB
NIH;BMS	3halo_pyridazine_2EWG	[#7;R1]1[#6]([$(S(=O)(=O)),$(C(F)(F)(F)),$(C#N),$(N(=O)(=O)),$([N+](=O)[O-]),$(C=O)])[#6]([F,Cl,Br,I])[#6][#6][#7]1	OTHER
CHEMBL	aromatic_Sulfonic_ester	[#6,#7]-S(=O)(=O)Oc	ALPHALISA
GSK_CTC	Thiadiazolesufinyl	O=Cc1nnsc1S(=O)[#6]	OTHER
Lily	biotin	[SD2r5]1-C-[C;R2]2-N-C(=O)-N-C2-C1	FRET, SPA & TRF
BMS	CH2_S#O_3_ring	[CH2]1[O,S]C1	ABSORB & OTHER
NIH	epoxide_aziridine_thioepoxide	[CH2]1[O,S,N]C1	ABSORB & OTHER
GSK_CTC	Nitrothienylpropene	[#6][C;!H0]=[C;!H0]c1ccc(s1)N(~[OD1])~[OD1]	OTHER
NIH	saponin	O1CCCCC1OC2CCC3CCCCC3C2	SPA
NIH;BMS	misc_7_aliphatic_OH;gte_7_aliphatic_OH	C[O;D1].C[O;D1].C[O;D1].C[O;D1].C[O;D1].C[O;D1].C[O;D1]	SPA
GSK_CTC	Thiazolequinolineamine	N(c1ccccc1)c1ccnc2cc(ccc12)-c1nccs1	ALPHALISA
PainsA	anil_di_alk_E	[CH]N([CH2])c1[cH][c;$([cH]),$(c[C!H0!H1])]c([C!H0;$([CH3]),$(C[C!H0])])[cH][cH]1	ALPHALISA
NIH	alkynyl_michael_acceptor2	[CH1]#CC(=O)-[#6,#7,#8]	ALPHALISA & OTHER
GSK_CTC	Nitrofuranylpropene	[#6][C;!H0]=[C;!H0]c1ccc(o1)N(~[OD1])~[OD1]	OTHER
GSK_CTC	Dihydroisoquinoline	[$([#6][C;!H0]=[C;!H0]C1NCCc2ccccc12),$([#6][CH]=[CH]C1=NCCc2ccccc12)]	ABSORB
NIH;BMS	aryl_phosphonate	P(=O)-[O;!R]-a	TRF
CHEMBL	Filter79_maleimide	[CH]1C(=O)NC(=O)[CH]=1	ABSORB & OTHER
NIH;BMS	maleimide_etc	[$([C;H1]),$(C(-[F,Cl,Br,I]))]1=[$([C;H1]),$(C(-[F,Cl,Br,I]))]C(=O)[N,O,S]C(=O)1	ABSORB & OTHER
GSK_CTC	Pyridoguanidine	[$([#7]-,=C(!@-,!@=N!@-,!@:c:1n:a:a:a:a:1)!@-,!@=N!@-,!@:c:1:a:a:a:a:a:1),$([nX2r6]([c!$(c[OH])!$(c=O)])c[N;$(N=C([NH2])[NH]a),$([NH]C([NH2])=Na)])]	ABSORB
GSK_CTC	Pyrimidinesulfone	[#6]S(=O)(=O)c1ncc([Cl,Br])c(n1)C([#7])=O	OTHER
Lilly	8_hydroxyquinoline	[OD1]c1cccc2cccnc12	ALPHALISA

References for "Sources" column in Table 2S, 3S, 4S and 5S

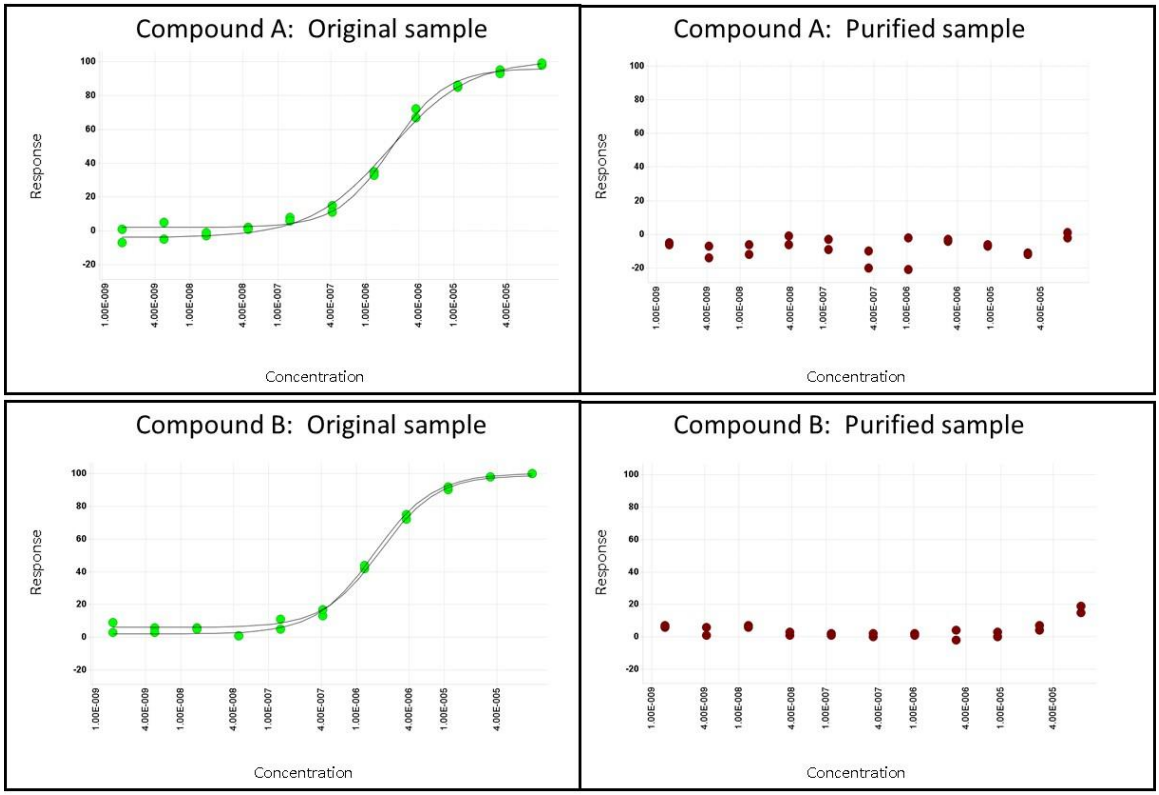
Source	**Reference** (Ref number in main text)
BMS	Pearce *et al* (15)
ChEMBL	Bento *et al* (17)
ChEMBL-fixes	Bento *et al* (17) (filters updated after private communication with ChEMBL)
GSK	Calculated physchem properties on GSK compounds
GSK_CTC	This article
Lilly	Bruns & Watson (16)
NIH	Jadhav *et al* (8)
PainsA	Baell & Holloway (10)
PainsB	Baell & Holloway (10)
ZINC15	Sterling & Irwin (18)

Figure 2S: IFI profile of Abbott Tiers in the GSK HTS collection



Supplementary information on analysis of active contaminant from ion-exchange purification

Figure 3S: Dose-response curves for two ZAP70 screening hits before and after re-purification



	ZAP70 pIC50 (original sample)	ZAP70 pIC50 (re-purified sample)
Compound A	5.7	<4.1
Compound B	5.8	<4.1

Experimental procedure

A 20 g acidic ion exchange cartridge was eluted with methanol 100 mL. The eluent was concentrated to give 2 mg of material.

LCMS of eluted material does not give a response (UV, ELSD) and 1H NMR (MeOD) showed a complex mixture: The sample was submitted to the assay (based on arbitrary molecular weight of 300). The dose response curve (Fig 4S) is typical of a reversible inhibitor.

Figure 4S: Dose-response in ZAP70 assay of eluate from acidic resin

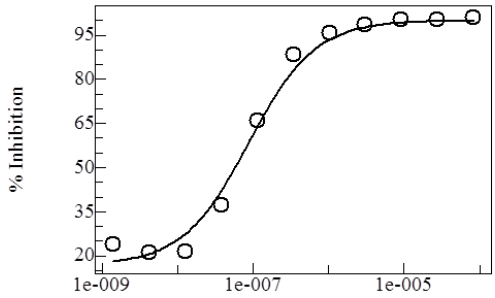


Table 6S: Analysis of IFI profile of noisy combinatorial chemistry arrays in GSK HTS collection

ARRAY_ID	IFI ≤ 0.00	0.00 < IFI ≤ 1.00	1.00 < IFI ≤ 2.00	2.00 < IFI ≤ 5.00	5.00 < IFI ≤ 10.00	10.00 < IFI	Number of Noisy Compounds	Total Compounds	%Noisy
103749192	1032	1114	611	832	415	203	1450	4207	34.5%
202145195	508	695	682	1400	858	183	2441	4326	56.4%
206524080	1701	918	811	590	164	131	885	4315	20.5%
206524171	13	27	45	153	177	123	453	538	84.2%
104281073	67	268	485	1327	898	118	2343	3163	74.1%
106803036	349	334	245	464	211	98	773	1701	45.4%
105308060	1748	1129	977	798	385	90	1273	5127	24.8%
105308053	338	285	488	1185	795	85	2065	3176	65.0%
108293006	919	459	429	571	316	80	967	2774	34.9%
206166171	1000	579	524	552	144	60	756	2859	26.4%