

Metadata standard and data exchange specifications to describe, model and integrate complex and diverse high-throughput screening data from the Library of Integrated Network-based Cellular Signatures (LINCS)

Uma D. Vempati¹, Caty Chung¹, Chris Mader¹, Amar Koleti¹, Nakul Datar¹, Dušica Vidović¹, David Wrobel², Sean Erickson², Jeremy Muhlich³, Gabriel Berriz³, Cyril Benes⁴, Aravind Subramanian⁵, Ajay Pillai⁶, Caroline E. Shamu^{2,3}, Stephan C. Schürer^{*,1,7}

¹Center for Computational Science, University of Miami, Miami, FL; ²ICCB-Longwood Screening Facility, Harvard Medical School Boston MA; ³Dept. of Systems Biology, Harvard Medical School, Boston MA; ⁴Center for Molecular Therapeutics, Massachusetts General Hospital, Boston MA; ⁵Broad Institute, Cambridge, MA; ⁶National Human Genome Research Institute, National Institutes of Health, Rockville, MD; ⁷Department of Molecular and Cellular Pharmacology, University of Miami, Miami, FL.

*corresponding author, contact: sschurer@med.miami.edu

Supporting information

(I) Supporting Tables S1 to S5

Table S1. Sources of annotations and controlled vocabularies used in LINCS reagent metadata standards.

Metadata category / descriptor	Annotation source / controlled vocabulary	Reference URL
Cell Line	Cell Line Ontology (CLO)	http://www.clo-ontology.org
Cell line mutation	Catalogue Of Somatic Mutations In Cancer (COSMIC)	http://cancer.sanger.ac.uk/cancergenome/projects/cosmic
Cell type	Cell type Ontology (CL)	http://cellontology.org
Organism	NCBI Taxonomy	http://www.ncbi.nlm.nih.gov/taxonomy
Organ and tissue	Uber Anatomy Ontology (UBERON)	http://uberon.org
Gene	NCBI Gene, HUGO	http://www.ncbi.nlm.nih.gov/gene , http://www.genenames.org
Protein	UniProt	http://www.uniprot.org
siRNA / shRNA	NCBI Probe	http://www.ncbi.nlm.nih.gov/probe
Small molecule	PubChem, DrugBank, ChEBI, and others	http://pubchem.ncbi.nlm.nih.gov , http://www.drugbank.ca/ , http://www.ebi.ac.uk/chebi/
Antibody	NIF antibody registry	http://antibodyregistry.org/
Disease	Human Disease Ontology (DOID)	http://bioportal.bioontology.org/ontologies/1009
Biological process	Gene Ontology (GO)	http://www.geneontology.org/

Table S2. Excerpt of annotation of cell lines from LINCS assays with the metadata standards and corresponding terminology.

Cell Name	Alternate Name	Alternate ID	Provider	Provider Catalog ID	Organism	Organ	Cell type	Disease term	DOID	Growth properties	Organism gender
MCF7		MGH:403; COSMIC:905946	ATCC	HTB-22	Homo sapiens	breast	epithelial	breast fibrocystic disease	3458	adherent	female
MCF 10A	MCF10A		ATCC	CRL-10317	Homo sapiens	breast	epithelial	breast adenocarcinoma	10354	adherent	female
SK-BR-3	SKBr3		ATCC	HTB-30	Homo sapiens	breast	epithelial	breast adenocarcinoma	3458	adherent	female
MDA-MB-231	MDAMB231; MDA-MB231	COSMIC:905960	ATCC	HTB-26	Homo sapiens	breast	epithelial	breast carcinoma	3458	adherent	female
BT-20		MGH:408; COSMIC:906801	ATCC	HTB-19	Homo sapiens	breast		prostate adenocarcinoma	3459	adherent	female
PC-3 [JPC-3]	PC3	ATCC:CRL-1435; MGH:877; COSMIC:1240202	JHSF	JCRB0077	Homo sapiens	lung		non-small cell lung carcinoma	2526	adherent	female
A549		MGH:677; COSMIC:905949	ATCC	CCL-185	Homo sapiens	lung		skin melanoma	3908	adherent	male
A-375	A375	COSMIC:906793	ATCC	CRL-1619	Homo sapiens	skin	epithelial	hepatocellular carcinoma	8923	adherent	female
Hep G2	HEPG2	MGH:650	ATCC	HB-8065	Homo sapiens	liver	epithelial	prostate cancer	684	adherent	male
Vcap	VCaP		ATCC	CRL-2876		prostate	epithelial	non-small cell lung carcinoma	10283	adherent	male

Table S3. Example annotation of LINCS protein reagents using proposed metadata standards.

Protein (by gene symbol)	Phosphorylation status	Mutation	UniProt accession	Alternate names
AAK1			Q2M2I8	AP2 associated kinase 1; AAK1; DKFZp686F03202; DKFZp686K16132; FLJ23712; FLJ25931; FLJ31060; FLJ42882; FLJ45252; KIAA1048; MGC138170; MGC164568; MGC164570
ABL1			P00519	c-abl oncogene 1, receptor tyrosine kinase; ABL; ABL1; bcr/abl; c-ABL; JTK7; p150; v-abl
ABL1		E25K	P00519	c-abl oncogene 1, receptor tyrosine kinase; ABL; ABL1; bcr/abl; c-ABL; JTK7; p150; v-abl
ABL1	Phosphorylated	E25K	P00519	c-abl oncogene 1, receptor tyrosine kinase; ABL; ABL1; bcr/abl; c-ABL; JTK7; p150; v-abl
ABL1		F317I	P00519	c-abl oncogene 1, receptor tyrosine kinase; ABL; ABL1; bcr/abl; c-ABL; JTK7; p150; v-abl
ABL1	Nonphosphorylated	F317I	P00519	c-abl oncogene 1, receptor tyrosine kinase; ABL; ABL1; bcr/abl; c-ABL; JTK7; p150; v-abl
ABL1	Phosphorylated	F317I	P00519	c-abl oncogene 1, receptor tyrosine kinase; ABL; ABL1; bcr/abl; c-ABL; JTK7; p150; v-abl
ABL1		F317L	P00519	c-abl oncogene 1, receptor tyrosine kinase; ABL; ABL1; bcr/abl; c-ABL; JTK7; p150; v-abl
ABL1	Nonphosphorylated	F317L	P00519	c-abl oncogene 1, receptor tyrosine kinase; ABL; ABL1; bcr/abl; c-ABL; JTK7; p150; v-abl
ABL1	Phosphorylated	F317L	P00519	c-abl oncogene 1, receptor tyrosine kinase; ABL; ABL1; bcr/abl; c-ABL; JTK7; p150; v-abl
ABL1		H396P	P00519	c-abl oncogene 1, receptor tyrosine kinase; ABL; ABL1; bcr/abl; c-ABL; JTK7; p150; v-abl

Table S4. Selected annotation of LINCS small molecules using LINCS metadata standards and including information about DrugBank indication (for a full list see <http://lifekb.org/index.php/data-standards>).

LINCS ID	Name	Alternative Names	PubChem CID	SMILES	DrugBank Indication	NCBI MLP Probe ID	NCBI Probe URL	PubChem Library category	Molecular Weight	Molecular Formula	clogP
LSM-1001	(R)- Roscovitine	(R)- Roscovitine;CVC 202;Seliciclib	160355	<chem>CC(C@H)[C@H](CO)N1C1=NC2=C(C(=N1)NCCC3=CC=CC=C3)N=CN2C(C)C</chem>					354.2	C19H26N6O	2.89
LSM-1002	ALW-ll-38-3	ALW-ll-38-3	24880028	<chem>CC1=C(C(=C(C=C1)N(C)=O)C2=CC=CC(=C2)N3C=C(N=C3)C(F)N(C)=O)C4=CC=NO4</chem>					469.1	C23H18F3N5O3	2.51
LSM-1004	AT-7519	AT-7519	11338033	<chem>C1CNCCC1NC(=O)C2=C(C(=NN2)N(C)=O)C3=C(C(=CC=C3)C)I</chem>					381.1	C16H17Cl2N5O2	-0.12
LSM-1005	AV-951	AV-951;Tivozanib	9911830	<chem>CC1=CC(=NO1)N(C(=O)N2C=C(C=C(C=C2)O)C3=C4C=C(C(=CC4=NC=C3)O)O)C1</chem>					454.1	C22H19ClN4O5	3.57
LSM-1006	AZD7762	AZD7762		<chem>C1C(C@H)[CNC1]N(C)=O)C2=C(C(=C(C(=C2)C3=CC(=CC=C3)F)N(C)=O)N</chem>					362.1	C17H19FN4O2S	2.40
LSM-1007	AZD8055	AZD8055	25262965	<chem>C(C@H)[C@H](CO)N1C2=NC(=NC3=C2C=CC(=N3)C4=CC(=C(C=C4)O)O)N5CCOC(C@H)5C</chem>					465.2	C25H31N5O4	3.53
LSM-1008	BAY-439006	BAY-439006;Sorafenib	216239	<chem>CNC(=O)C1=NC=CC(=C1)OC2=CC=C(C(=C2)N(C(=O)N(C3=CC(=C(C=C3)C)C(F)C(F)C)C(=O)N(C3=CC(=C(C=C3)O)O)N5CCOC(C@H)5C</chem>	Antineoplastic;For the treatment of patients with advanced renal cell carcinoma.			MLSMR	464.1	C21H16ClF3N4O3	4.34
LSM-1009	CP466722	CP466722	44551660	<chem>COC1=C(C=C2C(=C1)C(=NC=N2)N3C(=NC(=N3)C4=CC=CC(=N4)O)C</chem>					349.1	C17H15N7O2	2.35
LSM-1011	Flavopiridol	Flavopiridol;Alv oxidib;HMR-1275;L868275	5287969	<chem>C=C(C3=C2O(C(=CC3=O)C4=CC=CC=C4)C1O)O</chem>					401.1	C21H20ClNO5	2.46
LSM-1012	GSK429286A	GSK429286A	11373846	<chem>CC1=C(C(C(C(=O)N1)C2=CC=C(C(=C2)C(F)F)F)C(=O)N(C3=C(C=C4(C=C3)C=NN4)F)C(=O)N(C3=CC(=C(C=C4)CN</chem>					432.1	C21H16F4N4O2	1.93
LSM-1013	GSK461364	GSK461364	15983966	<chem>C(C@H)(C1=CC=CC=C1C(F)F)OC2=C(S(C(=C2)N3C=NC4=C3C(=C(C=C4)CN5CCN(C)C(=O)N</chem>					543.2	C27H28F3N5O2S	3.66
LSM-1020	Dasatinib	Dasatinib;BMS-354825;Sprycel	3062316	<chem>CC1=C(C(=CC=C1)C)N(C(=O)C2=CN=C(C2)N(C3=NC(=NC(=C3)N4CCN(C)C4)CCO)C</chem>	For the treatment of adults with chronic, accelerated, or myeloid or lymphoid blast phase chronic myeloid leukemia with resistance or intolerance to prior therapy. Also indicated for the treatment of adults with Philadelphia chromosome-positive acute lymphoblastic leukemia with resistance or intolerance to prior therapy.			MLSMR	487.2	C22H26ClN7O2S	2.52
LSM-1323			16190867	<chem>CCCNC1(C(=C(N1)CN2CCCC(C2)C(=O)C3=CC(=CC=C3)OC(C)C</chem>		ML157	https://pub.ncbi.nlm.nih.gov/pubmed/15151	MLSMR	383.3	C23H33N3O2	4.02

Table S5. Annotation of disease sub-type and mutation information of ovarian cancer cell lines used at LINCS.

Cell line	Disease	Gene mutations
SW 626	ovary adenocarcinoma	APC; KRAS; SMAD4; TP53
TOV-21G	ovarian clear cell adenocarcinoma	ARID1A; KRAS; PIK3CA; PTEN
OVISE	ovary adenocarcinoma	ARID1A; PPP2R1A
OVTOKO	ovary adenocarcinoma	ARID1A; PPP2R1A
ES-2	ovarian clear cell adenocarcinoma	BRAF
OV-90	papillary serous adenocarcinoma	BRAF
KGN	granulosa cell tumor	CDKN2A; CDKN2a(p14); FOXL2
SK-OV-3	ovary adenocarcinoma	CDKN2A; CDKN2a(p14); MLH1; PIK3CA; TP53
TYK-nu	ovarian carcinoma	CDKN2A; CDKN2a(p14); NRAS; TP53
TOV-112D	endometrioid ovary carcinoma	CTNNB1
Caov-3	ovary adenocarcinoma	FAM123B; PTEN; STK11; TP53
MCAS	ovarian mucinous cystadenocarcinoma	KRAS; PIK3CA; SMAD4
OAW28	ovarian cystadenocarcinoma	MAP2K4
EFO-27	ovarian mucinous adenocarcinoma	MSH2; PTEN; TP53
PA-1	teratocarcinoma	NRAS; RB1
OAW42	ovarian cystadenocarcinoma	PIK3CA
A2780	ovary adenocarcinoma	PTEN
OVCAR-3	ovary adenocarcinoma	RB1; PIK3R1; SMARCA4; TP53
KURAMOCHI	ovarian carcinoma	TP53
EFO-21	ovarian cystadenocarcinoma	TP53
OC-314	ovarian cystadenocarcinoma	TP53
OVCAR-4	ovary adenocarcinoma	TP53
Caov-4	ovary adenocarcinoma	TP53; RB1
RMG-I	ovarian clear cell adenocarcinoma	
COV644	ovarian mucinous adenocarcinoma	
A2780ADR	ovary adenocarcinoma	
A2780cis	ovary adenocarcinoma	
COLO-704	ovary adenocarcinoma	
OVKATE	ovary adenocarcinoma	
OVMIU	ovary adenocarcinoma	
OVSAYO	ovary adenocarcinoma	
RKN	ovary leiomyosarcoma	
OV7	ovary mixed epithelial carcinoma	
FU-OV-1	papillary serous adenocarcinoma	

(II) Supporting Figure S1 to S2

```

{
  "assayDescription": "The KINOMEScan assay platform is based on a competition binding ...",
  "assayProtocol": "1. T7 kinase-tagged phage strains are grown in parallel in 24-well or ...",
  "assayProtocolReference": "KINOMEScan website: http://kinomescan.com/Technology ...",
  "assayTitle": "Sorafenib KINOMEScan",
  "bioAssay": "KINOMEScan",

  "datapointFile": {
    "cols": [
      "dataRecordID",
      "hmsDatasetID",
      "smCenterCompoundID",
      "smSalt",
      "smCenterSampleID",
      "smLincsID",
      "smName",
      "cName",
      "cLincsID",
      "cCenterSpecificID",
      "ppName",
      "ppLincsID",
      "recordedPlate",
      "recordedWell",
      "controlType",
      "datapointName",
      "datapointUnit",
      "datapointValue"
    ],
    "datapoints": [
      {
        "cell": null,
        "comments": null,
        "dataType": "Text",
        "datapointName": "screeningConcentrationUnit",
        "datapointUnit": null,
        "description": null,
        "displayName": "Conc unit",
        "precision": null,
        "protein": null,
        "readoutType": null,
        "replicate": null
      },
      {
        "cell": null,
        "comments": null,
        "dataType": "Numeric",
        "datapointName": "screeningConcentration",
        "datapointUnit": null,
        "description": null,
        "displayName": "Assay compound conc",
        "precision": 0,
        "protein": null,
        "readoutType": null,
        "replicate": null
      },
      {
        "cell": null,
        "comments": null,
        "dataType": "Numeric",
        "datapointName": "percentControl",
        "datapointUnit": null,
        "description": null,
        "displayName": "% Control",
        "precision": 1,
        "protein": null,
        "readoutType": "Quantitative PCR",
        "replicate": null
      }
    ]
  },
  "noCols": 18,
  "noDatapoints": 3,
  "uri": "http://lincs.hms.harvard.edu/db/api/v1/datasetdata/20020"
},

  "dateDataReceived": "2011-01-25",
  "dateLoaded": "2011-07-15",
  "datePubliclyAvailable": "2011-07-15",
  "hmsDatasetID": "20020",

  "safVersion": "0.1",

  "screeningFacility": "HMS",
  "screeningLabInvestigator": "Qingsong Liu",
  "screeningPrincipalInvestigator": "Nathanael Gray",
  "usageMessage": null
}

```

Legend:

Color coded boxes to identify the logical sections in a SAF:	Color coded text to identify SAF fields:
 Header	BAO/LIFE Ontology Concepts
 BioAssay Description	LINCS Metadata Standards
 Data Description	House Keeping Data Exchange

Figure S1. Illustration of the assay Simple Annotation Format (SAF) to describe assay and result metadata and to facilitate the exchange of this information via web services.

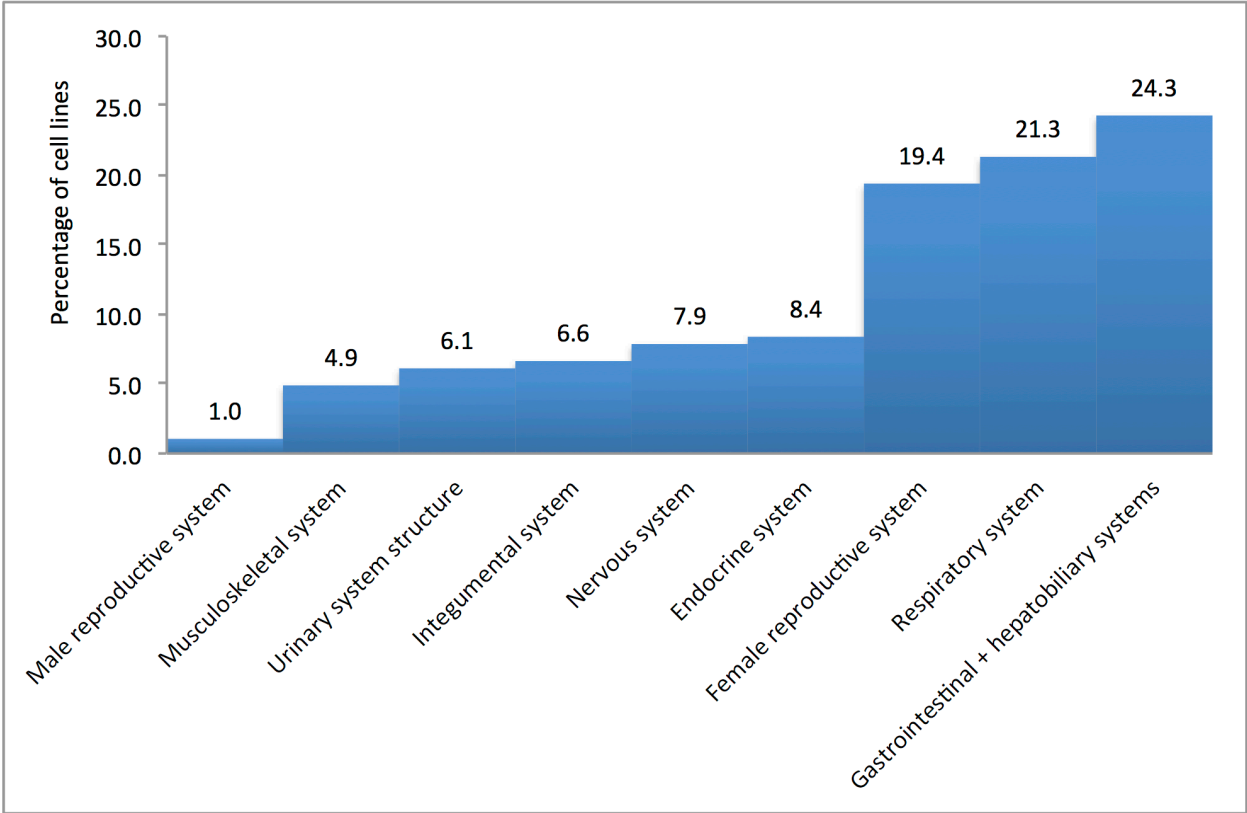


Figure S2. Distribution (percentage) of cancer cell lines tested in LINCS assays by their organs of origin.