Standard Operating Procedure For Upload Download

Sysbiocube Upload Procedure:

Sysbiocube upload depends on different data type and experiment type across different disease and species. Currently the SysBioCube system is familiar with known data type for both PTSD and Plague studies for mouse, human and non-human primate (Rehesus macaque) species.

Upload procedure for Transcriptomics (mRNA microarray & miRNA microarray) and Methylation data:

SysBioCube system accepts transcriptomics data for all species types from PTSD and Plague studies at this moment. However system in not limited to specific disease, species or study type. To upload data sets from different studies other than PTSD and Plague, SysBioCube team has to review the data files so system can learn the new data type. For new data type please contact Sysbiocube@mail.nih.gov

Transcriptomics raw data

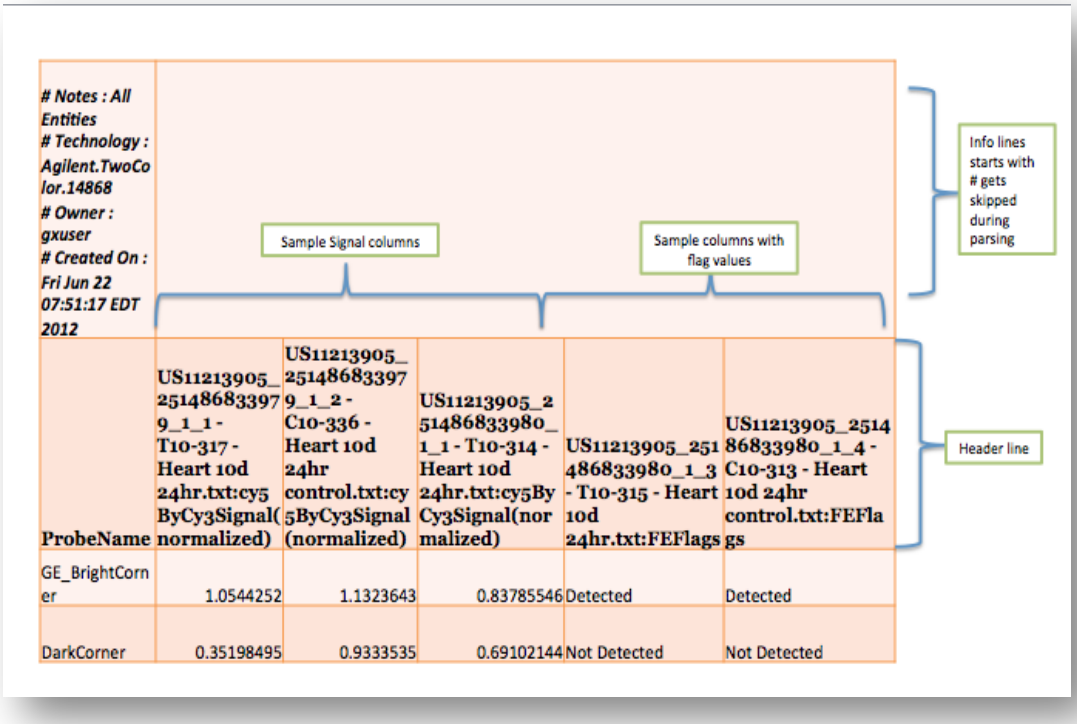
Data Type Raw:

- File name format: Instrument-Tag_study_species_strain_tissue_group_raw.txt (in mouse case)
  - For ex: 20190192_ptsd_mouse_C57_Amygdala_ T10R1 _raw.txt
- Instrument-Tag_study_species_ tissue_group_raw.txt (in other species case)
  - For ex: 20190192_ptsd_human_blood_52-52_Training_raw.txt

Transcriptomics normalized data

Data Type Normalized:

- File name format: study_species_strain_tissue_group_normalized_ normalizedMethodTag.txt (in mouse)
  - For ex: ptsd_mouse_BALBc_Amygdala_T10R1 _normalized_ quantile.txt
- File name format: study_species_strain_tissue_group__normalizedMethodTag_normalized.txt (in other species)
  - For ex: ptsd_human_all_Blood_52-52_Training_normalized_ quantile.txt



# Notes : All Entities # Technology : Agilent.TwoColor.14868 # Owner : gxuser # Created On : Fri Jun 22 07:51:17 EDT 2012					
ProbeName	US11213905_251486833979_1_1 - T10-317 - Heart 10d 24hr.txt:cy5ByCy3Signal(normalized)	US11213905_251486833979_1_2 - C10-336 - Heart 10d 24hr control.txt:cy5ByCy3Signal(normalized)	US11213905_251486833980_1_1 - T10-314 - Heart 10d 24hr.txt:cy5ByCy3Signal(normalized)	US11213905_251486833980_1_3 - T10-315 - Heart 10d 24hr.txt:FEFlags	US11213905_251486833980_1_4 - C10-313 - Heart 10d 24hr control.txt:FEFlags
GE_BrightCorner	1.0544252	1.1323643	0.83785546	Detected	Detected
DarkCorner	0.35198495	0.9333535	0.69102144	Not Detected	Not Detected

Transcriptomics analyzed data

- File name convention
  - Data Type Analyzed:
    - File name format: study_species_strain_method_parameter_analyzed.txt (in mouse)
      - For ex: ptsd_mus_C57_methodName_parameters_analyzed.txt
    - File name format: study_species_normalizedMethodTag_normalized.txt (in other species)
      - For ex: ptsd_homo_methodName_parameters_analyzed.txt

Entrez	Probe	Symbol	Name	log(FC)	P-value	FDR
75552	A_52_P450934	Paqr9	progestin and adipoQ receptor family member IX	-1.686810357	0.000388903	0.999868825
12877	A_52_P462296	Cpeb1	cytoplasmic polyadenylation element binding protein 1	-1.915865875	0.000415534	0.999868825

Notes : Created from Advanced Analysis operation: significance Analysis.
#EntityList : Filtered on Flags [Detected, Not Detected] C57B6 5d24h
#Interpretation : C57B/6 5d24h
#Experiment: Mouse PTSD C57B6 BALBc D8A
#corrected p-value cut-off:0.05
#Selected Test : Moderated T-Test
#p-value computation: Asymptotic
#Multiple Testing Correction: No Correction
Technology : Agilent.TwoColor.14868
Owner : gxuser
Created On : Thu May 03 10:50:28 EDT 2012

Probe Name	p	Regulation	FC (abs)	FC	Log FC	[C57B6_C_5d_24h] (normalized)	[C57B6_SS_5d_24h] (normalized)	Gene Symbol	Description	Chromosome End Index_Avadis_	Chromosome Number_Avadis_	Chromosome Start Index_Avadis_	Chromosome Strand_Avadis_	ControlType	Cytoband	EntrezGeneID	GenbankAccession	Gene Name	GenomicCoordinates	RefSeqAccession	UniGeneID
A_51_P430630	5.52E-06	up	5.3299003	5.3299003	2.4141085	-0.6797078	1.7344007	Gpr33		53124437	chr12	53124378	-	FALSE	mm|12qC1	14762	NM_008159	G protein-coupled receptor 33	chr12:53124437-531248	NM_008159	Mm.12890

miRNA Microarray Data

Analyzed File Format

* Required File Name: ptsd_mouse_BalBc_Heart_T10R1_analyzed.txt
* Platform Support: Agilent miRNA microarray/Exiqon LNA[a]

Analyzed file for miRNA microarray

No	ProbeID	Symbol	C5-127	C5-128	C5-129	C5-130	T5-135	T5-136	T5-137	T5-138	logFC	AveExpr	PA call	t	P.Value	FDR (BH)
1	4390	mmu-miR-7b/rno-miR-7b	5.840771897	5.84792758	5.761580841	5.711284816	6.04409246 1	5.9716431 73	5.8590964 61	5.832224449	0.136372853	5.85857771	P	2.46523851 1	0.037601108	0.116787725
2	4610	hsa-miR-126/ mmu-miR-126-3p/ rno-miR-126	9.444115922	9.656317756	9.532229164	9.439893976	9.4995795 29	9.4965992 71	8.9823970 12	9.08876986	-0.251302786	9.392487811	P	-2.02266226 86	0.075999875	0.183679 17
3	4700	hsa-miR-140-5p/ mmu-miR-140/ rno-miR-140	5.922212774	6.02494625	5.706855107	5.858864063	5.2008297 29	5.3952408 51	5.2726099 05	5.560283764	-0.520978486	5.617730305	A	-5.78658805	0.000343952	0.010487033

Metabolomics data:

SysBioCube system is familiar with all kinds of metabolomics data from Metabolon Company. Currently system reads the standard metabolon spreadsheet.
For automated upload one has to convert ".`xls` "spreadsheet to tab delimited text file.

- Data Type (Metabolon Standard Data)
  - Excel File with multiple spread sheets
    - study_species_strain_metabolon.xls (in mouse)
      - For ex: ptsd_mus_C57_metabolon.xls
    - study_species_metabolon.xls (in other species)
      - For ex: ptsd_homo_metabolon.xls

  - Required information:
    - Sample annotation
    - Species build
    - Scientific question (optional)
    - Experiment design (if metabolon report is not submitted)

- Data Type (Metabolon Standard Data to text file)
  - Excel File converted to text file
  - Excel file first spread sheet (convert to text)
    - Naming convention for first spread sheet text file
      - study_species_strain_metabolon_raw.txt(in mouse)
      - For ex: ptsd_mus_C57_metabolon_raw.txt
      - study_species_metabolon_raw.txt (in other species)
      - For ex: ptsd_homo_metabolon.txt
  - Excel file second spread sheet (convert to text)
    - File naming convention for rest of the spread sheets to text file for Metabolon original xls file

- study_species_strain_metabolon_analyzed.txt(in mouse)
- For ex: ptsd_mus_C57_metabolon_analyzed.txt
- study_species_metabolon_analyzed.txt (in other species)
- For ex: ptsd_homo_metabolon_analyzed.txt

Metabolon File Format

										SAMPLE_NAME	WALT-00001-LC-01	WALT-00002-LC-01	WALT-00003-LC-01	WALT-00004-LC-01	WALT-00002-LC-01	WALT-00003-LC-01
										CLIENT_IDENTIFIER	C05-244	C05-245	C05-246	C05-247	C05-245	C05-246
										PARENT_SAMPLE_ID	625226	625227	625228	625229	625227	625228
										TREATMENT	Control	Control	Control	Control	Control	Control
										TREATMENT_DURATION	5d	5d	5d	5d	5d	5d
										COLLECTION_TIME	1d	1d	1d	1d	1d	1d
										TUBE_ID	C5244	C5245	C5246	C5247	C5245	C5246
										EXP_NO	14	14	14	14	14	14
										COLLECTION_DATE	5/19/10 0:00	5/19/10 0:00	5/19/10 0:00	5/19/10 0:00	5/19/10 0:00	5/19/10 0:00
										DAY_TO_RUN	1	2	1	2	2	1
										TIMEPOINT	T1	T1	T1	T1	T1	T1
										TIMEPOINT2	EARLY	EARLY	EARLY	EARLY	EARLY	EARLY
BIOCHEMICAL	SUPER_PATHWAY	SUB_PATHWAY	LIB_ID	COMP_ID	PLATFORM	RI	MASS	CAS	KEGG	Group HMDB	1	1	1	1	1	1
1,2-propanediol	Lipid	Ketone bodies	50	38002	GC/MS	1041	117	57-55-6;	C00717,C02912,C00583,C01506,C02917	HMDB01881	25268992	23503840	23458874	23616094	23503840	23458874

Proteomics analyzed data:

RBM File Format

* Required File Name: ptsd_mouse_C57_Plasma_rbm_raw.txt
* Platform Support: RBM

RBM file for Proteomics

Analytes	Apolipoprotein A-I (Apo A-I)	CD40 (CD40)	CD40 Ligand (CD40-L)	C-Reactive Protein Mouse (CRP Mouse)
Units	ug/mL	pg/mL	pg/mL	ug/mL
Least Detectable Dose	2.4	4.5	414	0.90
Samples				
C5-244	159	43	1890	3.6
C5-245	130	41	1400	4.3

Methylation Array Data

Analyzed File Format

* Required File Name: ptsd_mouse_C57_Hemi-brain_T10R1_analyzed.txt
* Platform Support: AgilentÕs Human/Mouse CpG Microarrays

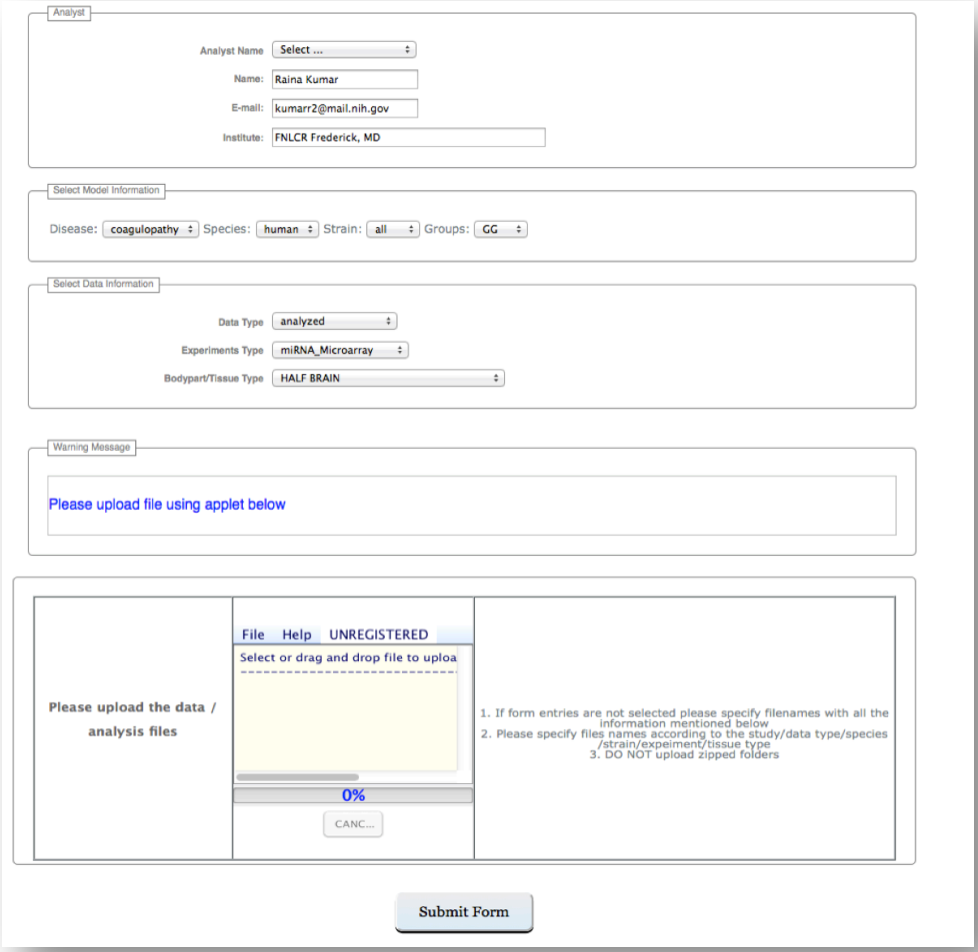
Analyzed file for Methylation microarray

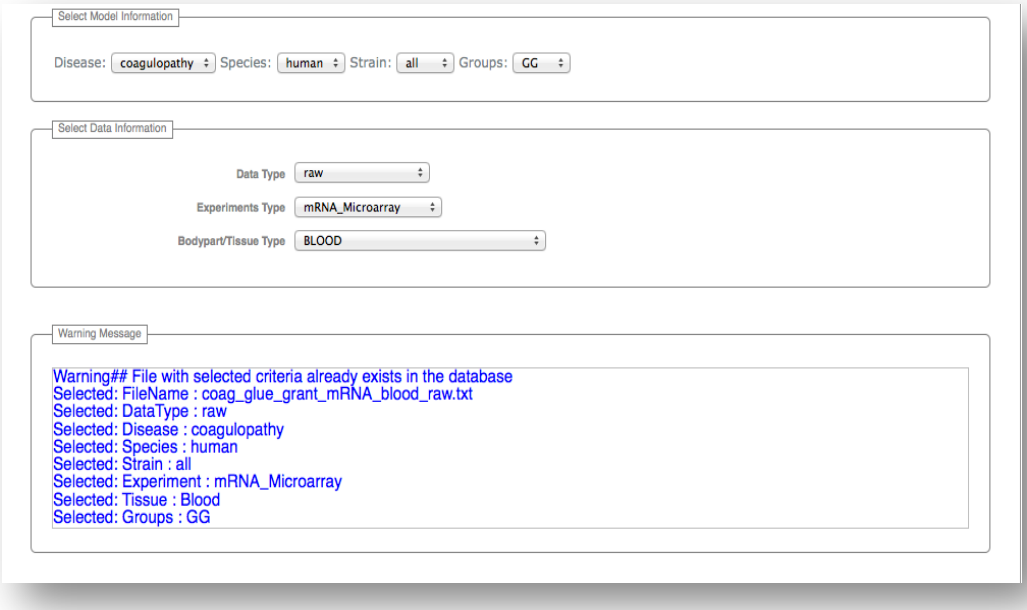
Methylation Analyzed File Format

							sample	251527910715_251527910715_1_1 T10-107 10day 24hrs	251527910715_251527910715_1_2 C10-93 10day 24hrs Control
							recovery_treatment	SD_10d24h	C_10d24h
							treatment_recovery vs C	SD_10d24h	C
							treatment	SD	C
							recovery_treatment_id	SD_10d24h_107	C_10d24h_93
							sample	251527910715_251527910715_1_1 T10-107 10day 24hrs	251527910715_251527910715_1_2 C10-93 10day 24hrs Control
ProbeName	GeneName	Mm_mm9_CpGIsland_20090908	ChrName	Description	Start	Stop			
A_68_P20002076	Xkr4		chr1	ag_LACat|INSIDE	3521583	3521642		0	0
A_68_P20002763	Xkr4		chr1	ag_LACat|INSIDE	3660621	3660667		0.122632094	0.125661001

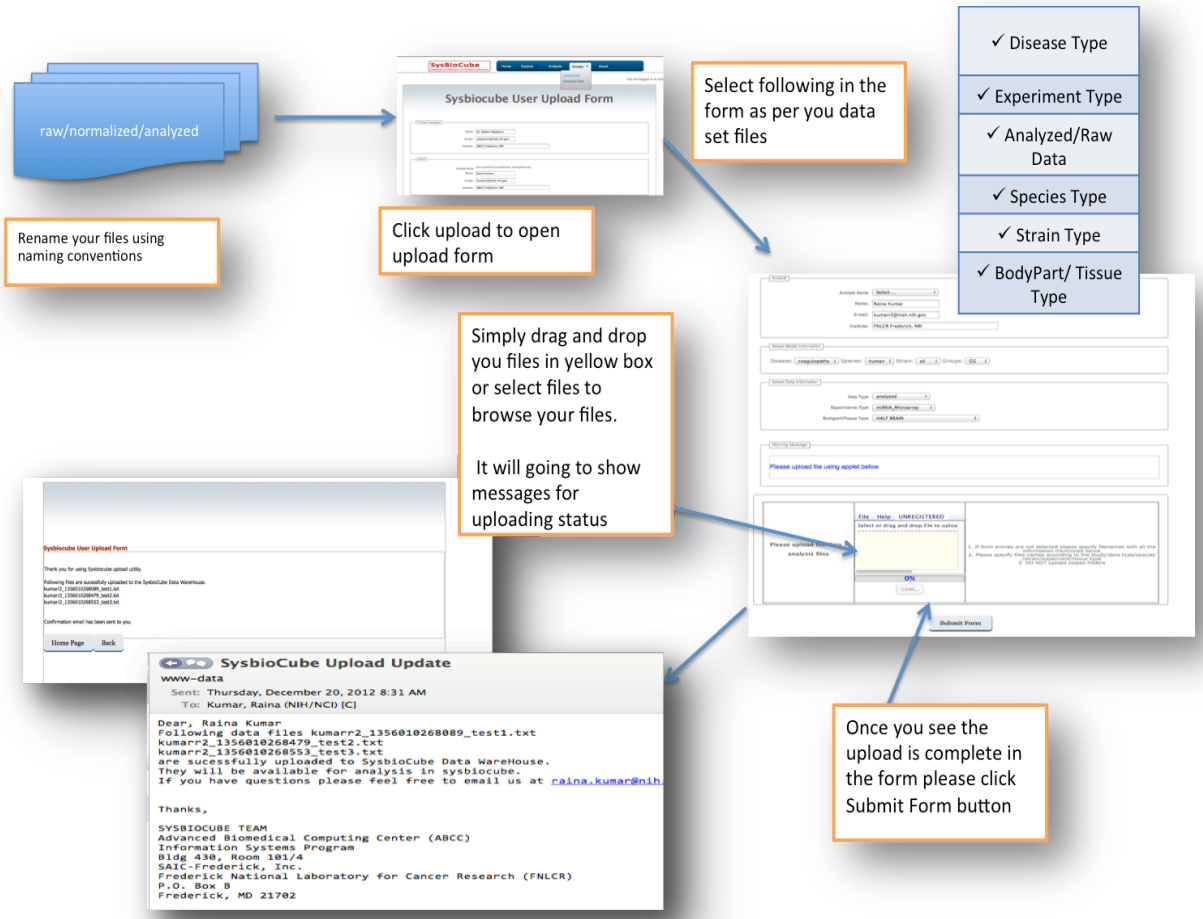
Standard Upload Workflow:

Upload Form





Workflow for uploading files:


raw/normalized/analyzed
Rename your files using naming conventions
Sysbiocube User Upload Form
Click upload to open upload form
Select following in the form as per you data set files
✓ Disease Type
✓ Experiment Type
✓ Analyzed/Raw Data
✓ Species Type
✓ Strain Type
✓ BodyPart/ Tissue Type
Simply drag and drop you files in yellow box or select files to browse your files.
It will going to show messages for uploading status
Please upload file using applet below
Submit Form
Sysbiocube User Upload Form
Thank you for using Sysbiocube upload utility.
Following files are sucessfully uploaded to the SysbioCube Data WareHouse.
Confirmation email has been sent to you.
Home Page
Back
SysbioCube Upload Update
www-data
Sent: Thursday, December 20, 2012 8:31 AM
To: Kumar, Raina (NIH/NCI) [C]
Dear, Raina Kumar
Following data files kumarr2_1356010268089_test1.txt
kumarr2_1356010268479_test2.txt
kumarr2_1356010268553_test3.txt
are sucessfully uploaded to SysbioCube Data WareHouse.
They will be available for analysis in sysbiocube.
If you have questions please feel free to email us at raina.kumar@nih.
Thanks,
SYSBIOCUBE TEAM
Advanced Biomedical Computing Center (ABCC)
Information Systems Program
Bldg 430, Room 101/4
SAIC-Frederick, Inc.
Frederick National Laboratory for Cancer Research (FNLCR)
P.O. Box B
Frederick, MD 21702
Once you see the upload is complete in the form please click Submit Form button