Guidelines for Depositing ICCB-L Screening Data

Overview:

1. Before any raw data are generated, meet with the Data Analyst and Data Curator to discuss the specifics of your template, data analysis, and data annotation.
2. Every time you generate new raw data, you should submit them via e-mail to the Data Curator, who will convert it into an Excel template for you. Use this template to perform analysis of your data and annotate your screening positives. Since this annotated template will be entered into our database, the data and the description of how it was analyzed must be as clear and complete as possible.
3. Along with the annotated data, you should provide a Primary Screen Report, which will describe the assay protocol and data analysis associated with your screen. This required document will be sent to you to be filled out and will be used by future researchers to decipher your data.
4. Cherry-pick requests may be submitted after annotated screening data and the Primary Screen Report are submitted. Cherry-pick requests should be e-mailed to the Data Analyst as an Excel file, including plate and well location. All requests must include a screen number. Please see the Cherry Pick Guidelines for additional information.
5. Follow-up data from cherry picks and secondary analyses should be submitted as they become available.

Step 1: Initial Meeting with Data Analyst

During your first visit to the screening facility, you should meet with the Data Analyst (David Wrobel) and Data Curator (Jen Splaine) to discuss data template and data deposition procedures. The purpose of this meeting is to describe the template that will be used to convert your raw data into a format for analysis and result annotation.

If your data will be acquired using a plate reader or image analysis software and produce numeric values, you will receive a template containing the raw results. If you do not generate numeric data (e.g., if you are conducting a visual screen), you will be provided with a template for entering your data. The Excel file will contain the data for all of the plates screened, with each well of each plate listed in a separate row.

Step 2: Submitting Raw Data

Raw data from your screen should be submitted before you begin to analyze it. Please use the following guidelines when submitting raw data:

a. Raw data should be submitted to Jen Splaine by e-mail, as Excel file attachments (.xls or .csv). Jen's e-mail address is:

   jennifer_splaine@hms.harvard. edu

b. Data should be in plate-grid format, except for imaging data. The file from the plate reader should not be edited, other than including the plate number for each grid within the file (this is optional, and most screeners don't do it). For imaging screens, raw data generated by microscope analysis software should be in column format, with plates on separate worksheets and wells in separate rows.

c. The e-mail accompanying the data files must include the following information:

i) Order of library plate numbers and replicates within each raw data file. You may specify ranges of plates and replicates (e.g., 1568A, 1568B, 1569A... 1571B) as long as it is clear how the replicates are ordered.

Examples:

"The file screen100_1568-1571.xls contains the data from plates 1568-1571, with replicates A and B for each plate grouped together. The order is thus 1568A, 1568B, 1569A, 1569B, and so on through 1571B". This example makes clear the order of grids within the specified file.

In contrast, "1568-1571" or "1568A-1571B" do not make it entirely clear how replicates are positioned in the file.

When possible, please give ranges for long series of plates rather listing each plate. If a long list of plates is given, it's very important to note any breaks in the sequence.

All of this information can be provided in the e-mail message or as plate keys in worksheets within the raw data files.

ii) Well positions of positive and negative controls. You may specify a column (e.g., "positive controls in column 24") if all wells in the column contain the same type of control.

iii) Description of positive and negative control well content

Examples:

"Negative control: DMSO in column 23; Positive control: 30 uM taxol in column 24"

"Negative control: (Wells C22 and D22) non-targeting siRNA #2 (Dharmacon). Positive controls: (Wells E22 and F22) Plk1 siRNA pool (Dharmacon Cat#: M-003290-01); (Wells G22 and H22) Eg5 Smartpool (Dharmacon Cat#: xxxxxxx-01)."

iv) Type of readout (e.g., fluorescence intensity, luminescence, absorbance at 405 nm, imaging)

v) Description of any anomalies (e.g., plates that should be omitted, plates out of sequence, control position differences among plates, reverse order of plates, etc.).

vi) Files for different phenotypes and readouts should be clearly specified. For example, if you read both fluorescence polarization and intensity, then you should specify which files correspond to each readout mode.

vii) Please state the screen number for all data files. This is particularly important if you have a counterscreen or multiple screening projects.

Step 3: The Data Template

The data template you receive will show one row per well screened. For multiple data points, each raw data reading will be placed in a separate column. If you run your screen in duplicate, there will be two result columns, one for each raw result, all in the same row. This template is the mandatory starting point for returned analysis of your data.

The columns in the template you receive will be Plate, Well, Type, Exclude, and at least one raw result column per data reading. The default name for each raw result column is the data type and unit of the raw result. In addition, you may add a title for each pass of the plate screened (e.g. Run 1, Run 2; or 24hrs, 48hrs; or Reporter Cells, Control Cells). If you did not use a plate reader or imaging analysis software, there will be no raw result columns, and you will need to fill in your data as explained in the following section.

The options for Well Type are P = Positive Control (control that mimics a screening positive), N = Negative Control (baseline), X = Experiment, E = Empty, or O = Other. The well types will be defined by you and the Data Curator before the template is created.

As noted above, please specify the well locations for positive and negative controls when submitting raw data.

Please see Figure 3.1 for a template example.

Figure 3.1: An example of the first 4 data rows for one plate of a screen run in duplicate:

Plate	Well	Type	Exclude	Intensity_Run1	Intensity_Run2
1989	A01	P		5563	5758
1989	A02	P		6559	6309
1989	A03	X		1621	1624
1989	A04	X		1523	1626

Step 4: Analyzing and Annotating your Data

You may use Excel or other software programs to analyze your results, but your analysis results **must** be returned in the template provided to you. This is the file that will be used for entry into our database. The columns in the analysis result file must include Plate, Well, Type, Exclude, and all raw data. With the exception of visual data, this information is included in the template provided. All rows should be returned, including those for control wells, empty wells, excluded wells, and wells that are not screening positives.

You are expected to add annotations for visual data, analyzed results, screening positives, and comments. Additional information, such as calculated data necessary for determining your hits (such as z-scores, etc.) should be added as new columns to the template.

Figure 4.1: Possible column headers added to the template

Column Name	Required	Values	Purpose
Exclude	Yes	Y (for Yes), rep A or repB if just one replicate, or blank	If you exclude a well from further analysis because of well-based or data issues, enter a Y.
Result Numeric	Required if there is one final numeric result.	Numeric	Analyzed result (Fold induction, z-score, etc.)
Result Text	Required if results are images	Text	Annotations about visual data
Analysis	No	Numeric	Intermediate analysis values
Positive	Yes	S/M/W/blank, or Y/blank	Designate screening positives as Strong/Medium/Weak, or Yes. Note that this designation reflects the confidence in a positive, not the potency.
Comment	No	Text	Add comments where appropriate

Exclude: the purpose of the Exclude column is to note wells that you are excluding from your analysis due to well-based problems or data issues (e.g., certain wells are known not to have received appropriate reagents, or severe edge effects).

Result Numeric: in addition to the columns provided in the template please add a Result Numeric column with the desired column header containing any final numeric result, if applicable (e.g., FI, z-score, etc.). You may include additional, appropriately titled Analysis columns to show intermediate steps used to calculate results (e.g., plate-normalized values, percent of control values, etc.). Please see below for examples.

Result Text: if your analysis includes visual data, enter analysis text results from visual data in a Result Text column with the desired column header. If your screen includes plate reader data and visual data, enter analyzed plate reader data in a Result Numeric column and analysis results from visual data in a Result Text column. All data must be accounted for in the file. You may include additional, appropriately titled Analysis columns to show intermediate steps used to calculate results. Please see below for examples.

Screening Positive: a Screening Positive column is required. Please note that the designations for a screening positive reflect your degree of confidence in the positive, not the potency of the positive. The preferred options

for the Positive column are S = Strong, M = Medium, or W = Weak. For wells considered as screening positives, enter S, M, or W. If it is not possible to separate your positives into S/M/W ranges, you may simply designate positive wells using Y for Yes. If a well is not considered a screening positive, leave the field blank.

If you are scoring for multiple phenotypes, create a separate column for each phenotype. For instance, if your screen seeks to describe both Enhancers and Inhibitors, create an "Enhancer" column and an "Inhibitor" column, and annotate both columns appropriately.

If you designate your screening positives using cut-off ranges, please provide the criteria you used to designate positives as W, M, or S. This information should be included in your Primary Screen Report. For example:

"Strong positives are those wells with a z-score ≥ 5. Medium positives are those wells with $5 >$ z-score ≥ 3. Weak positives are those wells with $3 >$ z-score ≥ 2. If replicates differ in their strength the final determination is based on the weaker scoring replicate."

If you have wells that fall into one of the screening positive cut-off ranges but are not annotated accordingly, please add a comment explaining why the well is excluded (see "Comments" below). **Note**: if the well is excluded because of a well-based problem or data issue, please enter "Y" in the Exclude column.

For high-content screens, add a Positive column for each phenotype measured (e.g., Positive Actin, Positive Mitosis) and add a Comment column for each phenotype measured (see Example 2 below).

Comments: in the Comment column, include any comments relevant to the well. Some examples of comments include the following:

"lower consistency (not a screening positive in all replicates)"
"lower strength"
"systematic errors on this plate"
"compromised cell integrity"
"out-of-focus image" "insoluble
compound"
"fluorescent compound"
well-based problems and irregularities, such as "bad mixing" or "well debris"

Sample Analysis Result Files

Figure 4.2: Example of intermediate analysis columns

Columns in template
Columns added by screener

Plate	Well	Type	Exclude	Intensity /cps_A	Intensity /cps_B	(FI)_A	(FI)_B	Result Numeric (FI)	Positive	Comment
1920	A01	N		9930	9790	1.032	1.022	1.027		
1920	A02	P		888	1020	0.125	0.152	0.138		
1920	A03	X		10012	9745	1.062	1.011	1.037		
1920	A04	X		8990	9540	0.965	0.979	0.972		
1920	A05	X	Y	4050	5210	0.452	0.512	0.482		Bad mixing in well
1920	A06	X		6650	6500	0.652	0.639	0.645	W	
1920	A07	X		2321	2120	0.225	0.214	0.220	S	

This example shows the first 7 rows of an analysis result file with two intermediary analyses (fluorescence intensity divided by plate median, or FI) columns, one for each replicate plate. A positive is a well whose FI average value is below a certain cut-off. The criteria for designating positives should be described in the accompanying Primary Screen Report. Note that control wells are included and that well A05 is excluded due to bad mixing in the well.

Figure 4.3: Example of numeric results

Columns in template
Columns added by screener

Plate	Well	Type	Exclude	RLU/counts_A	RLU/counts_B	Result Numeric (FI)	Positive	Comment
1649	A01	X		1600	1730	1.68		
1649	A02	X		20970	19890	21.84	S	
1649	A03	X		930	950	0.97		
1649	A04	X		15840	16220	16.50	M	
1649	A05	X		680	720	0.71		

This example shows the first 5 rows of an analysis result file with a numeric result. The plate median of the experimental wells is determined for replicates A and B. The fold-induction divided by the plate median is then calculated for each well. The Result Numeric is the average of the fold-induction for both replicates. All calculations should be described in the Primary Screen Report.

Figure 4.4: Example of visual phenotypic screen

Columns in template
Columns added by screener

Plate	Well	Type	Exclude	Positive, Kills Cells	Positive, Anti-Mitotic	Result Text Kills Cells	Result Text Anti-Mitotic
1661	A01	X		W		Many cells killed in one replicate	
1661	A02	X			M		Weak metaphase arrest
1661	A03	X	Y				Image slightly out of focus
1661	A04	X			S		Metaphase arrest
1661	A05	X					

This example shows the first 5 rows of an analysis result file for a phenotypic screen. There is a separate Screening Positive and Comment column for each phenotype monitored. Annotations are added in the Comment column. Well A03 is excluded.

Guidelines for Template Analysis

- **Do not delete any raw data from the file.** All raw data must be in the final analysis result file that you submit.

- **Do not delete any wells from the file.** Data and results must be returned for all wells screened, regardless of whether the well is considered a screening positive. Controls and wells that you want to exclude from your analysis should be included and annotated appropriately in the Exclude column.

- Do not insert rows before the column headings in the transformed file. Plate should be the first cell in the upper left corner of each worksheet.

- Sort the file first by plate, then by well. Please see the Excel Help menu for instructions on how to sort data.

- The Screening Positive and Comment columns should be the last two columns for each plate.

- The data for each plate should be contained in a separate worksheet, with the tabs for each worksheet titled with the appropriate plate number. The same columns should appear in the same order for all plates, and analysis methods and criteria for scoring positives should be consistent.

Step 5: Primary Screen Report

In addition to the annotated screen result file(s) containing data for all plates screened, a completed Primary Screen Report is required. This important document contains your assay protocol summary, information on controls, and a detailed text description of how numeric results were calculated and how to interpret them. If you determined your screening positives using result value cut-off ranges, you should define the ranges in the report. We suggest that you write this document as if you are describing your screen to an individual who is completely unfamiliar with your research. In this way, we hope that important information will not inadvertently be left out.

Please submit this report and your screen result data to the Data Analyst immediately after you have completed the analysis for your screen. Generally, this should be within 1 month of completing your primary screen. Cherry pick requests will NOT be completed until the required information is returned.

Important Contact Information:

Jennifer Smith, Ph.D.
Director of ICCB-L
617-432-5735
jennifer_smith@hms.harvard.edu

David Wrobel
Data Analyst
617-432-6266
david_wrobel@hms.harvard.edu

Jen Splaine
Data Curator
617-432-6266
jennifer_splaine@hms.harvard.edu

Screening Facility Staff
617-432-5815
iccb_screen@hms.harvard.edu