## P1 –Medicinal Chemist and Group Lead – Broad Institute

Link to session video <https://docs.google.com/file/d/0ByCluwBeg2GgaXl5NGZZMGVkck0/edit>

| Topic area | Comments |
| --- | --- |
| *Background* | P1 is a group leader for 8 Medicinal Chemists. The focus of his group’s work is hit validation and preliminary in-vitro optimization |
| *Tools participant uses for early stage drug discovery BioAssay research and how they use these tools.* | Goal of his work is to verify research other teams have done on particular compounds. His team replicates research in their own lab to verify findings. When undertaking a new project P1 might spend up to 2 full days doing preliminary online research on a compound. “My goal is to learn everything about a compound”.  His most frequent online/offline resources include:   * PubChem (makes heavy use of Kimball DB) * SciFinder (to search citations) * USPTO patent searches * Google (keywords vary by the compound in question) * Open literature – he uses a “berry picking” approach where he follows citations |
| *Frustrations with these tools* | While PubChem is his “go to” resource, it is time consuming to use. “I know the data is there but it is hard to find.” For example, when doing promiscuity analysis on a compound might want to search other assays that have used this compound or research results on related structures, or learn about work that has been done on small molecules for this target. Getting this full picture is a very tedious process.  Before embarking on new experiments “I want to know every single thing about this small molecule against this particular target”. |
| *Key learning from the participant’s walkthrough of the tool they use most often* | Note: due to technical difficulties we were unable to see P1’s screen during his walkthrough. The following are comments he made as he walked through is use of PubChem.   * Does not use SAR tables because the MLPCN compounds are diverse and not systematically created * “There should be 1 compound to 1 assay”. Right now the same assay might be used 10x and would appear as 10 unique entries. He has to manually sort through the results to group assays * Key questions include “what are the list of assays this list of compounds have come through? * P1 needs to be able to search a database via chemical structures and substructures * He downloads data for analysis within Excel * Wishes to combine assays |
| *How participant connects with colleagues* | P1 uses LinkedIn to connect with others in his line of work. He would like to receive an RSS alert when new information has been added on compounds/assays/other items of interest |
| *Ideal online Bio Assay research tool* | A tool with more chemistry functions |
| *Index to interesting video snippets* | 22:57 walks through a typical PubChem search (though you won’t see his screen unfortunately!)  27:43 describes how he uses data from search results to run an in-house validation  39:09 describes how he would do a small molecule search associated with a target |

PARTICIPANT’S PRIORITIZATION OF USER STORIES



## P2 – Biologist and group leader – Johns Hopkins

Link to session video <https://docs.google.com/file/d/0B-WTFEl6SqmsTTQxa25Sbmg5bkk/edit>

| Topic area | Comments |
| --- | --- |
| *Background* | P2 is a biologist and Director of Operations for a small research lab at Johns Hopkins. His team consists of post docs, students, and PhDs. The focus of his lab’s work is running screens for molecular libraries and reporting results. |
| *Tools participant uses for early stage drug discovery BioAssay research and how they use these tools.* | * Chem Card * DotMaddox * ISIS * OMEN (for genes) * Chem Bank * Spot Fire * US Patent Office * Google Patents * PubMed * PubChem   P2 used Chem Card and DotMaddox most often in his work, extracting data to remote access files that he can work with locally. When initiating new research, he performs a due diligence by running a literature search on PubMed to find out what is available. He will do a patent search on targets within the USPTO and then go to Google Patents to get an image. |
| *Frustrations with these tools* | Having to deal with non-standard terminology when doing searches. He searches targets by gene names, common names, and the “semi-trivial” name linked to a gene name. |
| *Key learning from the participant’s walkthrough of the tool they use most often* | We did not do a product walk through during this call. However, P2 spent a good deal of time discussing the user stories (queries). Please refer to the screenshot of the spreadsheet from his interview for details. |
| *How participant connects with colleagues* | P2 uses LinkedIn and belongs to multiple professional organizations such as the Society of General Physiology and the Biophysical ASPNET. |
| *Ideal online Bio Assay research tool* | He places a great deal of value in having a “decent export” 44:05. He would also like to maintain a viewing and an export template for a project. P2 also likes how the CID and SID identifiers have a pattern which allows you to analyze key characteristics |
| *Index to interesting video snippets* | 25:03 discusses user story 3 which concerns seeing a summary of a project/screening campaign  35:47 discusses user story 10 which concerns cloning an existing assay definition, assay registration. This function needs to be tightly controlled  40:27 discusses user story 14, ability to see all compounds that were active against a particular target  56:00 discusses output format for story 6 |

PARTICIPANT’S PRIORITIZATION OF USER STORIES



## P3 – Biologist and Group Leader – Broad Institute

Link to session video <https://docs.google.com/?authuser=0#home>

| Topic area | Comments |
| --- | --- |
| *Background* | P3’s group work is in the HTS group, work with MLP/Broad MLPCN library and a variety of assay types, most of which are cell-based: florescence, luminescence read outs on primary assays. Secondary assays are more diverse and often involve live imaging.  Currently working with the cholera virus. Up till this past January P3 worked mostly in the lab. Now his managerial duties take up most of his day. He now oversees the HTS lab and several PhDs. |
| *Tools participant uses for early stage drug discovery BioAssay research and how they use these tools.* | * Electronic notebooks to manage protocols and start assay in the Broad DB * GeneData Screener: Assay Analyzer (to examine and clean up raw data), Conversao (IC50 & AC50) * Spotfire (generate cherry pick lists – adds in new compounds) * CBIP * Seurat * PubChem * PubChem Promiscuity * Excel (small scale experiments)   He likes using Excel because it is easy to share and he knows how to use it. as he demoed various products he uses all the time, he mentioned that he likes to be able to produce files that he can download locally. |
| *Frustrations with these tools, and in general with research resources* | The new workflow using the above tools is much easier. PubChem. Would like IC50 curves to have a higher resolution, graphics should be large enough to place in a Powerpoint, and should be able to control how a structure is drawn. Also wants to go from structure graphic to an SAR table (22:00)  Also mentioned PubChem issue with needing to manually collapse experiments that use the same Assay ID. Would like to see a description for the assay ID. He also mentioned later on (37:00) about the confusion caused by non-standard nomenclature. There is a lot of turnover so it is sometimes hard to reach the research team who did the original work |
| *Key learning from the participant’s walkthrough of the tool they use most often* | P3 demoed the Pubchem Promiscuity tool, described how you can filter out by attributes such as luminescence. He uses this when he is doing HTS.  He also uses GeneData or Seurat because he can enter the SID/CID to retrieve all assays as well as activity levels over a particular assay. Can do a quick search within Seurat by typing in a Broad ID.  CBIT also allows him to retrieve assay data by entering a Broad ID |
| *How participant connects with colleagues* | He uses LinkedIn for work and Facebook in his personal life  Experiments with social media within Broad provide the opportunity to learn who else within the organization has an interest in a particular topic. When people move on, it is easier to reach people because they keep their contact information current within social media  Like’s PubChem’s feature of sending an alert to scientist when compound is retested |
| *Ideal online Bio Assay research tool* | Going from a particular compound to look at its behavior within other Assays, and being able to show compounds that show activities for a particular target, protein or phenotypic assay. Should also incorporate the pathway view.  Queries (stories) embody his needs. Substructure queries are important and being able to put in a list of compounds and see their activity. Should also have a general overview of the protocol to see how the assay was run with contact information. |
| *Index to interesting video snippets* | 14:34 demos use of Pubchem promiscuity  17:19 demos Seurat  18:23 demos CBIT  20:12 critiques PubChem Promiscuity IC50 curves and other values  27:40 his ideal Bio Assay features. |

PARTICIPANT’S PRIORITIZATION OF USER STORIES (will update with top 5)

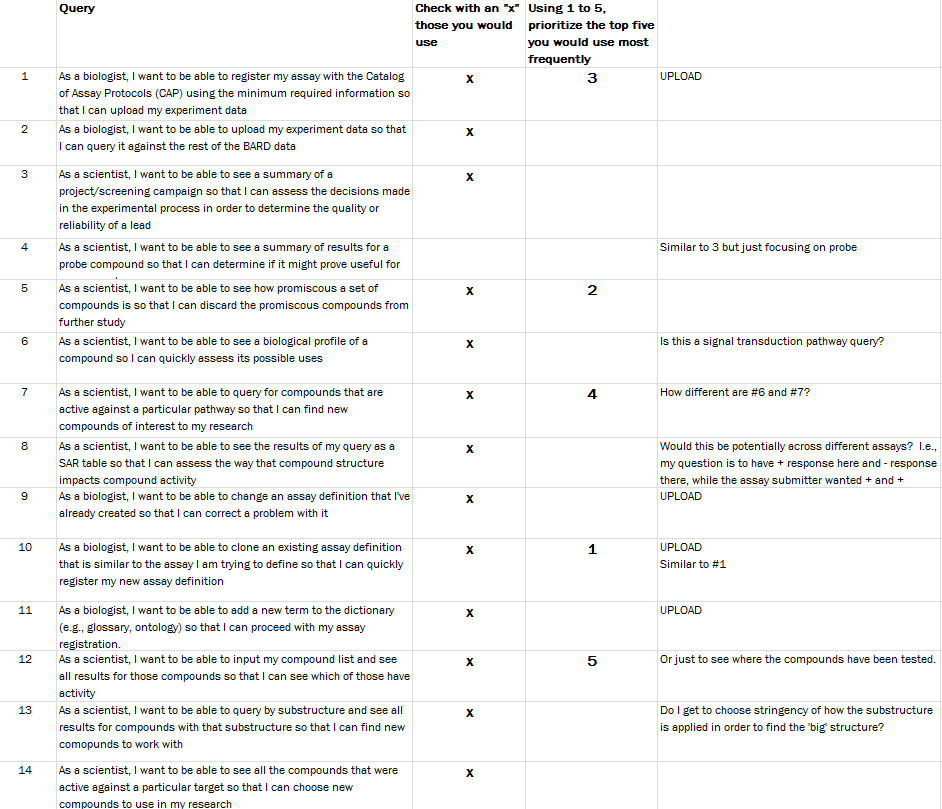


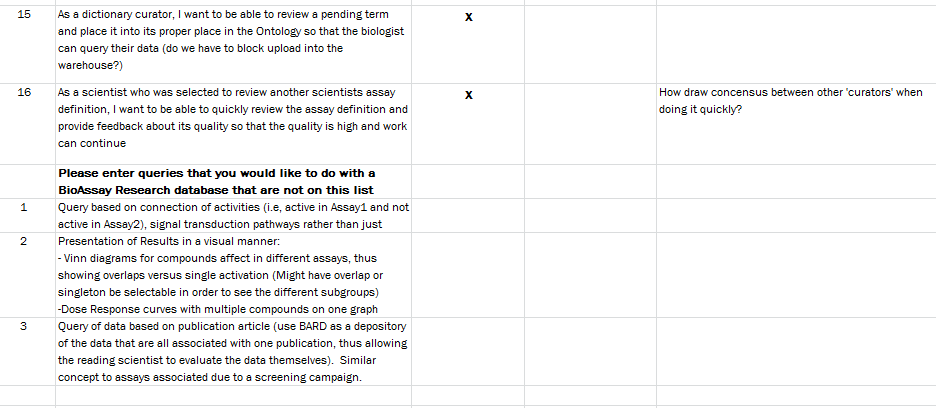
## P4 – BioInformatics Professional (uploads data to PubChem) UNM

Link to session video <https://docs.google.com/file/d/0B-WTFEl6SqmsQ05SY2NCeHMzZ0U/edit>

| Topic area | Comments |
| --- | --- |
| *Background* | P4 works in a specialty center that focuses on biologic screening. They work with assay providers to optimize assays and tweak smaller libraries so they can use them in a HTS. The do mostly fluorescence-based analysis. P4 pools data from different experimental stages to generate a probe report. She also assembles all the data into an Excel spreadsheet before uploading it to PubChem.  She has a PhD in chemical engineering but most of her work has been in biology. She does most of her work is done on the computer. |
| *Tools participant uses for early stage drug discovery BioAssay research and how they use these tools.* | * Excel (most widely used tool to analyze flat files to compare hit lists and create a rank ordering) * Prism (dose response curve, second mostly widely used) * Pipeline Pilot (visualization + basic analysis) * PubChem   P4’s work focuses on learning “which compounds have been hitting a lot of assay”. Her seeks to optimize the hits and remove false positives from the Assay results. She orders and cherry picks research results before placing it on PubChem.  She works with UNM Chem Informatics group to generate nascent SAR tables. The Chem Informatics group extracts dose response data from PubChem to build the tables. |
| *Frustrations with these tools, and in general with research resources* | “Each screen and assay presents itself with its own challenges”. Her chief frustrations arise from malfunctioning test equipment (vs. software systems) which causes media to spill over from one well to the other.  Later in the conversation (29:36) describes how she has to manually dig into the data to locate key details that explain different results (e.g., type of buffer used in the Assay)  They upload results to PubChem, but the owner is the assay provider who does not have the ability to understand and visualize the results of their data on PubChem. The providers come back to her asking her to help them visualize the data. They ask “can you make a graph of this” or show them “what kind of promiscuity analysis are there in the hits that come out of this assay, vs other assays that are out there”.  The providers want insight as well as data they can use in their own PowerPoint presentation The providers want comparisons across several assays, and provide citations to related work. She will manually generate results that give them insights. |
| *Key learning from the participant’s walkthrough of the tool they use most often* | P4 shows us how she manually prepares the Description File (an Excel spreadsheet “validation sheet” provided by PubChem) for upload to PubChem. She really likes having all this information in one spot. However, once uploaded if PubChem reviewers will ask her to make changes to the uploaded Description, the uploaded data is now in XML format so it needs to be reformatted back into CSV before she can share it with an Assay Provider.  Entering data can be tedious (e.g., entering mutant proteins & targets, dealing with many decimal points). She will merge data from experiments into the spreadsheet. If everything goes well this process takes about an hour. |
| *How participant connects with colleagues* | Uses LinkedIn, no Twitter. Belongs to several ListServs of groups who use the same brand of testing equipment. They help each other overcome glitches with these products. |
| *Ideal online Bio Assay research tool* | Would like the ability to reuse downloaded data for other purposes than the original goal. Would also like more visual displays of the data, in addition to the numbers “helps people to see that they are on the right track when they are querying the database”. While we were discussing her query prioritization (48:21) she talks about storing supporting data with a publication. |
| *Index to interesting video snippets* | 12:25 Background  23:00 Describes her use of SAR tables during a similarity search  24:52 Describes key data elements she seeks during a search (e.g., substructure elements, activity on targets)  28:00 Describes her ideal BioAssay research tool  30:51 Describes upload process  33:08 Sometimes it is difficult to isolate one target when putting together her Excel spreadsheet for data upload to PubChem  45:50 Wants to see connections between assays |

PARTICIPANT’S PRIORITIZATION OF USER STORIES



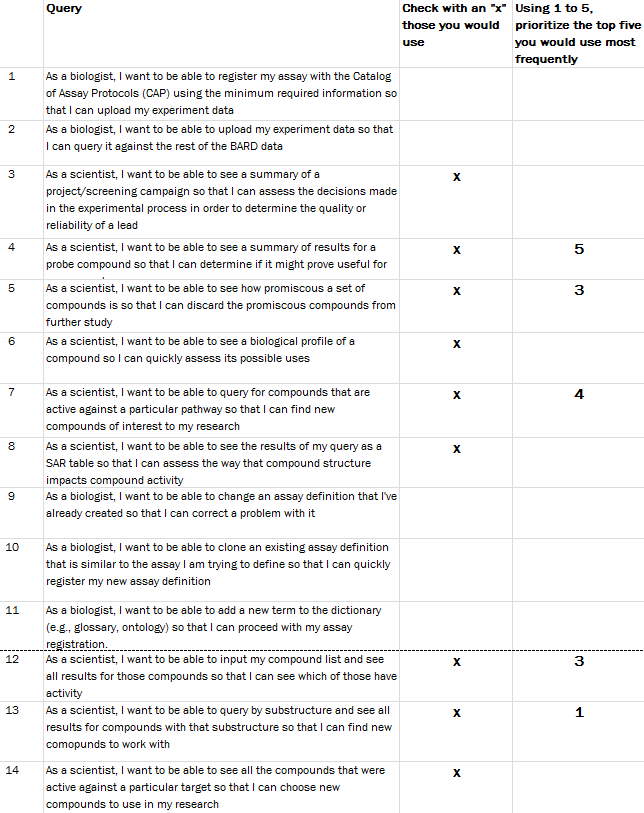


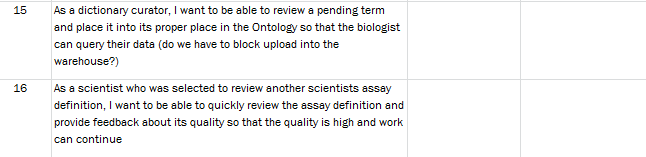
## P5 – Medicinal Chemist - Broad

Participant did not want us to video record his session

| Topic area | Comments |
| --- | --- |
| *Background* | P5 is a Medicinal Chemist who takes assay data from the biologists to develop probe compounds that others can use. At the beginning of the project he uses PubChem to do an exhaustive search for “prior art” compounds so he can pinpoint areas for optimization that become criteria for new projects. He uses PubChem later on in the process to validate compounds known to be promiscuous, so he can winnow to the best research candidtes. His team needs to prove it has produced a superior compound from what is in the PubChem database.  He spends about 60% of his time in the hood and the rest using computers. |
| *Tools participant uses for early stage drug discovery BioAssay research and how they use these tools.* | * Excel * SciFinder * Enterez * PubChem, PubMed * ISI Web of Knowledge * USPTO and the World IP Database * Reaxyx * Google and Google Scholar (text-based searches for prior art or chemical of interest has a standard name)   Like the other P’s we’ve spoken with P5 downloads data in Excel format for local analysis. PubChem has a limit of 100 compounds. |
| *Frustrations with these tools, and in general with research resources* | He does frequent multiple ADI searches. During his walk through of doing PubChem searches P5 noted many frustrations with common search sequences (see below). PubChem treats everything very generically |
| *Key learning from the participant’s walkthrough of the tool they use most often* | * Seems you need to know how to construct a Boolean search for more complex queries (e.g., AND-ing Assay IDs). He is not so comfortable with Boolean and unsure if he is getting the results he expects * Each search is very slowwwww, and he needs to do sequences of them to get to the data he really needs * He spent several minutes searching for a compound within the “BioAssay” tab (did not seem to notice this one was active) * Search on a compound produces 342 compounds with chemical information but it doesn’t tell you what the activities are in each assay. It just gives you the chemical structures * Search returns all the assays that every compound was screened in, which is not useful * He wants to find and compare compounds that exceed particular thresholds, and needs to see this activity by individual assay. He had to carry out several searches to get to this * P5 gets lost at the end when he couldn’t find the way to specify the particular assays he’s interested in and then do a union of the 3 * Another frustration is in addition to the SID and CID PubChem assigns a third ID that is unique to the institution that generated the compound, so you need to also do a search by institution to get complete results. Further confusion occurs when people end up entering the wrong ID into the wrong DB. |
| *How participant connects with colleagues* | Google Reader of RSS feeds, would like a watch list for new assays, compounds, etc |
| *Ideal online Bio Assay research tool* | As a chemist, primary assay data is less interesting to him. He wants to see Assay data that is run in dose. He wants to be able to segregate out to see just the info that he needs. Needs to see at a high level what “Active” means (e.g., within search results). Within search results he needs to look within each result that comes back active to learn the details about its activity. |
| *Index to interesting video snippets* | N/A |

**PARTICIPANT’S PRIORITIZATION OF USER STORIES**





## P6 – Director of Medicinal Chemistry – Sanford Burnham

Link to session video: <https://docs.google.com/file/d/0B-WTFEl6SqmsMWdCLUVpQnVnMEE/edit>

| Topic area | Comments |
| --- | --- |
| *Background* | Participant is a BARD product owner, member of the MLPCN, and Director of Medicinal Chemistry at Sanford Burnham. Prior to taking this position he worked in private industry for 10 years. His needs are very specific – P6 seeks to look at screening data to decide which compounds are good candidates for his team to pursue |
| *Tools participant uses for early stage drug discovery BioAssay research and how they use these tools.* | * Seurat * SciFinder (to cross check where else compound has been cited) * CBIT (internal DB to see list of internal compounds) * ChemDraw (for substructures) * PubChem |
| *Frustrations with these tools, and in general with research resources* | Having to manually go through PubChem to get his data. Takes 1 hour. He seems to avoid frustrations by keeping the list of candidate compounds small (10-20) as he goes to use PubChem.  PubChem gives you potency number, but you need to make many mouse clicks to get to curve. |
| *Key learning from the participant’s walkthrough of the tool they use most often* | * Winnows candidate compounds to 10 to 20 to make PubChem research less onerous, cross checks with SciFinder to see where else that compound has been discussed in other papers. * Shows us an internally-generated spreadsheet displaying dose/response curves. Works with biologists to get dose/response data. (spreadsheet usually also contains promiscuity analysis) * Key data is dose response curve, potency number, and picture of structure. Pays a lot of attention to reading curves. “subjective to some extent but not totally” * He will cross-reference the compound ID for promising looking compounds with CBIT to get the CID * He makes his list of compounds (10-20) and researches each within PubChem * Scrolls down to the BioAssay results. Looks first for promiscuity * Then reviews activity for compound, would like to drill down to see activity for each assay, but this is too cumbersome |
| *How participant connects with colleagues* | Uses LinkedIn. Having a public resource which offers notifications of things coming off embargo would be valuable. Currently he uses SciFinder for this information, but subscriptions to SciFinder are very expensive.  His experience in the private sector has made him very wary of sharing information. Chemists are different, because this is where the IP and value creation is… chemists are close to the vest in discussing the compounds we are working on” |
| *Ideal online Bio Assay research tool* | (forgot to ask him this question – have sent Anthony a note requesting this and will place this info here once received) |
| *Index to interesting video snippets* | 13:28 Walks us through his use of PubChem  15:05 Describes how he looks through a list of compounds within an Excel spreadsheet to pick most interesting compounds.  22:18 Can’t find biological activity, not sure why  27:00 Discusses feelings towards use of social media within his profession  32:00 Walks through many steps to get to the curve- which is always the goal. Promiscuity analysis is key for him |

**PARTICIPANT’S PRIORITIZATION OF USER STORIES**

