User stories from Krauthammer Lab at Yale., as captured after the 11/23/2010 meeting.

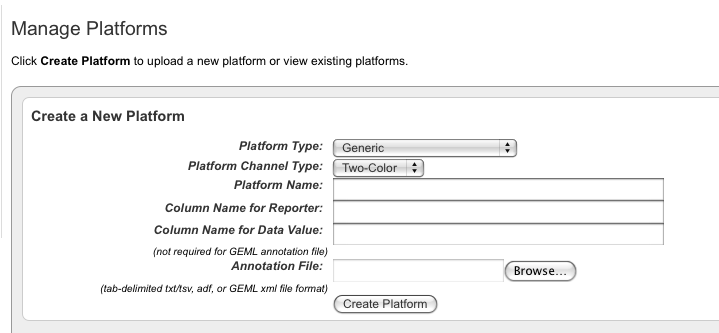
=====================================================

Priority – 1

User Story Title: Generic Data Type Support.

The basic idea here is to support gene-based genomic data that is not Gene Expression or Copy Number. They would like a generic way to specify platforms for, and the querying of, a generic platform and datatype. Their data is closely modeled to gene expression.

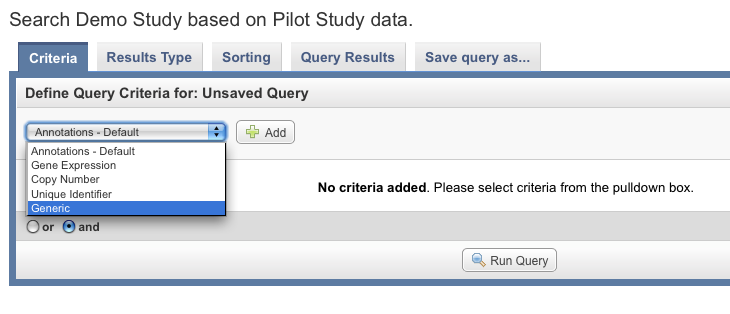
As a **platform manager** role, USBAT upload generic platform annotation file, and provide the datatype. (Current datatypes are copy number and gene expression.) USBAT see a new data type called “Generic”. Selecting Generic will present the user with an empty text box for providing a custom platform name. The file format will be the same as is currently used by Agilent Gene Expression. In addition to the file name to be uploaded, USBAT see a new field for providing the column name for the location of the platform’s reporter IDs. USBAT see a new field for providing the column name for the generic data type’s Data Value.

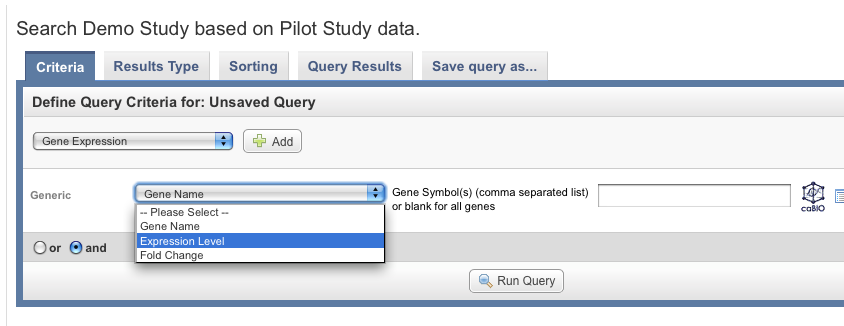


As a **study manager** role, USBAT select a platform of generic type on the Configure Genomic data source screen, similar to the way Gene Expression and Copy Number platforms are selected. USBAT select “Data Type = Generic” and “Vendor=Generic”. Selecting “Data Type = Generic” will display the list of custom platform names. The “Array Data Loading Type” will be “Supplemental Single” and “Supplemental Multiple.”

As a **study investigator** role, USBAT see a link in the left menu for “Data Plot”. This link will function very similarly to “Gene Expression Plot” with one key difference being that that on the page, the user will select which generic platform, they would like to plot.

As a study investigator role, USBAT see a new query criteria for the generic data type that they have added.





=========================================================

Priority – 2

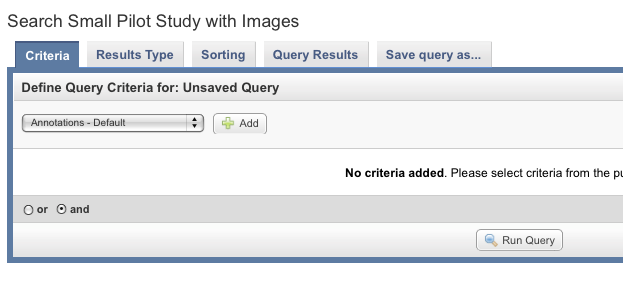
User Story Title: Annotation on sample data.

The basic idea here is allow the study manager to provide annotation data on each sample. This annotation data would then be queryable. Examples of annotation data are “disease state”,

As a study manager role, USBAT upload a file providing annotation for a sample ID. The file format will be similar to the subject annotation file that is currently used except that the sampleID will be used as the identifier. The process will work similarly to the way that image annotation works in that first a genomic source is configured and loaded, and second, the genomic sample annotations are added. If user selects the checkbox for “Create a new definition if one is not found” , and the definition does not exist for this annotation name, then one will be created automatically. The sample annotation name will be added to a default annotation group.

As a study manager role, USBAT group the sample annotation names using the existing Annotation Groups facility. USBAT see a new annotatio source called “sample”.

As a study investigator role, USBAT add sample annotation to a query by selecting the corresponding annotation group name which contains the sample annotations. (like is done at present).



Continuing the query creation, by selecting a sample annotation name and value, will then result in the only those matching samples being considered in the query.

Additional info captured after 12/15/2010 demonstration and meeting with KL.

1 - The users asked about the source for the mapping of the probe coordinates to the genome.  Presently, the information is obtained from caBIO by caIntegrator.  The feature request was to have a method for the user to upload these data from an external file.  The reason this was desired, was because the users had their own mapping information for genomic platforms that may not correspond with the resources in caBIO, for example with custom array platforms.

[JPM]  I thought their question was about being able to create a custom mapping from genes to position on the genome.  (not probe to position on the genome)  If they mean probe to position on the genome, then that comes from the platform design file that they can currently upload.  They have already created and uploaded custom platform files so I assume this will continue to work.

2- Enable queries of two platforms with the same data type.  For example, if samples from one subject were run on different gene expression platforms, the users would like to be able to query data from both platforms.  This could occur if the arrays were updated during the course of a long experimental program and the users wanted to look at results run on two array designs.  Presently, the UI prevents queries of two platforms within the same data type.  This could be a general requirement where samples from one subject were run on two copy number platforms, etc.

3- For caIntegrator, in any gene picker list window, display the number of genes in the list at the top of the table.

4-Is it possible to create a summary results section, where an advanced user can store information, final graphs, so that a novice user can just go to the study summary data?

[JPM]  This seemed to boil-down to a need for more collaboration and sharing features.  For any given study that they share with colleagues, they wanted to be able to send out a link to a page which answered "Here is the scientific point of this study and here is the chart/table which best illustrates the point".  Definitely need to refine their feedback before implementing anything.