BaseSpace® Correlation Engine API Overview
Connecting and accessing data and functionality in BaseSpace Correlation Engine.

Introduction
Illumina provides a suite of application programming interfaces (APIs) to its enterprise-level BaseSpace Correlation Engine customers. These APIs are designed to enable several processes:
- Data can be easily imported/exported.
- External applications can access BaseSpace Correlation Engine search results.
- BaseSpace Correlation Engine can link to external applications of interest.
- External applications can use the platform ontology and link to each individual result page.

BaseSpace Correlation Engine provides the following types of APIs:
- APIs to import data in bulk into the platform.
- APIs to perform auto-complete searches and tagging using platform ontology.
- APIs to search BaseSpace Correlation Engine programmatically.
- Mechanisms for linking into BaseSpace Correlation Engine.
- Mechanisms for linking out from BaseSpace Correlation Engine.
- API for exporting raw and processed data.

This technical note describes each of these APIs in more detail.

Bulk Import of Data into BaseSpace Correlation Engine
This API is useful for automated bulk import of enterprise data.
Data import is facilitated in two ways:
- First, Illumina provides the customer with a secure file transfer protocol (SFTP) site, from which all new data are automatically imported into BaseSpace Correlation Engine using BaseSpace APIs (Figure 1). An enterprise customer formats the data according to the BaseSpace Extensible Markup Language (XML) API standard and deposits it on the FTP site. Periodically, BaseSpace Correlation Engine polls the FTP site, imports the data, and deposits a summary document of the upload status. The FTP parameters, such as URL, username, password, are managed through an administrator page for the domain, which is accessible only to authorized personnel.
- Second, an enterprise customer can directly post the XML data to BaseSpace Correlation Engine using a secure hypertext transfer protocol (HTTPS) connection. For additional security, the enterprise can configure BaseSpace Correlation Engine to accept submissions from limited named IP addresses.

Data Format
BaseSpace Correlation Engine uses XML for all its API communications. An example of the study format (Figure 2) and bioset format (Figure 3) are provided. Each study consists of a description, metadata, associated files (attachments), and one or more biosets. A bioset consists of a description, metadata, associated tags, and a list of data.

Data Tagging
Data tags are used for data search, presentation, and correlations. More detailed documentation on data tagging is provided separately (see Tagging Cheat Sheet for BaseSpace Correlation Engine). The next section also covers the auto-complete API, which provides vocabulary keyword services for tagging or for use within any other internal application.

![Figure 1: BaseSpace Correlation Engine Data Import API — Schematic overview for the data import process](image)

For Research Use Only. Not for use in diagnostic procedures.
BaseSpace Correlation Engine Ontology Sources

BaseSpace Correlation Engine supports the following ontology sources for tagging:

- Gene Index (developed by Illumina)
- Disease (Systematized Nomenclature of Medicine [SNOMED] and Illumina internal)
- Anatomy — tissues/organs (Foundation Model of Anatomy [FMA] and Illumina internal)
- Compounds (PubChem, DrugBank, and other sources)
- Custom tags (see Tagging Cheat Sheet for the BaseSpace Correlation Engine)

XML Sample File

A sample XML file represents the BaseSpace Correlation Engine API data file format. A schema is also provided.

KEY


2  <nbs:description>
   <short>murine naive CD8+ T-cell samples for control or treated with 100ng/ml of IL-15 and/or IL-21 for 4 hours</short>

For Research Use Only. Not for use in diagnostic procedures.
<long>GSE2059 Status Public on Dec 11, 2004 Title IL-21/L-15 regulation of CD8 T-cells http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE2059 Organism(s) Mus musculus Type other Summary: murine naive CD8+ T-cell samples for control or treated with 100ng/ml of IL-15 and/or IL-21 for 4 hours Contributor(s) Zang R, Spolski R, Finkelstein SE, Oh S, Kovinen PE, Hintichs CS, Pise-Masison CA, Radonovich MF, Brady JN, Restifio NP, Berzofsky JA, Leonard WJ Submission date Dec 10, 2004 Contact name John Brady E-mail(s) bradyj@mail.nih.gov Phone 301 496 0986 Organization name NCI Lab LCO/VTB Street address City Bethesda State/province MD ZIP/Postal code 20892 Country USA Platforms (1) GPL339 Affymetrix GeneChip Mouse Expression Array 430A and Mouse Genome 430A 2.0 Array Samples (20)</long>
</nbs:description>

3
<nbs:contributor userId="?"
domainId="?">login@basespace.com</login>
</nbs:contributor>

4

5

6

7

<nbs:associatedFile>
<name><![CDATA[DenstyPlots_bioset_1.png]]></name>
</nbs:associatedFile>

</nbb:bioset>

8

9

<nbb:tag>
<tag>inflammatory response</tag>
</nbb:tag>

10

<nbb:description>
<short>Bioset_summary = GSE2059 murine naive CD8+ T-cells treated with 100ng/mL IL-15 and IL-21 vs untreated !Analysis_summary = Genes with statistically significant differences between murine naive CD8+ T-cells treated with 100ng/mL IL-15 and IL-21 for 4 hours vs untreated cells. Statistical tests — parametric test, variances not assumed equal (Welch t-test). P-value cutoff 0.05. 22111 probesets passed an initial filter requiring a minimal mean raw (MAS5) score of 2.7 (corresponding to the 5th percentile of all measurements) in at least one of the two conditions. Additional fold cutoff of 1.2 was applied to generate the final list of genes. !Change_description = CHANGING — murine naive CD8+ T-cells treated with 100ng/mL IL-15 and IL-21 vs untreated Test expression — average raw (MAS5) expression of murine naive CD8+ T-cells treated with 100ng/mL IL-15 and IL-21 Control expression —

average raw (MAS5) expression of untreated murine naive CD8+ T-cells</short>
</nbs:description>

11

12

For Research Use Only. Not for use in diagnostic procedures.
Key #1 — The study element type field indicates the type (GeneExpression, SequenceCentric, SNP) of the XML document. This is used to determine which sequence of Standard fields is used by the <ribb:dataset> element.

Key #2, 10 — The description element is used to provide descriptive text for use by both the study and the biosets. The short study description is shown whenever a study comes up in a query.

Key #3 — The contributor identifies the user with whom the import is associated. This element is required.

Key #4 — The study name uniquely identifies the study in the user’s domain. If one already exists with this name, it is updated (i.e., biosets are appended); otherwise, a new study is created.

Key #5 — The library element requires the ID provided to the customer by the BaseSpace Correlation Engine API. This element directs the data import to the target library accessible only by the customer.

Key #6 — The project element requires the ID provided to the customer by the BaseSpace Correlation Engine API within the library specified.

Key #7 — This element specifies any associated files (word-processing documents, presentations, images, etc.). Multiple instances of this element can appear in the document.

Key #8 — The bioset data type is required. Available data types can be found using the BaseSpace Correlation Engine API.

Key #9 — Tags can be generated using the Auto-Complete API. Tags that cannot be matched are ignored. The customer can add or change tags after import using the web user interface.

Key #10 — For GeneExpression type biosets, the best platform match is computed at import if the PlatformID = 0.

Key #12 — DataRow elements are repeated as often as needed. The fields available in the DataRow element depends on the type of study selected.

Notes:
The BaseSpace Correlation Engine provides additional services to look up IDs for various entities that are used in this API.

For additional information on BaseSpace API elements, use the BaseSpace Correlation Engine API document https://YOURDOMAIN.uscc.informatics.illumina.com/c/corp/queryAp

<row>
  <field name="speciesid">1</field>
</row>

Species ID Fetch Service Sample

Figure 3: Bioset Attributes — Solid lines indicate required fields, while dotted lines indicate optional fields. An enterprise client uses an API to return IDs corresponding to key fields (such as platform or species) to make sure of a precise database match. An empty field, such as a project without an ID, indicates creation of a new entity.

For Research Use Only. Not for use in diagnostic procedures.
<table>
<thead>
<tr>
<th>Platform ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>Affymetrix GeneChip Rat RAE230</td>
</tr>
<tr>
<td>20</td>
<td>Affymetrix GeneChip Rat RAE230A</td>
</tr>
<tr>
<td>22</td>
<td>Affymetrix GeneChip Rat RG-U34</td>
</tr>
<tr>
<td>21</td>
<td>Affymetrix GeneChip Rat RG-U34A</td>
</tr>
<tr>
<td>23</td>
<td>Affymetrix GeneChip Rat RN-U34</td>
</tr>
<tr>
<td>24</td>
<td>Agilent Human 1A G4110A</td>
</tr>
<tr>
<td>28</td>
<td>Agilent Human 1A G4110B</td>
</tr>
<tr>
<td>29</td>
<td>Agilent Human 1B G4111A</td>
</tr>
<tr>
<td>30</td>
<td>Agilent Human CGH 44A</td>
</tr>
<tr>
<td>31</td>
<td>Agilent Human CGH 44B</td>
</tr>
<tr>
<td>32</td>
<td>Agilent Human Whole Genome</td>
</tr>
<tr>
<td>33</td>
<td>Custom C. elegans</td>
</tr>
<tr>
<td>34</td>
<td>Custom D. melanogaster</td>
</tr>
<tr>
<td>35</td>
<td>GE Healthcare CodeLink ADME Rat 16-Assay</td>
</tr>
<tr>
<td>36</td>
<td>Custom Saccharomyces cerevisiae</td>
</tr>
<tr>
<td>37</td>
<td>GE Healthcare CodeLink Human Whole Genome</td>
</tr>
<tr>
<td>38</td>
<td>GE Healthcare CodeLink Human Whole Genome</td>
</tr>
<tr>
<td>39</td>
<td>Agilent Mouse G4121A</td>
</tr>
<tr>
<td>40</td>
<td>Agilent Mouse Whole Genome</td>
</tr>
<tr>
<td>41</td>
<td>Agilent Mouse Development G4130A</td>
</tr>
<tr>
<td>42</td>
<td>Agilent Mouse Development G4130B</td>
</tr>
<tr>
<td>43</td>
<td>Agilent Mouse Development G4131A</td>
</tr>
<tr>
<td>44</td>
<td>Agilent Mouse Development G4131B</td>
</tr>
<tr>
<td>45</td>
<td>Agilent Mouse Development G4132A</td>
</tr>
<tr>
<td>46</td>
<td>Agilent Mouse Development G4132B</td>
</tr>
</tbody>
</table>
<field name="name">GE Healthcare CodeLink Mouse</field>
</row>
</row>
</row>
</row>
<row>
<field name="platformid">50</field>
<field name="name">GE Healthcare CodeLink Rat Whole Genome</field>
</row>
<row>
<field name="platformid">44</field>
<field name="name">GE Healthcare CodeLink UniSet Human 20K</field>
</row>
<row>
<field name="platformid">45</field>
<field name="name">GE Healthcare CodeLink UniSet Human 1K</field>
</row>
<row>
<field name="platformid">47</field>
<field name="name">GE Healthcare CodeLink UniSet Mouse 20K</field>
</row>
<row>
<field name="platformid">48</field>
<field name="name">GE Healthcare CodeLink UniSet Mouse 1K</field>
</row>
<row>
<field name="platformid">51</field>
<field name="name">GE Healthcare CodeLink UniSet Rat 1K</field>
</row>
<row>
<field name="platformid">59</field>
<field name="name">Illumina Sentrix Human-6 BeadChip</field>
</row>
<row>
<field name="platformid">60</field>
<field name="name">Illumina Sentrix Human-6 v2 BeadChip</field>
</row>
<row>
<field name="platformid">69</field>
<field name="name">Illumina Sentrix Human-6 v2 (Target ID) BeadChip</field>
</row>
<row>
<field name="platformid">61</field>
<field name="name">Illumina Sentrix Human-8 BeadChip</field>
</row>
<row>
<field name="platformid">63</field>
<field name="name">Illumina Sentrix Human-8 V2 BeadChip</field>
</row>
<row>
<field name="platformid">70</field>
<field name="name">Illumina Sentrix Human-8 V2 (Target ID) BeadChip</field>
</row>
<row>
<field name="platformid">62</field>
<field name="name">Illumina Sentrix Human-8RD BeadChip</field>
</row>
<row>
<field name="platformid">65</field>
</row>
<row>
<field name="platformid">72</field>
<field name="name">Illumina Sentrix Mouse-6 v1 (Target ID) BeadChip</field>
</row>
<row>
<field name="platformid">64</field>
<field name="name">Illumina Sentrix Mouse-6 v1_1 (Target ID) BeadChip</field>
</row>
<row>
<field name="platformid">71</field>
<field name="name">Illumina Sentrix Mouse-6 v1_1 BeadChip</field>
</row>
<row>
<field name="platformid">67</field>
<field name="name">Illumina Sentrix Mouse-8 v1 BeadChip</field>
</row>
<row>
<field name="platformid">74</field>
<field name="name">Illumina Sentrix Mouse-8 v1 (Target ID) BeadChip</field>
</row>
<row>
<field name="platformid">66</field>
<field name="name">Illumina Sentrix Mouse-8 v1_1 BeadChip</field>
</row>
<row>
<field name="platformid">73</field>
<field name="name">Illumina Sentrix Mouse-8 v1_1 (Target ID) BeadChip</field>
</row>
<row>
<field name="platformid">68</field>
<field name="name">Illumina Sentrix Rat-12 v1 BeadChip</field>
</row>
<row>
<field name="platformid">75</field>
<field name="name">Illumina Sentrix Rat-12 v1 (Target ID) BeadChip</field>
</row>
<row>
<field name="platformid">58</field>
<field name="name">Trex IAS_RBG v3 BeadChip</field>
</row>
</ROOT>