

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.

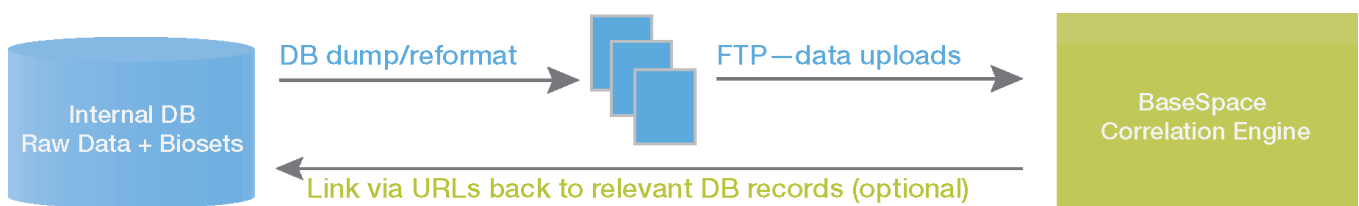


Figure 1: BaseSpace Correlation Engine Data Import API—Schematic overview for the data import process

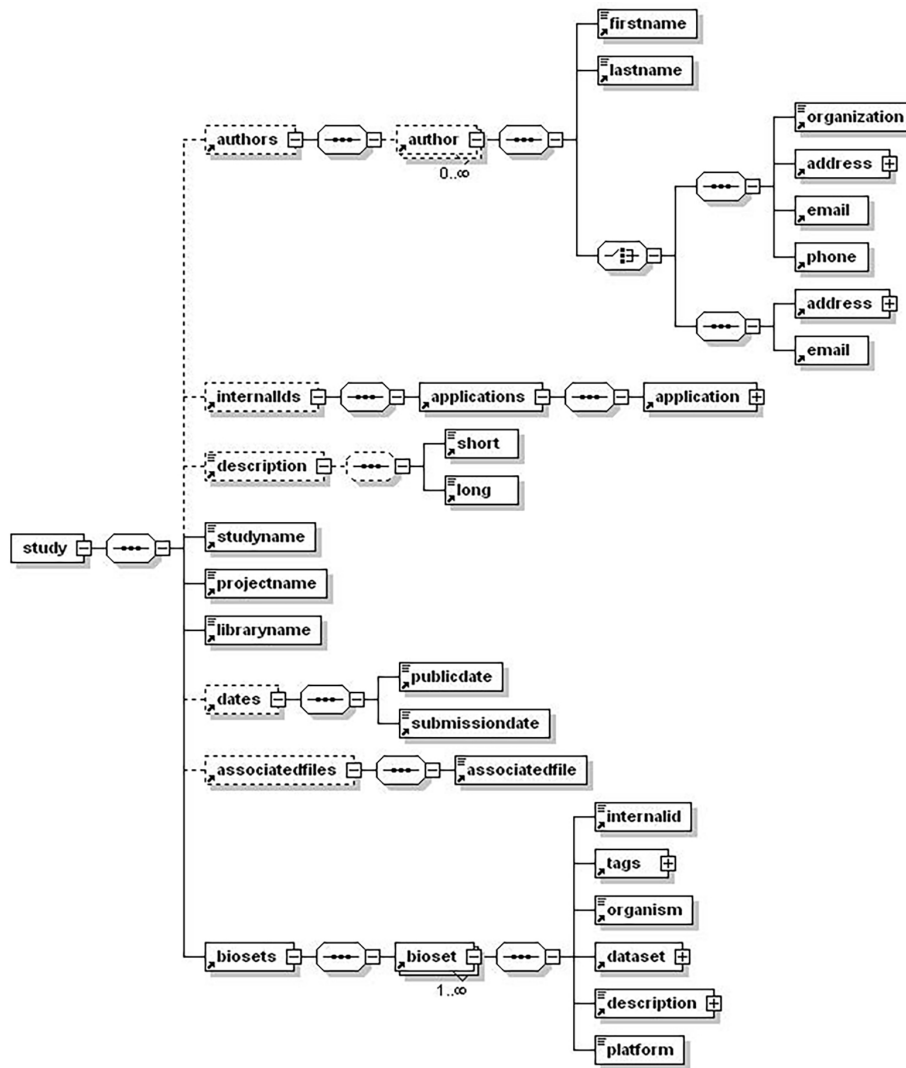


Figure 2: Study Attributes—Solid lines indicate required fields, while dotted lines indicate optional fields. An enterprise client uses an API to return IDs corresponding to key fields (such as platform or species) to make sure of a precise database match. An empty field, such as a project without an ID, indicates creation of a new entity.

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

```

1
  <nbs:study type="GeneExpression"
    xmlns:nbs="http://www.basespace.com/bioframework/schemas
/study_v2.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
    xmlns:nbb="http://www.basespace.com/bioframework/schemas/bio
set_v2.0"
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>

```

```

<long>GSE2059 Status Public on Dec 11, 2004 Title IL-21/IL-15
regulation of CD8 T-cells
http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE2
059 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) Zeng R,
Spolski R, Finkelstein SE, Oh S, Kovanen PE, Hinrichs CS, Pise-
Masison CA, Radonovich MF, Brady JN, Restifo 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 useri="?"
domainid="?"><login>example@basespace.com</login>
</nbs:contributor>
4
<studyName>IL-15 and IL-21 regulation of CD8 T-cells
</studyName>
5
<library id="1">BaseSpace Library</library>
6
<project id="1">Immune response and infection</project>
<nbs:dates>
<submissionDate>2008-01-01-08:00</submissionDate>
</nbs:dates>
7
<nbs:associatedFile> <name><![CDATA[DensityPlots_bioset_
1.png]]></name> </nbs:associatedFile>
<nbb:bioset
xmlns:nbb="http://www.basespace.com/bioframework/schemas/bio
set_v2.0"
xmlns:nbs="http://www.basespace.com/bioframework/schemas/stu
dy_v2.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
</nbb:bioset >
8
<biosetTypeid>1</biosetTypeid >
9
<nbb:tags>
<tag>inflammatory response</tag>
<tag>interleukin binding</tag>
<tag>IL15</tag>
<tag>IL21</tag>
</nbb:tags>
10
<nbs: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>
<long xsi:nil="true"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" />
</nbs:description>
<species id="2">Mus musculus</species>
11
<platform id="14">Affymetrix GeneChip Mouse MG_430A
2.0</platform>
<biosetName>CD8+ T-cells + IL-15 and IL-21 4hr_CHGN vs
control</biosetName>
<nbb:dataset>
<CustomFieldName1>test expression</CustomFieldName1>
<CustomFieldName2>control expression</CustomFieldName2>
12
<nbb:dataRow >
<standard><genename>19305</genename> <foldChange>-
1.23</foldChange> <pvalue>0.041438438</pvalue> </standard>
<custom><col1>127.964</col1> <col2>157.695</col2>
</custom>
</nbb:dataRow>
...
</nbb:dataset>
</nbb:bioset>
</nbb:biosets>
</nbs:study>

```

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 <nbb: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 (ie, 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 #11 — 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.ussc.informatics.illumina.com/c/corp/queryAp i.nb> (replace YOURDOMAIN with the customer’s BaseSpace domain name).

An XML result file indicates the status (success or failure) of the import. It also provides a list of any new IDs created.

Species ID Fetch Service Sample

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ROOT SYSTEM "species.dtd">
<ROOT xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
  <row>
    <field name="speciesid">1</field>
```

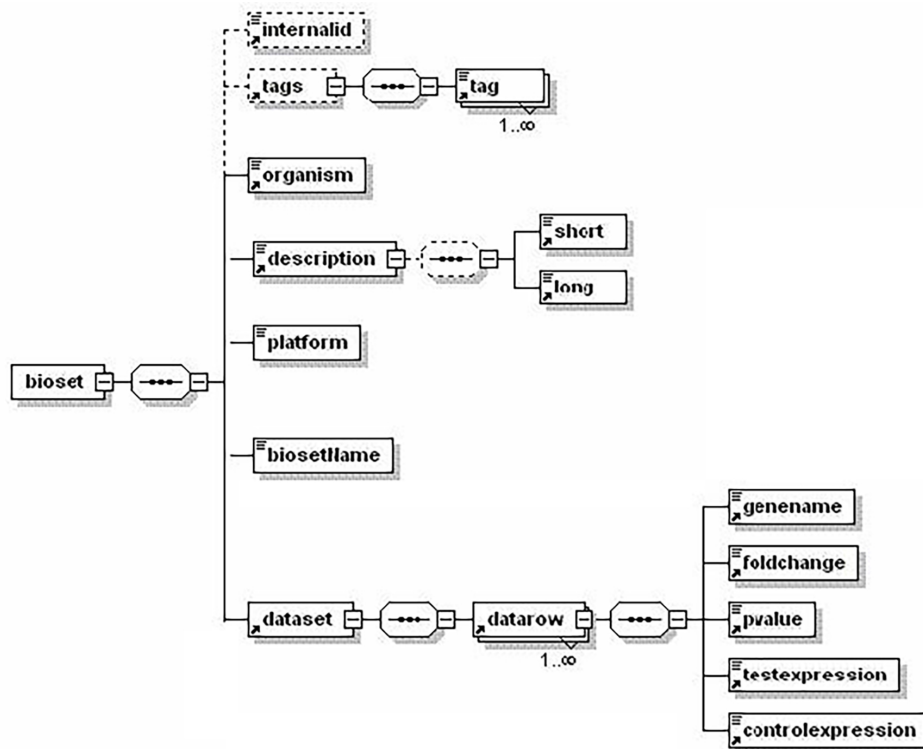


Figure 3: Bioset Attributes — Solid lines indicate required fields, while dotted lines indicate optional fields. An enterprise client uses an API to return IDs corresponding to key fields (such as platform or species) to make sure of a precise database match. An empty field, such as a project without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

```

    <field name="name">Homo sapiens</field>
</row>
<row>
    <field name="speciesid">2</field>
    <field name="name">Mus musculus</field>
</row>
<row>
    <field name="speciesid">3</field>
    <field name="name">Rattus norvegicus</field>
</row>
<row>
    <field name="speciesid">4</field>
    <field name="name">Arabidopsis thaliana</field>
</row>
<row>
    <field name="speciesid">5</field>
    <field name="name">C. elegans</field>
</row>
<row>
    <field name="speciesid">6</field>
    <field name="name">D. melanogaster</field>
</row>
<row>
    <field name="speciesid">7</field>
    <field name="name">Saccharomyces cerevisiae</field>
</row>
</ROOT>

```

Platform ID Fetch Service Sample

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ROOT SYSTEM "platform.dtd">
<ROOT xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
    <row>
        <field name="platformid">27</field>
        <field name="name">Affymetrix GeneChip C. elegans
Genome Array</field>
    </row>
    <row>
        <field name="platformid">25</field>
        <field name="name">Affymetrix GeneChip Drosophila
Genome 1.0 Array</field>
    </row>
    <row>
        <field name="platformid">26</field>
        <field name="name">Affymetrix GeneChip Drosophila
Genome 2.0 Array</field>
    </row>
    <row>
        <field name="platformid">1</field>
        <field name="name">Affymetrix GeneChip Human 35k
</field>
    </row>
    <row>
        <field name="platformid">2</field>
        <field name="name">Affymetrix GeneChip Human
Cancer HC-G110</field>
    </row>
    <row>
        <field name="platformid">4</field>
        <field name="name">Affymetrix GeneChip Human Full
Length HuGeneFL</field>
    </row>
    <row>
        <field name="platformid">5</field>
        <field name="name">Affymetrix GeneChip Human HG-
Focus Target Array</field>
    </row>
</row>

```

```

    <field name="platformid">7</field>
    <field name="name">Affymetrix GeneChip Human HG_
U133 Plus 2.0</field>
</row>
<row>
    <field name="platformid">8</field>
    <field name="name">Affymetrix GeneChip Human HG_
U133 Set</field>
</row>
<row>
    <field name="platformid">6</field>
    <field name="name">Affymetrix GeneChip Human HG_
U133A version [1 or 2]</field>
</row>
<row>
    <field name="platformid">10</field>
    <field name="name">Affymetrix GeneChip Human HG_
U95A version [1 or 2]</field>
</row>
<row>
    <field name="platformid">9</field>
    <field name="name">Affymetrix GeneChip Human HG_
U95A-E version [1 or 2]</field>
</row>
<row>
    <field name="platformid">3</field>
    <field name="name">Affymetrix GeneChip Human
Muscle Chip</field>
</row>
<row>
    <field name="platformid">76</field>
    <field name="name">Affymetrix GeneChip Human X3P
</field>
</row>
<row>
    <field name="platformid">12</field>
    <field name="name">Affymetrix GeneChip Mouse
GNF1M</field>
</row>
<row>
    <field name="platformid">13</field>
    <field name="name">Affymetrix GeneChip Mouse MG_
430 2.0</field>
</row>
<row>
    <field name="platformid">14</field>
    <field name="name">Affymetrix GeneChip Mouse MG_
430A 2.0</field>
</row>
<row>
    <field name="platformid">15</field>
    <field name="name">Affymetrix GeneChip Mouse MG_
6500A</field>
</row>
<row>
    <field name="platformid">17</field>
    <field name="name">Affymetrix GeneChip Mouse MG_
U74</field>
</row>
<row>
    <field name="platformid">16</field>
    <field name="name">Affymetrix GeneChip Mouse MG_
U74A</field>
</row>
<row>
    <field name="platformid">18</field>
    <field name="name">Affymetrix GeneChip Mouse
Mu11K</field>
</row>

```

<pre> <row> <field name="platformid">19</field> <field name="name">Affymetrix GeneChip Rat RAE230 2.0</field> </row> <row> <field name="platformid">20</field> <field name="name">Affymetrix GeneChip Rat RAE230A</field> </row> <row> <field name="platformid">22</field> <field name="name">Affymetrix GeneChip Rat RG-U34 </field> </row> <row> <field name="platformid">21</field> <field name="name">Affymetrix GeneChip Rat RG- U34A</field> </row> <row> <field name="platformid">23</field> <field name="name">Affymetrix GeneChip Rat RN-U34 </field> </row> <row> <field name="platformid">24</field> <field name="name">Affymetrix GeneChip Rat RT-U34 </field> </row> <row> <field name="platformid">28</field> <field name="name">Affymetrix GeneChip Yeast Genome S98 Array YG-S98</field> </row> <row> <field name="platformid">29</field> <field name="name">Agilent Human 1A G4110A</field> </row> <row> <field name="platformid">30</field> <field name="name">Agilent Human 1A V2 G4110B </field> </row> <row> <field name="platformid">31</field> <field name="name">Agilent Human 1B G4111A</field> </row> <row> <field name="platformid">32</field> <field name="name">Agilent Human CGH 44A G4410A</field> </row> <row> <field name="platformid">33</field> <field name="name">Agilent Human CGH 44B G4410B</field> </row> <row> <field name="platformid">34</field> <field name="name">Agilent Human Whole Genome G4112A</field> </row> <row> <field name="platformid">35</field> <field name="name">Agilent Mouse CGH 44A G4414A </field> </row> <row> </pre>	<pre> <field name="platformid">36</field> <field name="name">Agilent Mouse Development 44K G2519A</field> </row> <row> <field name="platformid">37</field> <field name="name">Agilent Mouse Development G4120A</field> </row> <row> <field name="platformid">38</field> <field name="name">Agilent Mouse G4121A</field> </row> <row> <field name="platformid">39</field> <field name="name">Agilent Mouse Whole Genome G4122A</field> </row> <row> <field name="platformid">40</field> <field name="name">Agilent Rat G4130A</field> </row> <row> <field name="platformid">41</field> <field name="name">Agilent Rat V2 G4130B</field> </row> <row> <field name="platformid">42</field> <field name="name">Agilent Rat Whole Genome G4131A</field> </row> <row> <field name="platformid">56</field> <field name="name">Custom C. elegans</field> </row> <row> <field name="platformid">55</field> <field name="name">Custom D. melanogaster</field> </row> <row> <field name="platformid">52</field> <field name="name">Custom Human</field> </row> <row> <field name="platformid">53</field> <field name="name">Custom Mouse</field> </row> <row> <field name="platformid">54</field> <field name="name">Custom Rat</field> </row> <row> <field name="platformid">57</field> <field name="name">Custom Saccharomyces cerevisiae</field> </row> <row> <field name="platformid">49</field> <field name="name">GE Healthcare CodeLink ADME Rat 16-Assay</field> </row> <row> <field name="platformid">43</field> <field name="name">GE Healthcare CodeLink Human Whole Genome</field> </row> <row> <field name="platformid">46</field> </pre>
--	---

```

    <field name="name">GE Healthcare CodeLink Mouse
Whole Genome</field>
  </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 1</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 1</field>
  </row>
  <row>
    <field name="platformid">51</field>
    <field name="name">GE Healthcare CodeLink UniSet Rat
1</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</field>
  </row>
  <row>
    <field name="platformid">69</field>
    <field name="name">Illumina Sentrix Human-6 v2 (Target
ID)</field>
  </row>
  <row>
    <field name="platformid">61</field>
    <field name="name">Illumina Sentrix HumanRef-8
BeadChip</field>
  </row>
  <row>
    <field name="platformid">63</field>
    <field name="name">Illumina Sentrix HumanRef-8 V2
  </field>
  </row>
  <row>
    <field name="platformid">70</field>
    <field name="name">Illumina Sentrix HumanRef-8 V2
(Target ID)</field>
  </row>
  <row>
    <field name="platformid">62</field>
    <field name="name">Illumina Sentrix HumanRef-8RD
BeadChip</field>
  </row>
  <row>
    <field name="platformid">65</field>
    <field name="name">Illumina Sentrix Mouse-6 v1</field>
  </row>
  <row>
    <field name="platformid">72</field>
    <field name="name">Illumina Sentrix Mouse-6 v1 (Target
ID)</field>
  </row>
  <row>
    <field name="platformid">64</field>
    <field name="name">Illumina Sentrix Mouse-6 v1_1
  </row>
  <row>
    <field name="platformid">71</field>
    <field name="name">Illumina Sentrix Mouse-6 v1_1
(Target ID)</field>
  </row>
  <row>
    <field name="platformid">67</field>
    <field name="name">Illumina Sentrix MouseRef-8 v1
  </field>
  </row>
  <row>
    <field name="platformid">74</field>
    <field name="name">Illumina Sentrix MouseRef-8 v1
(Target ID)</field>
  </row>
  <row>
    <field name="platformid">66</field>
    <field name="name">Illumina Sentrix MouseRef-8 v1_1
  </field>
  </row>
  <row>
    <field name="platformid">73</field>
    <field name="name">Illumina Sentrix MouseRef-8 v1_1
(Target ID)</field>
  </row>
  <row>
    <field name="platformid">68</field>
    <field name="name">Illumina Sentrix RatRef-12 v1
  </field>
  </row>
  <row>
    <field name="platformid">75</field>
    <field name="name">Illumina Sentrix RatRef-12 v1
(Target ID)</field>
  </row>
  <row>
    <field name="platformid">58</field>
    <field name="name">Trex_IAS_RGI3_v3</field>
  </row>
</ROOT>

```

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.

© 2017 Illumina, Inc. All rights reserved. Illumina, BaseSpace, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners. Pub. No. 970-2014-026-B

