

# BaseSpace<sup>™</sup> Correlation Engine

Accelerate biological insights  
from genomic data

- Comprehensive tool for determining biological context based on a growing library of curated data sets
- Easy-to-use graphical user interface
- Proprietary algorithm automates and simplifies correlation analysis



## Introduction

All scientists need to put the results of their experiments into biological context. This is commonly achieved through literature-based searches, whether that be through PubMed, Google, or pathway-based tools. A drawback to these approaches is that only a small percentage of the data is included in a scientific paper.

To overcome this challenge, Illumina developed BaseSpace Correlation Engine, one of the largest biological databases in the world. Illumina spent over a decade curating raw data from whole-genome studies, normalizing it across platforms, and ingesting it into Correlation Engine.

As a result, Correlation Engine routinely identifies hundreds to thousands of studies featuring thousands of genes that have few or no search results in PubMed. Correlation Engine provides life science researchers with unprecedented access to vast numbers of high-quality whole-genome analyses and insightful scientific tools (ie, Body Atlas, Disease Atlas, Pharmaco Atlas, Knockdown Atlas, Genetic Markers, Meta-analysis).

A simple intuitive graphical interface (Figure 1) takes advantage of continuously expanding content and enables researchers to identify novel correlations with ease and efficiency. Because it is data driven, scientists are more likely to discover novel associations and find results that would be missed in a simple literature scan.

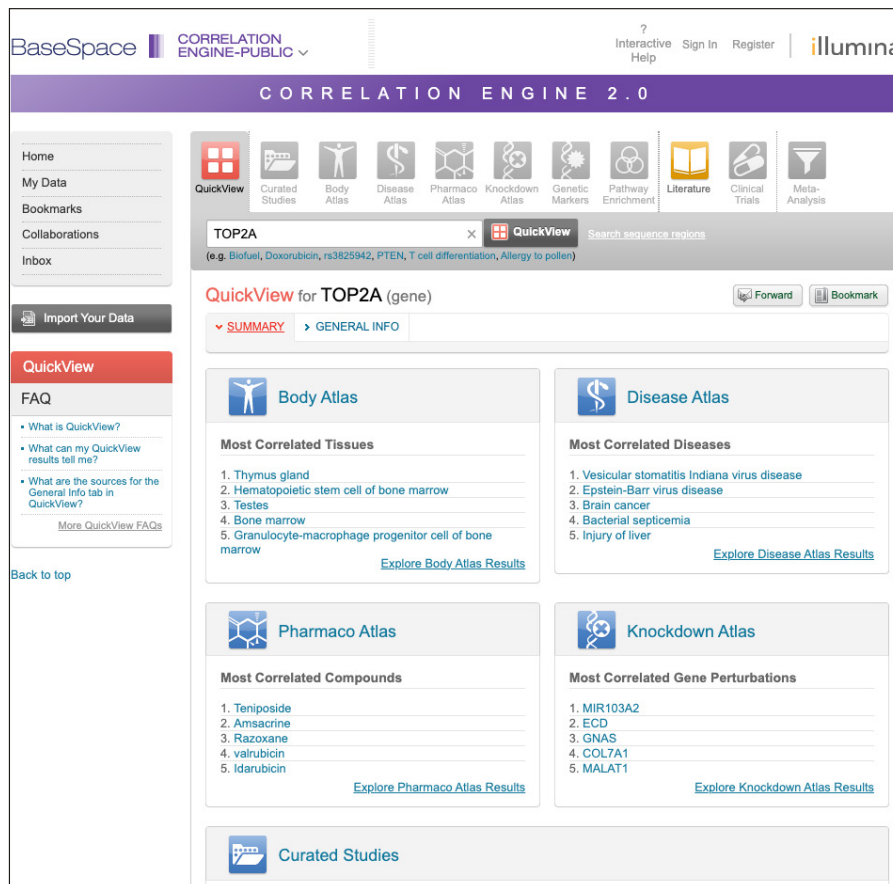


Figure 1: BaseSpace Correlation Engine user interface enables queries for numerous association types—Novel correlations and associations are quickly identified for a given query, revealing data-driven connections between genes, diseases, compounds, tissues, pathways and literature.

## Comprehensive platform

Adaptive learning processes take advantage of weekly updated content from public and proprietary data. Correlation Engine computes ranked association scores for tissues, diseases, compounds, and genetic perturbations. The content is standardized using accredited ontologies to create a platform of genomic studies covering more than 10,000 disease/phenotype, tissue, and compound concepts.

Many types of genomic studies are included in results, such as mRNA expression, miRNA expression, somatic mutations, copy number changes, DNA methylation, protein-DNA binding, histone modifications, and genome-wide association studies (GWAS). The rank-based enrichment algorithms make the framework agnostic of the technology platforms on which the genomic data were generated. This allows cross-analysis of data obtained from different methods, such as next-generation sequencing (NGS) and microarrays.

The platform is powered by interspecies comparisons as the framework has built-in ortholog mapping across 13 species (Figure 2). Researchers can compare and harness information to derive biological context from experimental results of human, mouse, rat, and other model organisms.

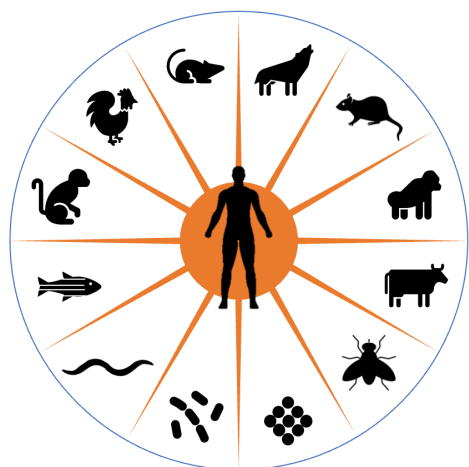
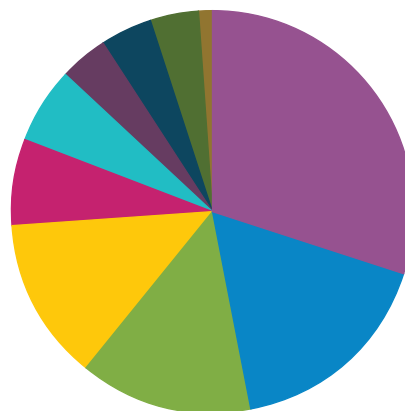


Figure 2: BaseSpace Correlation Engine maps orthologous data across 13 species—Genomic content in BaseSpace Correlation Engine is searched by ortholog gene names, synonyms, and features from NGS and array studies to provide comprehensive results.

## Mechanisms of disease

By comparing disease profiles across animal models, cohorts, and disease stages. Correlation Engine enables users to assess the pathways that play significant roles in disease development across multiple studies and data types. Correlation Engine contains over 135,000 analyses derived from standardized processing of more than 22,000 genomic studies spanning diverse diseases (Figure 3) from major public repositories such as Gene Expression Omnibus (GEO), Array Express, European Molecular Biology Laboratory (EMBL), Stanford Microarray Database (SMD), Encyclopedia of DNA Elements (ENCODE), Cancer Cell Line Encyclopedia (CCLE), the GenotypeTissue Expression (GTEx) project, and more.



- Oncology
- Stem cells, differentiation, and development
- Metabolic disorders and obesity
- Immune response and infection
- Neurological disorders
- Injury and inflammation
- Mutation and phenotypic data
- Pharmacology and toxicology
- Cardiovascular disorders
- Atlas—Organs, tissues, and cell types

Figure 3: BaseSpace Correlation Engine curated genomic studies by disease area—With over 135,000 analyses derived using a standardized pipeline from more than half a million samples, the BaseSpace Correlation Engine content is constantly growing. Data-driven analysis allows target assessment and validation, biomarker discovery, drug repositioning, etc.

## Mechanisms of gene function

Correlation Engine enables scientists to gain insights about where a gene is expressed in the Body Atlas and how a gene functions from billions of data points covering close to 5500 diseases across major disease areas and nearly 10,000 unique genetic perturbations.

## Mechanisms of drug action

In Correlation Engine, more than 50,000 analyses related to more than 4500 compounds exist in the system. Researchers can analyze proprietary candidate molecules for on-target mechanisms and toxicity profiles and compare the profiles with those of other compounds.

## From data to insights

BaseSpace Correlation Engine enables novel insights by interrogating billions of data points derived from standardized analyses of whole-genome studies. The public data available in Correlation Engine is just the starting point for discovery. Users can securely upload their data and query it against itself or against public data. Enterprise account holders can share results within their private domain, and add results to meta-analysis applications for generation of unique correlations. Private data is inaccessible across enterprise domains and results are kept safe and private in an ISO27001, SOC1, SOC2, SOC3, PCIDSS certified environment.

## Learn more

[BaseSpace Correlation Engine](#)

[Free trial](#)

## Ordering

[Contact us](#)

Special academic pricing is available



1.800.809.4566 toll-free (US) | +1.858.202.4566 tel  
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners.  
For specific trademark information, see [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).  
M-GL-02840 v1.0