# BaseSpace Correlation Engine

Bringing big data and biology together with a comprehenisve tool for mining genomic data.

## 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 (formerly known as NextBio™ Research), 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 into BaseSpace Engine. As a result, BaseSpace Engine routinely identifies hundreds to thousands of studies for thousands of genes that have few or no search results in PubMed. BaseSpace 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) was designed to take advantage of continuously expanding content and enable 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.

## Comprehensive platform

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

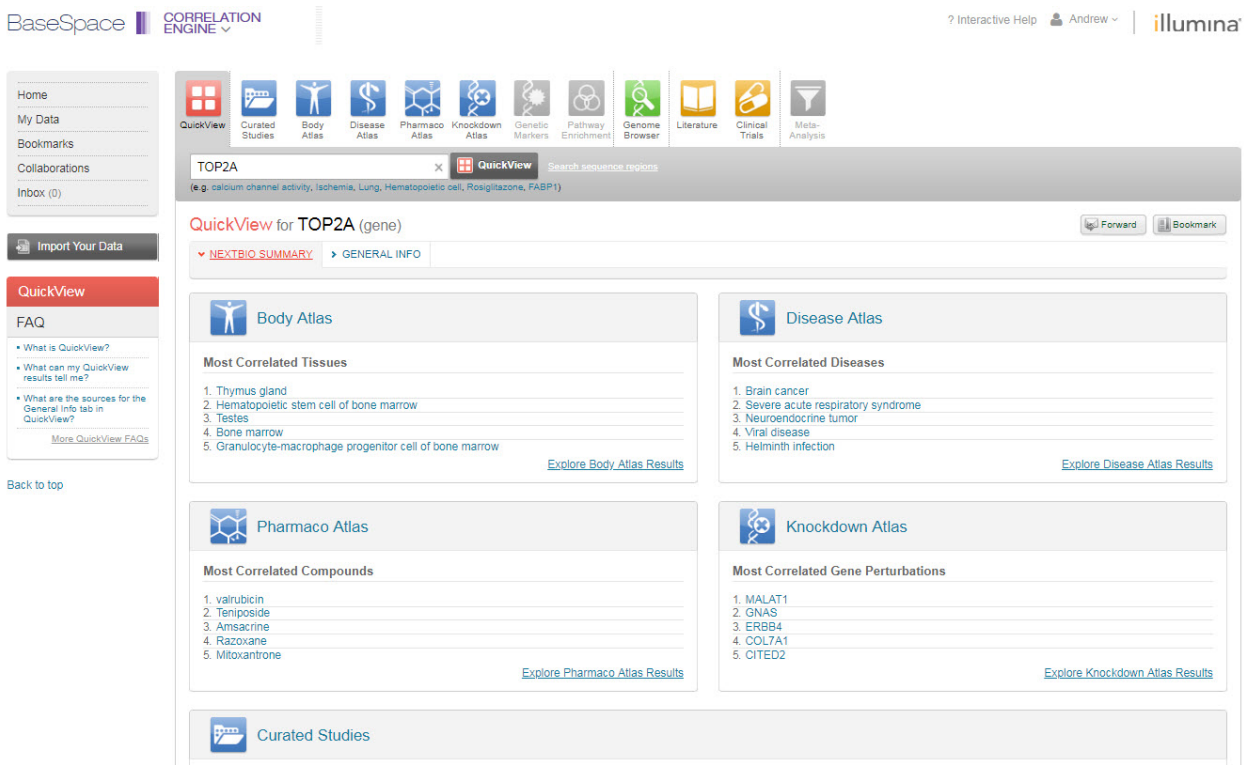


**Figure 1: BaseSpace Correlation Engine user interface enables queries for numerous association types —** Novel correlations and associations are quickly identified for a given query, revealing data driven connections between genes, diseases, compounds, tissues, pathways and literature.

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 GWAS. The rank-based enrichment algorithms used make the framework agnostic of technology platforms used to generate genomic data. This allows cross-analysis of data from different platforms, such as next-generation sequencing (NGS) and microarrays. BaseSpace Engine enables novel insights and discoveries by interrogating billions of data points derived from standardized analyses of whole genome studies.

The platform is powered by inter-species 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 —** Entire genomic content in BSCE is searched by orthologue 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. BaseSpace Engine enables users to assess the pathways that play significant roles in disease development across multiple studies and data types.

BaseSpace 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 Genotype-Tissue Expression (GTEx) project, and more.

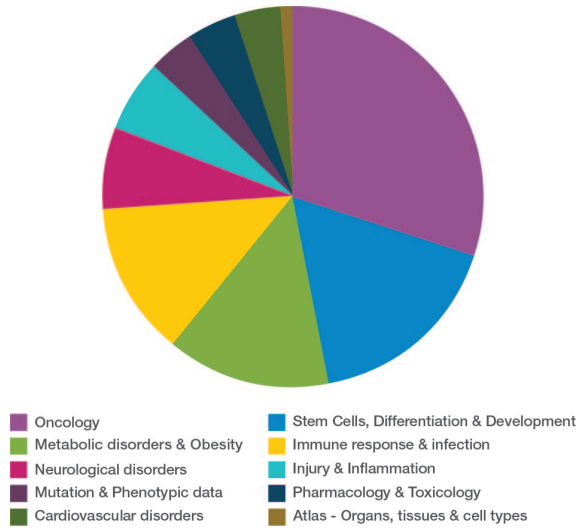


**Figure 3: BaseSpace Correlation Engine Curated Genomic Studies by Disease Area —** With over 135,000 analyses derived using standardized pipeline from more than half a million samples, the BSCE content is constantly growing. Data-driven analysis allows target assessment and validation, biomarker discovery, drug repositioning, etc..

### Mechanisms of gene function

BaseSpace 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 5,500 diseases across major disease areas and nearly 10,000 unique genetic perturbations.

### Mechanisms of drug action

In BaseSpace Engine, more than 50,000 analyses related to more than 4,500 compounds exist in the system. Researchers can analyze proprietary candidate molecules for on-target mechanisms and toxicity profiles comparing with profiles of other compounds.

## Start today

The public data available in BaseSpace 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 metanalysis 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, PCI DSS certified environment.

## Learn more

To purchase, start a free trial and learn more go to www.illumina.com/basespacecorrelationengine

For additional contact information, please see www.illumina.com/company/contact-us.html

Special academic pricing is available

**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.**