

# Biothings APIs: high-performance bioentity-centric web services

Cyrus Afrasiabi<sup>1</sup>, Jiwen Xin<sup>1</sup>, Ginger Tsueng<sup>1</sup>, Andrew I. Su<sup>1\*</sup>, Chunlei Wu<sup>1\*</sup>

<sup>1</sup> The Scripps Research Institute, 10550 N Torrey Pines Rd, La Jolla, CA 92037

\*Email: [cwu@scripps.edu](mailto:cwu@scripps.edu), [asu@scripps.edu](mailto:asu@scripps.edu)

**Project Website:** <http://biothings.io>, also see <http://mygene.info> and <http://myvariant.info>

**Source Code:** <https://github.com/sulab/biothings.api>

**License:** Apache License v2

The accumulation of biological knowledge and the advance of web and cloud technology are growing in parallel. The latest computation technology allows us to modernize the way we collect, organize and disseminate the growing biological knowledge. Just like what has been happening in the software-engineering field, biological data providers start to provide web-based APIs (Application Programming Interfaces) for accessing data in a simple and reliable manner, in addition to the traditional raw flat-file downloads. Web APIs provide many benefits over traditional file downloads. For instance, users can request specific data such as list of genes of interest without having to download the entire dataset, thereby providing the latest data on demand and reducing compute and data transfer times. Web APIs are also more likely to return data that conforms to common standards (e.g. JSON or XML). This means that programmers can spend less time on wrangling data, and more time on analysis and discovery.

Building and deploying scalable and high-performance web APIs requires sophisticated software engineering techniques that may not be known to many bioinformatics developers. We previously developed high-performance and scalable web APIs for gene and genetic variant annotations, accessible at [MyGene.info](http://MyGene.info) and [MyVariant.info](http://MyVariant.info). These two services are a tangible implementation of our expertise and collectively serve over 6 million requests every month from thousands of unique users. Crucially, the underlying design and implementation of these systems are in fact not specific to genes or variants, but rather can be easily adapted to other biomedical data types such drugs, diseases, pathways, species, genomes, domains and interactions, collectively, we refer them as “**BioThings**”.

Based on the existing MyGene.info and MyVariant.info APIs, we now provide a generic framework (or called **BioThings SDK**) for building the same high-performance APIs for other BioThings data types. This SDK enables other developers to build their own BioThings API for their specific data types. Users can take advantage of the abstracted technical layers we built into the SDK, and produce a high-performance API, which follows the best practice and community standards. We also adopted JSON-LD technology to form the connections between different data types, so that the set of BioThings APIs will form a network of linked programmatic-accessible biological knowledge. As these BioThings APIs are preferred to be built under the cloud environment, the BioThings SDK is essentially serving as a component of “Software as the Service”.