Scaling Up and Out: Data and Informatic Architectures to Meet High Volumes of Diverse Sequencing Applications

Seth Rhoades (srhoades@elevate.bio), ElevateBio, **Seth Rhoades**, ElevateBio, **Will Arnold**, ElevateBio, **Christina Codden**, ElevateBio, **Adriana Geldart**, Elevate Bio, **Cherylene Plewa**, ElevateBio

A high-throughput sequencing (HTS) service that often encounters low sample volumes, narrow scopes, and fixed sequencing technologies is sufficiently met with commensurate sample-tracking, sequence analysis, and data delivery mechanisms. Solutions that scale with massive experiment loads, diverse sequencing applications, and fast data turnaround justify coupling information management systems with laboratory and bioinformatic automation approaches. Herein, we describe a software-oriented series of workflows and pipelines designed to receive sequencing requests and return analysis for large experiments (hundreds to thousands of samples) for diverse applications on amplicons, mRNA, and whole exomes within days. Design decisions in these workflows stem from a desire to augment, not replace, existing solutions for laboratory management systems, cloud computing, notification services, and bioinformatic pipelines. In doing so, software engineers and bioinformatics focus their efforts on meeting tailored bioinformatics and statistical analytical requests from the scientist-clients. Ultimately, this system imbues a robust, scalable, and highly automated sample supply chain and analysis framework, enabling scientists to make faster decisions and design subsequent experiments. Internal key performance indicators, such as turnaround times, sample throughput, and sequencing quality, quickly identify bottlenecks and opportunities for improvement.