

A Bioinformatics Ecosystem for High-Throughput MAVE-SGE Experiments



Matthew Forbes^{1,*}, Anna Shtanova¹, Elvira Hernandez Gutierrez¹, Shriniwas Kulkarni¹, Swee Hoe Ong¹, Miguel Rada¹, Zakiya Salemahomed¹, Helen Schuilenburg¹, Priyanka Surana¹, Dijue Sun¹, Leah Wells¹, Jamie Billington¹, Adam Hunter¹, Ian Lobb¹, Rahul Sadanandan Nair¹, Wendy Thorpe¹, Charlotte Beaver¹, Sebastian Gerety¹, Andrew Waters¹, Sanger MAVE Project Team¹, David Parry-Smith¹, Sónia Gonçalves¹, David Jackson¹

1 – Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom

Motivation - Scaling MAVE-SGE

Saturation Genome Editing (SGE) - A CRISPR-based MAVE that measures the functional impact of thousands of variants in parallel, most commonly as loss-of-function effects via changes in cell survival.

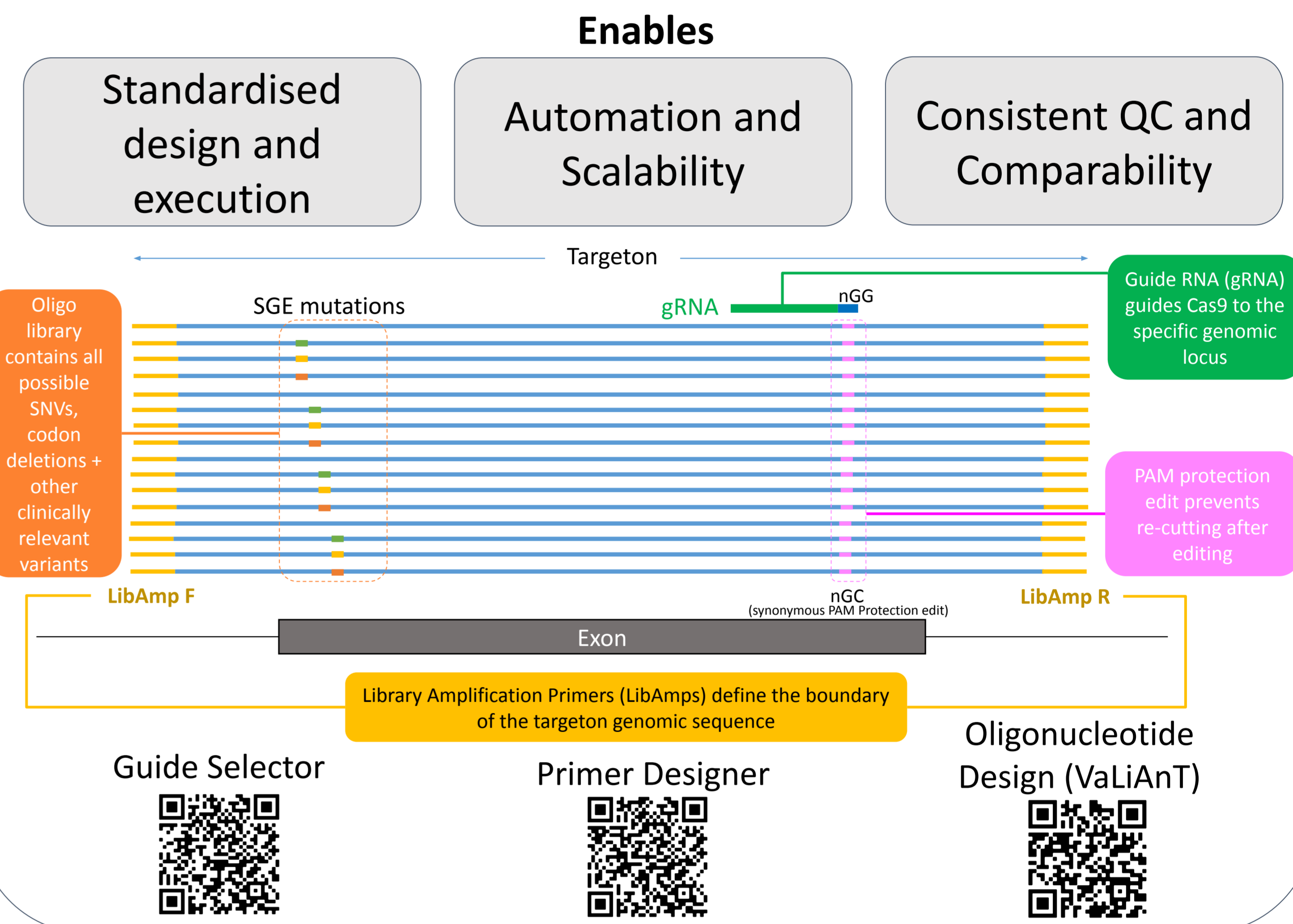
Vision - Generate variant effect maps for all ~20,000 genes in the human genome as part of the Atlas of Variant Effects.

Current Aim - Generate high quality variant effect maps for approximately 800 essential genes

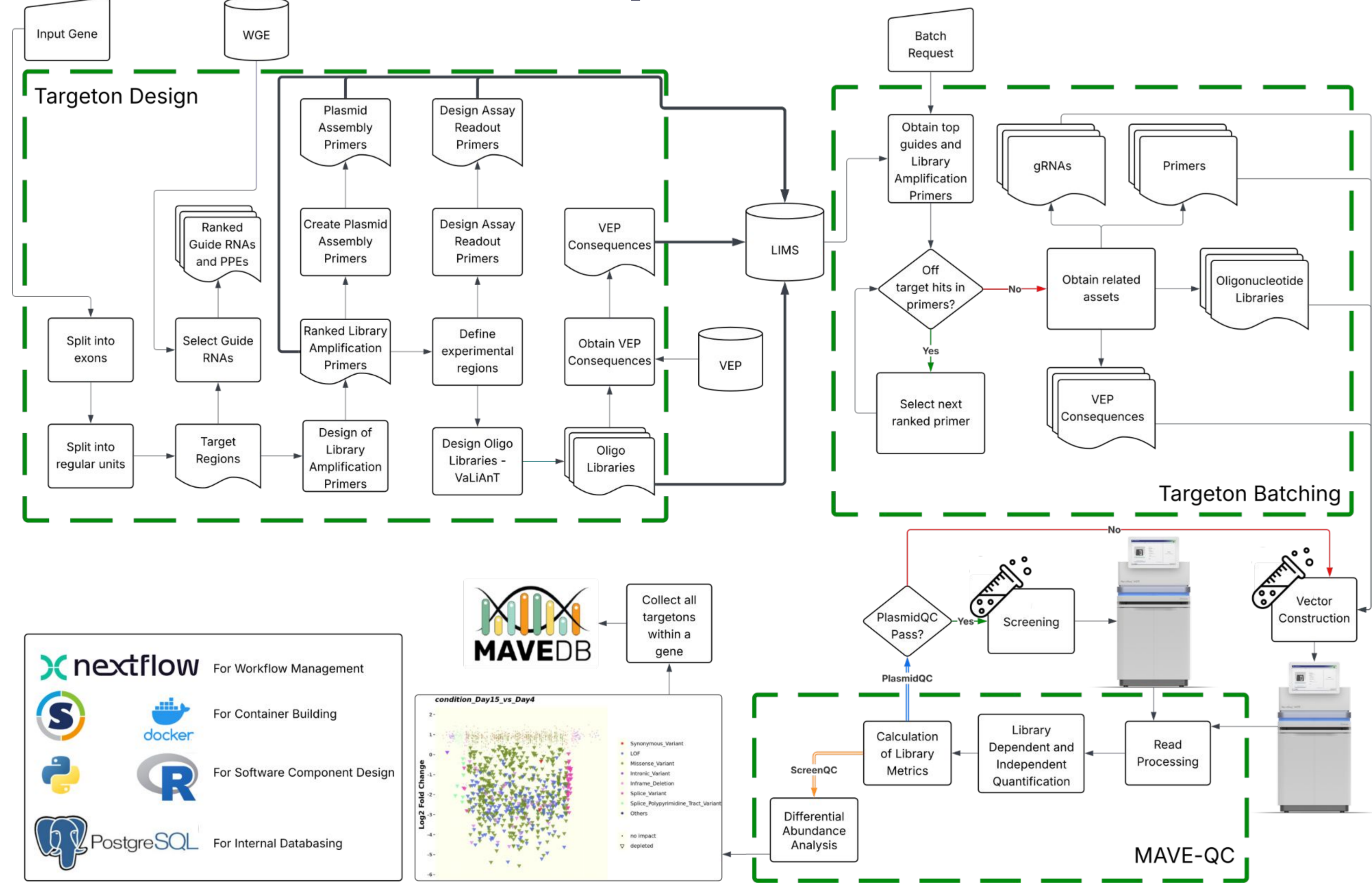
What will we need from Informatics? - A scalable ecosystem that **standardises** and **automates** MAVE workflows, and **integrates** with high-throughput laboratory processes—thereby enabling **consistent, high-quality** data generation with **low turnaround-time**.

The Targeton

A consistently sized genomic segment that acts as the fundamental unit of design, execution, and analysis for MAVE-SGE experiments.



From Concept to Products - End-to-End Informatics



What do our systems need?

- Reproducibility
- Speed
- Robustness
- Automation
- Scale
- Auditability
- Integration with Lab Processes
- Maintainability
- Standardisation

Summary

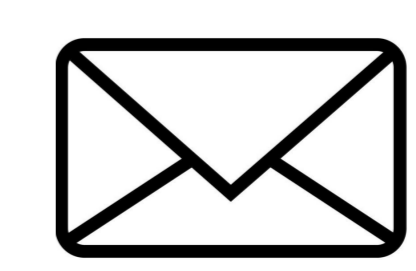
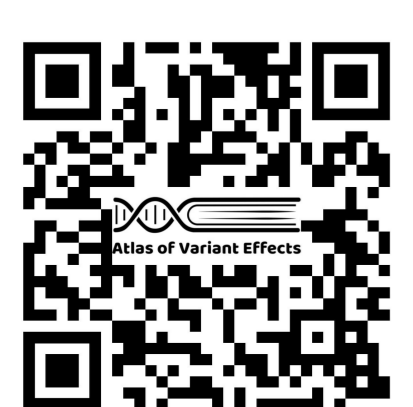
We are building a software ecosystem for end-to-end MAVE-SGE workflows based around the targeton as a reproducible unit

Aiming for seamless integration of automated experimental design and downstream analysis with experimental workflows

Standardisation, automation, and QC will ensure reproducible, high-quality data delivered with low turnaround time

Acknowledgments

The authors thank colleagues across the MAVE Programme, including MAVE Operations and faculty partners in the Human Genetics and Somatic Genomics Programmes. We also acknowledge operational teams in Scientific Operations, Cellular Operations and Informatics and Digital Systems for maintaining and running the infrastructure that enables this work. This work was supported by the Wellcome Trust [220540/Z/20/A].



Email: ms65@sanger.ac.uk
Or scan QR code

