A Genotype calling and haplotype/phenotype translation system supporting PGx implementation

# Main idea

Single base primer extension method (SBE) is an effective and sensitive approach that can cover over 30 loci in a single reaction. As a cost-effective approach, this method has been widely used in many applications such as forensic, disease diagnosis, and pharmacogenetics (PGx).

Several studies confirmed the benefit of panel based PGx testing in clinical settings and identifying star-alleles (haplotypes) from customed panels plays a role in PGx implementation. Many tools have been developed such as PharmCAT, Stargazer, Astrolabe, and Aldy and all of them mainly focus on analyzing Next-Generation Sequencing data. Moreover, genotyping software does not support matching haplotypes (GeneMapper in our case). Therefore, identifying haplotypes of customed PGx test panels remains challenging.

In a conventional way, PGx experts often have two options (Figure 1A): (1) they manually identify the haplotypes from raw intensity data (chromatogram plots); or (2) they identify genotype result using a genotyping software and upload the result to other tools to match haplotypes. Then, they must manually manage the data, analyzed result, and history for further steps. This way raises several following risks: firstly, manual processes increase human biases; secondly, it is hard to record and track the history of the whole process; finally, it is difficult to scale at many testing samples because it requires more efforts to manage all those data and processes.

In this work, we developed a cloud system that integrates all processes to overcome these challenges (Figure 1B). Once PGx experts upload the raw intensity data to the system, they can do all further steps: identifying genotype, quality controlling, and matching haplotypes/phenotypes. All activity logs have also been recorded for any further tracking steps.

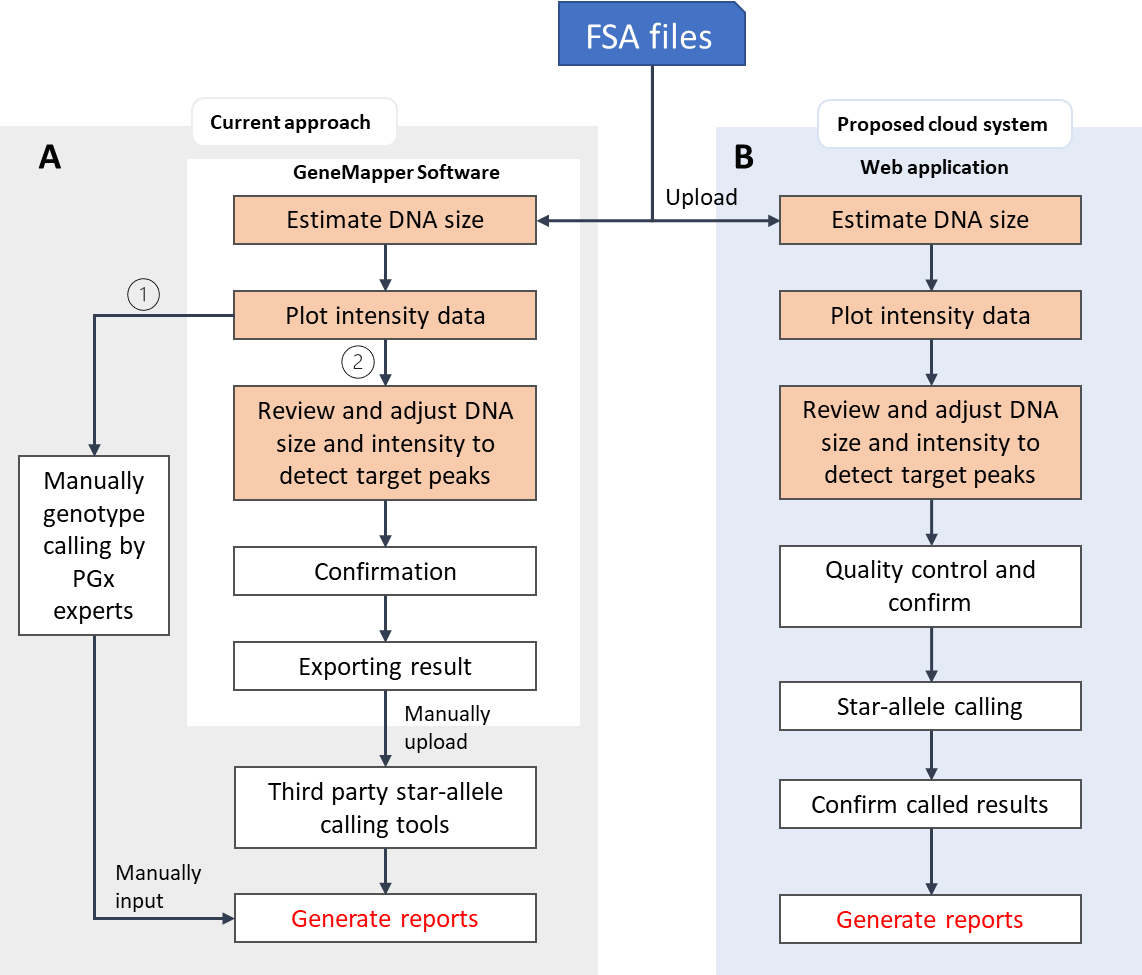


Figure 1. Overview about current approach and purpose solution

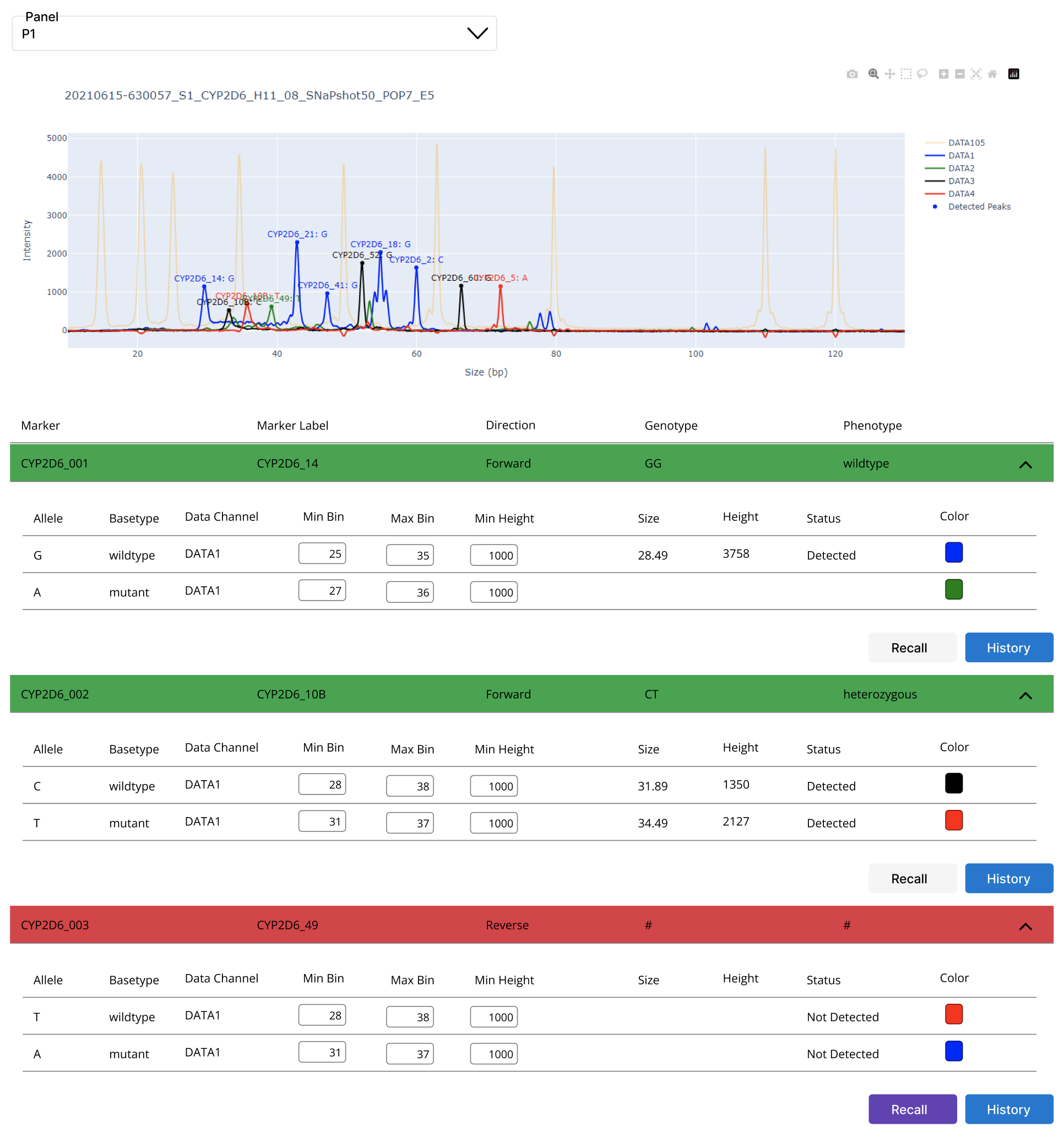


Figure 2. Prototype of the system implementing on web application; (A) raw intensity plot with detected marker; (B) present; (C) checking history by each marker; (D) user can adjust data and re-identify genotype for a specific marker.

Because of customizable, Laboratory-developed test (LDT) genotyping panel is critical and cost-effective for PGx implementation in current setting. However, development of star allele calling function (genotyping) is needed for this panel to prevent human bias and increase efficiency.

Identification of haplotype (star allele)/phenotype of metabolizer plays an important role in PGx implementation. However, the genotyping software package for genotyping data (GeneMapper) does not include a module to identify star-allele based on the customized panel data.

In fact, it requires many extra steps from genotyping to haplotype/phenotype translation, and report generation since software and systems are separated.

We briefly describe here steps of the current approach and purpose integrated cloud system as a solution. For current approach, FSA file obtained from (AAAA) was processed and analyzed by GeneMapper. From that software, users can do many processes for genotype calling such as DNA sizing, adjust bin range and minimum intensity to detect the peak, after confirmation, user can get the plot intensity. There are two options for haplotype/phenotype translation.

Firstly, the user will see the intensity data and manually match haplotype/phenotype based on their experiences. Then the result will be manually input to the system to generate reports.

Secondly, the users can review the genotype calling data on the software. In this process, bin range and minimum intensity of each marker can be adjusted manually. The genotype calling result will then be exported as file after confirmation. The exported file can be loaded to star-allele calling tools to identify the haplotype/diplotype. Then the result can be automatically or manually put into the system to generate reports.

The two above options have some of the following drawbacks

It is difficult to track change and manage because of separate systems.

Increase the risk of human bias.

Difficult to maintain

Low efficiency

We proposed a cloud system that covers all these steps in one system.

We developed functions that support DNA sizing, automatically peak detection based on pre-defined bin range and minimum intensity. It also allows user to manual update these values and auto-update the genotype calling result.

Single base primer extension method (SBE) is an effective and sensitive approach that can cover over 30 loci in a single reaction. This method has been widely used in a broad range of applications such as forensic, molecular diagnosis, disease diagnosis, and pharmacogenetics/pharmacogenomics (PGx). SNaPhot multiplex system is a primer extension-based method for the analysis of single nucleotide polymorphism (SNPs)

Single base primer extension is a cost-effective approach