star_allele_calling

October 6, 2022

— title: "Customized star allele calling"

—

1 Introduction

- Instroduction about SNAPShot
- The process of genotype calling
- Vary each run time
- Benefit of cloud base system

Star-allele calling from genotyping results is crucial for PGx implementation in clinical practice. Additionally, a laboratory-developed test (LDT) panel for PGx implementation is also critical and cost-effective for a specific population; however, the star-alleles calling tools for the panel need to be developed. In this work, therefore, we developed a star-allele calling tool applied for the developed testing panel. The tool also was integrated into the system and website application.

Taken all into consideration, we develop a cloud base system that support user can call genotype and matching with our knowledge database.

2 Methods

2.1 Loading FSA data

Raw data from fragment analysis sofware were loaded using package Biopython (version) package. The intensisty data were stored in different predefined chanel name such as DATA105 for refrernce chanel, for other chanels defined for 4 nucleotites (DATA1: A; DATA2:C; DATA3:G and DATA4:T). Those intensisty data were primary used in peak detection process.

2.2 Reference peak detection

As mentioned aboved, the DATA105 chanel as refrence intensity was used to detect the reference peak. Usually, the information of reference peak depends on the experimen desgin. In this study, we used ??? GeneScan 120 Lize dye size standard (CYP2D6 kit developed by SPMED Co. Ltd. (GTR link)).

We used peak finding function of scipy (version) with default height 800, width of peak, … However, the setting information can be adjusted by experties.

The number of detected peaks should be equal to number of reference size standard. In case of LIZ 120, there are 9 sizes, therefore the number of peaks should be 9 peaks. Only qualifed peak

detection was used to develop Least Square model as the next step, otherwise, it need to be adjusted by user to detect the correct peaks.

2.3 Sizing model development

As described in the manual of gene mapper software, there were three method to interfer length of DNA based on intensity points (Local Southern, Global Southern and Least Square). In the work, we used Least Square as advance method to develop a model for sizing identification.

The model was developed using numpy (version) python package. Users can choose the order of least sequare method (second-order or third-order). The performance of this model was evaluated using R square index from actual size and precited size. Based on R square, we found that third-order resulted in a slightly higher than second order, therefore, thrid order was set as default of the model. The model development process will be done by each input FSA file.

Workflow

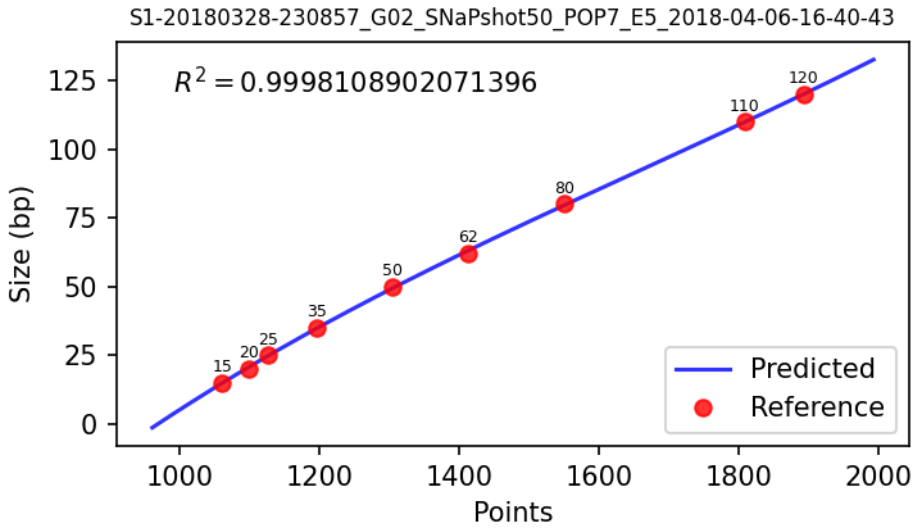
Please see here: https://app.diagrams.net/#G1HzMMyRkVhr8Y_M1wMmXwmZagIEpak1Bb

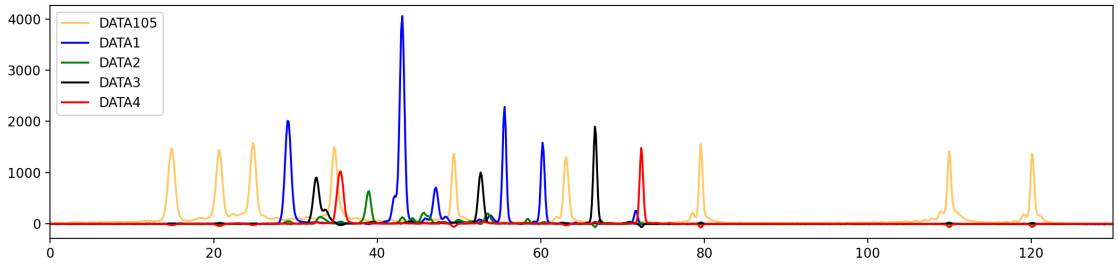
2.4 Load defined marker with colors

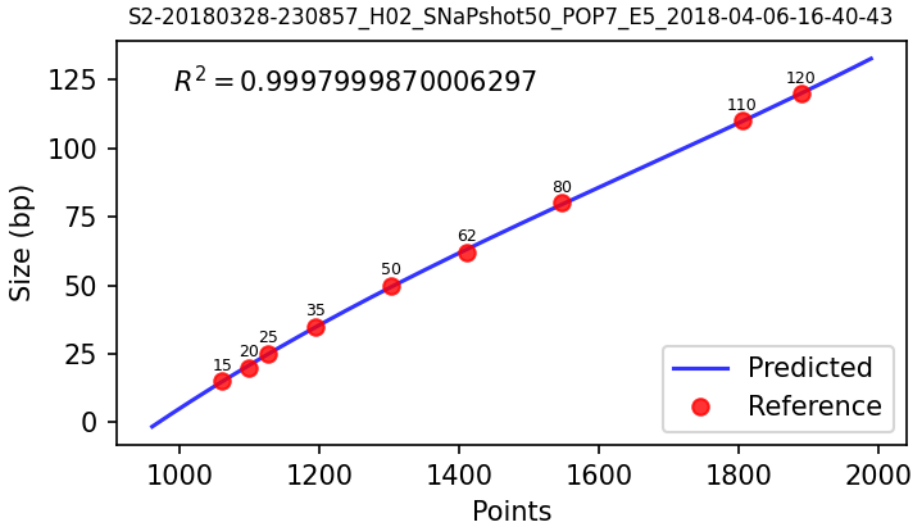
This form is test form and it was based on the format from PGx team.

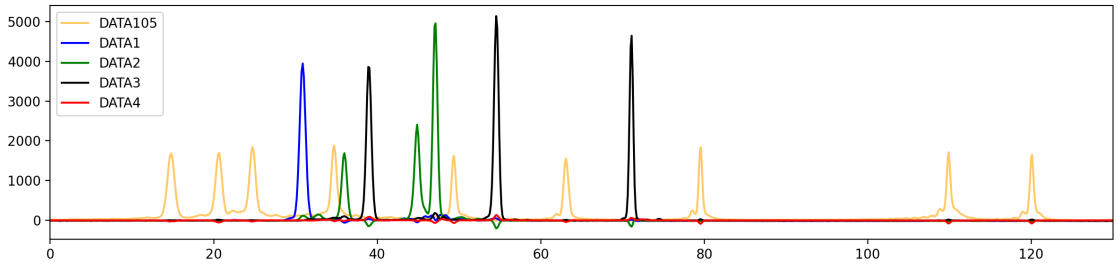
However this form have a disavantage if testing allele is not biallelic for example triallelic (A > C/T). and this requires a new format to overcome this issue.

2.5 Example call from fsa file with CYP2D6 and 2 panels

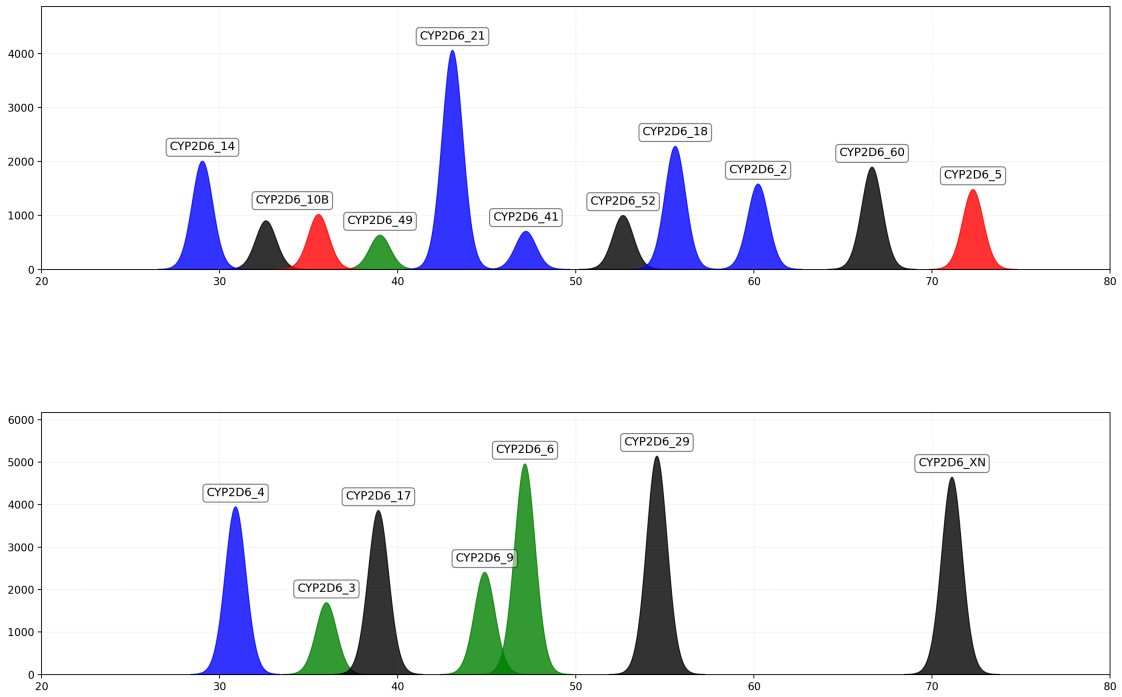








20180328-230857:*1/*10B



2.6 Example call from bin file

20180328-230857:*1/*10B
PTC:*1/*1
NTC:

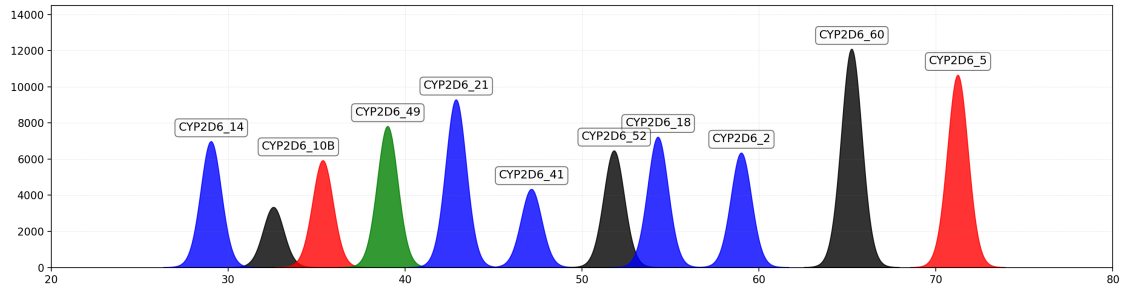
	sample	gene	marker	label	panel	direction	genotype	\
0	20180328-230857	CYP2D6	CYP2D6_001	CYP2D6_14	S1	Forward	GG	
2	20180328-230857	CYP2D6	CYP2D6_002	CYP2D6_10B	S1	Forward	CT	
4	20180328-230857	CYP2D6	CYP2D6_003	CYP2D6_49	S1	Reverse	TT	
6	20180328-230857	CYP2D6	CYP2D6_004	CYP2D6_21	S1	Forward	GG	
8	20180328-230857	CYP2D6	CYP2D6_005	CYP2D6_41	S1	Forward	GG	
10	20180328-230857	CYP2D6	CYP2D6_006	CYP2D6_52	S1	Reverse	GG	
12	20180328-230857	CYP2D6	CYP2D6_007	CYP2D6_18	S1	Forward	GG	
14	20180328-230857	CYP2D6	CYP2D6_008	CYP2D6_2	S1	Reverse	CC	
16	20180328-230857	CYP2D6	CYP2D6_009	CYP2D6_60	S1	Reverse	GG	
18	20180328-230857	CYP2D6	CYP2D6_010	CYP2D6_5	S1	Reverse	AA	
20	20180328-230857	CYP2D6	CYP2D6_011	CYP2D6_4	S2	Forward	GG	
22	20180328-230857	CYP2D6	CYP2D6_013	CYP2D6_17	S2	Forward	CC	
24	20180328-230857	CYP2D6	CYP2D6_012	CYP2D6_3	S2	Forward	AA	
26	20180328-230857	CYP2D6	CYP2D6_015	CYP2D6_6	S2	Reverse	TT	
28	20180328-230857	CYP2D6	CYP2D6_014	CYP2D6_9	S2	Forward	AA	
30	20180328-230857	CYP2D6	CYP2D6_016	CYP2D6_29	S2	Reverse	GG	
32	20180328-230857	CYP2D6	CYP2D6_017	CYP2D6_XN	S2	Reverse	GG	
34	PTC	CYP2D6	CYP2D6_001	CYP2D6_14	S1	Forward	GG	

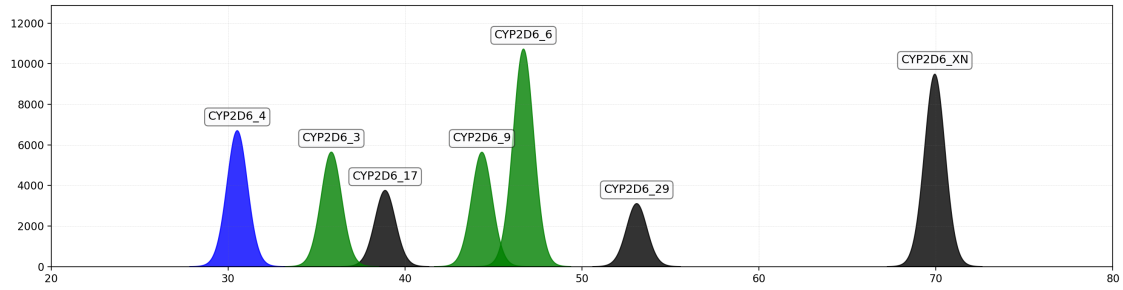
36	PTC	CYP2D6	CYP2D6_002	CYP2D6_10B	S1	Forward	CC
38	PTC	CYP2D6	CYP2D6_003	CYP2D6_49	S1	Reverse	TT
40	PTC	CYP2D6	CYP2D6_004	CYP2D6_21	S1	Forward	GG
42	PTC	CYP2D6	CYP2D6_005	CYP2D6_41	S1	Forward	GG
44	PTC	CYP2D6	CYP2D6_006	CYP2D6_52	S1	Reverse	GG
46	PTC	CYP2D6	CYP2D6_007	CYP2D6_18	S1	Forward	GG
48	PTC	CYP2D6	CYP2D6_008	CYP2D6_2	S1	Reverse	CC
50	PTC	CYP2D6	CYP2D6_009	CYP2D6_60	S1	Reverse	GG
52	PTC	CYP2D6	CYP2D6_010	CYP2D6_5	S1	Reverse	AA
54	PTC	CYP2D6	CYP2D6_011	CYP2D6_4	S2	Forward	GG
56	PTC	CYP2D6	CYP2D6_013	CYP2D6_17	S2	Forward	CC
58	PTC	CYP2D6	CYP2D6_012	CYP2D6_3	S2	Forward	AA
60	PTC	CYP2D6	CYP2D6_015	CYP2D6_6	S2	Reverse	TT
62	PTC	CYP2D6	CYP2D6_014	CYP2D6_9	S2	Forward	AA
64	PTC	CYP2D6	CYP2D6_016	CYP2D6_29	S2	Reverse	GG
66	PTC	CYP2D6	CYP2D6_017	CYP2D6_XN	S2	Reverse	GG
68	NTC	CYP2D6	CYP2D6_001	CYP2D6_14	S1	Forward	
70	NTC	CYP2D6	CYP2D6_002	CYP2D6_10B	S1	Forward	
72	NTC	CYP2D6	CYP2D6_003	CYP2D6_49	S1	Reverse	
74	NTC	CYP2D6	CYP2D6_004	CYP2D6_21	S1	Forward	
76	NTC	CYP2D6	CYP2D6_005	CYP2D6_41	S1	Forward	
78	NTC	CYP2D6	CYP2D6_006	CYP2D6_52	S1	Reverse	
80	NTC	CYP2D6	CYP2D6_007	CYP2D6_18	S1	Forward	
82	NTC	CYP2D6	CYP2D6_008	CYP2D6_2	S1	Reverse	
84	NTC	CYP2D6	CYP2D6_009	CYP2D6_60	S1	Reverse	
86	NTC	CYP2D6	CYP2D6_010	CYP2D6_5	S1	Reverse	
88	NTC	CYP2D6	CYP2D6_011	CYP2D6_4	S2	Forward	
90	NTC	CYP2D6	CYP2D6_013	CYP2D6_17	S2	Forward	
92	NTC	CYP2D6	CYP2D6_012	CYP2D6_3	S2	Forward	
94	NTC	CYP2D6	CYP2D6_015	CYP2D6_6	S2	Reverse	
96	NTC	CYP2D6	CYP2D6_014	CYP2D6_9	S2	Forward	
98	NTC	CYP2D6	CYP2D6_016	CYP2D6_29	S2	Reverse	
100	NTC	CYP2D6	CYP2D6_017	CYP2D6_XN	S2	Reverse	

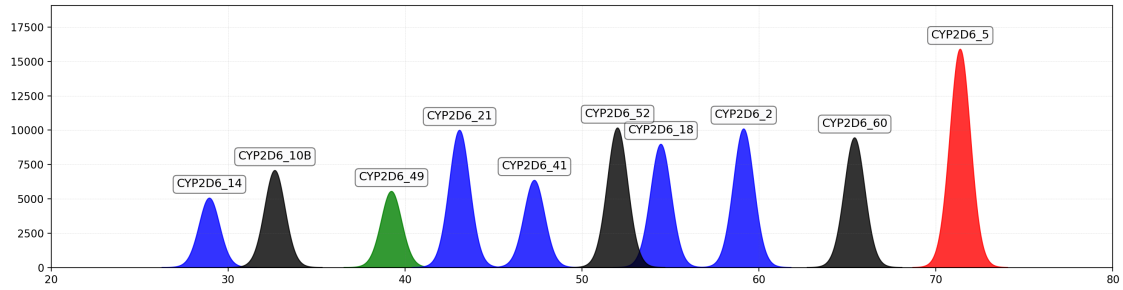
	phenotype
0	wildtype
2	heterozygous mutant
4	wildtype
6	wildtype
8	wildtype
10	wildtype
12	wildtype
14	wildtype
16	wildtype
18	wildtype
20	wildtype
22	wildtype

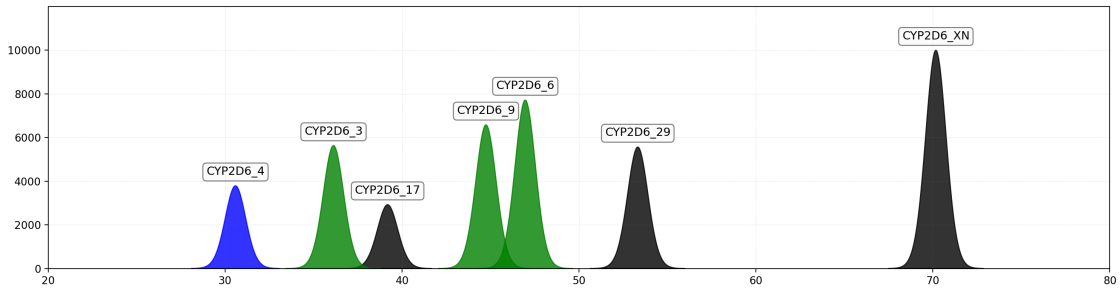
```
24              wildtype
26              wildtype
28              wildtype
30              wildtype
32              wildtype
34              wildtype
36              wildtype
38              wildtype
40              wildtype
42              wildtype
44              wildtype
46              wildtype
48              wildtype
50              wildtype
52              wildtype
54              wildtype
56              wildtype
58              wildtype
60              wildtype
62              wildtype
64              wildtype
66              wildtype
68
70
72
74
76
78
80
82
84
86
88
90
92
94
96
98
100

            sample    gene      marker      label panel direction genotype  \
0  20180328-230857  CYP2D6  CYP2D6_001  CYP2D6_14    S1   Forward       GG

  phenotype
0  wildtype
```

```
            sample    gene      marker      label panel direction base   size  \
0  20180328-230857  CYP2D6  CYP2D6_001  CYP2D6_14    S1   Forward    G  29.04

   height color
0  6967.0  blue
```



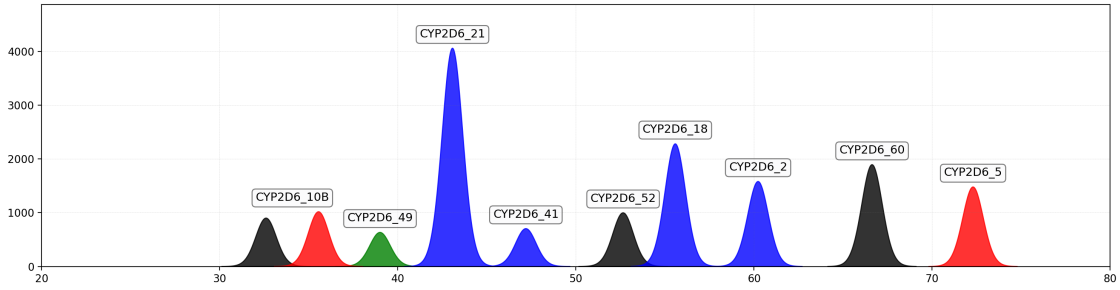


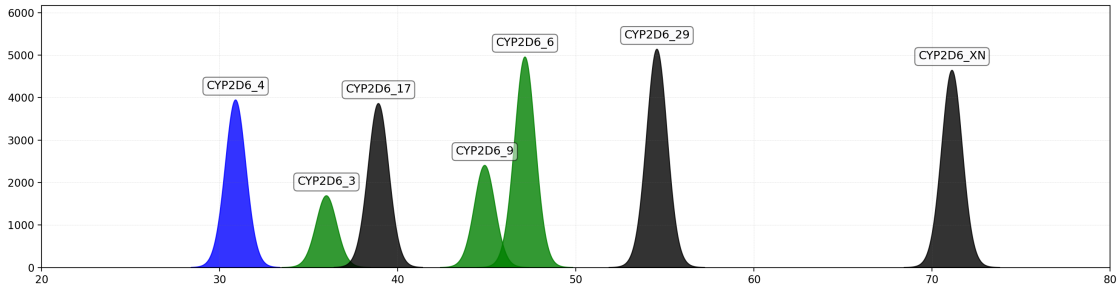


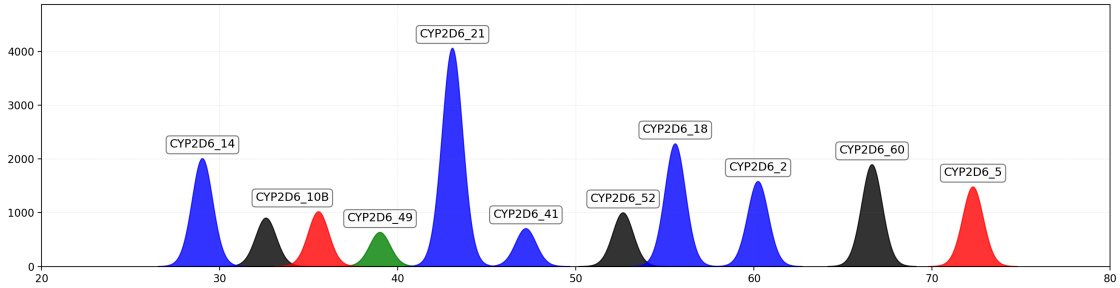


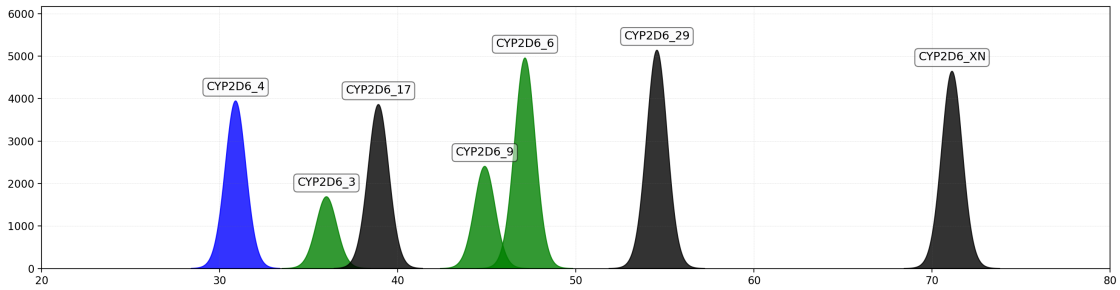
3 Update bin range to call genotype example

```
           sample    gene      marker       label panel direction base  \
0  20180328-230857  CYP2D6  CYP2D6_001   CYP2D6_14    S1   Forward    G
1  20180328-230857  CYP2D6  CYP2D6_001   CYP2D6_14    S1   Forward    A
2  20180328-230857  CYP2D6  CYP2D6_002  CYP2D6_10B    S1   Forward    C
3  20180328-230857  CYP2D6  CYP2D6_002  CYP2D6_10B    S1   Forward    T

   basetype  min_bin  max_bin  min_height  is_forward  is_detected peak  \
0  wildtype       20       25         500           1        False
1    mutant       27       36         500           1        False
2  wildtype       28       38         500           1         True   32
3    mutant       31       37         500           1         True   32

    size  height status                                          message  \
0                     Peak(s) could not be detected. Please check pe…
1                     Peak(s) could not be detected. Please check pe…
2   32.6   901.0     ok
3  35.55  1019.0     ok

   color
0   blue
1  green
2  black
3    red
```

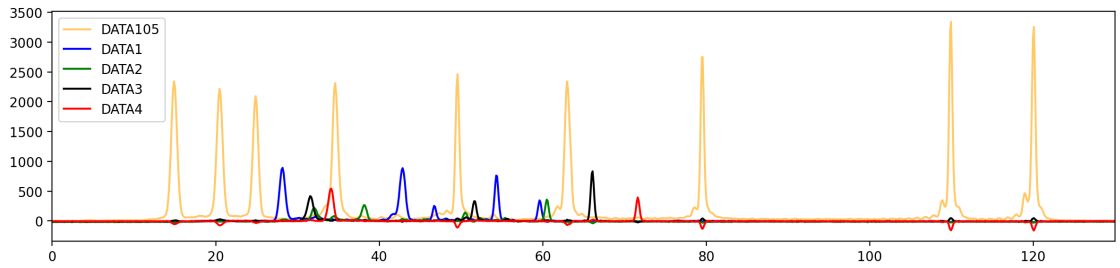


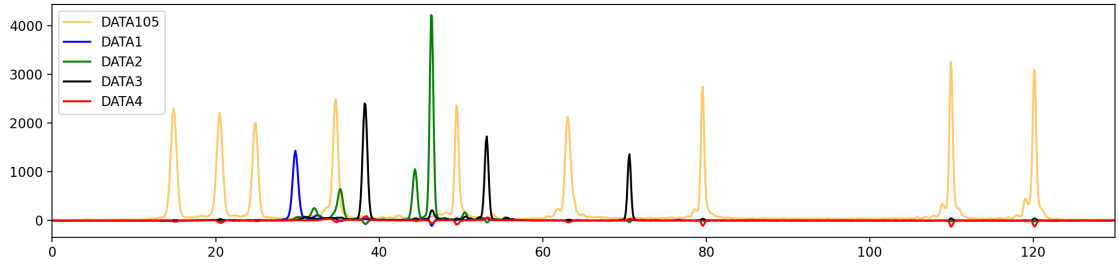


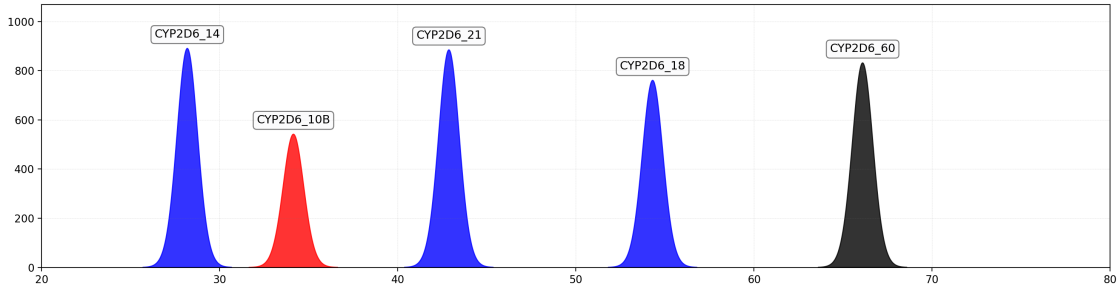


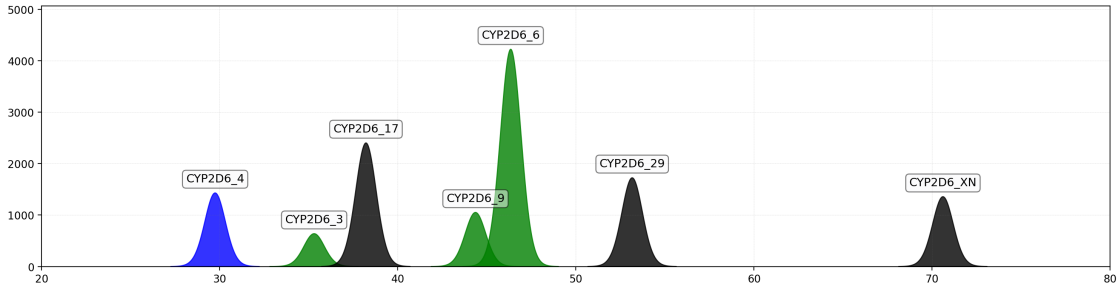


4 Example of adjusting height





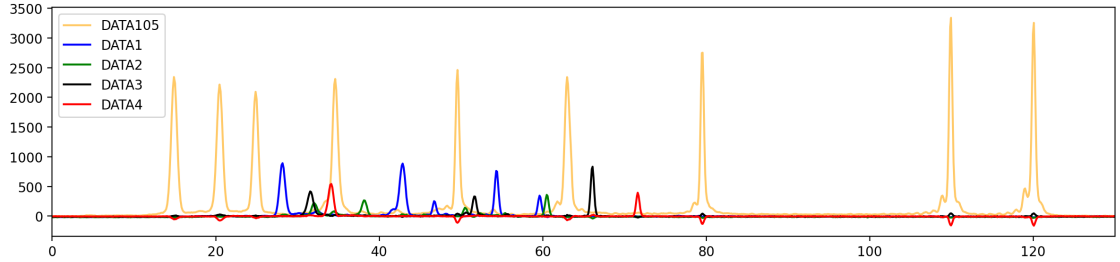


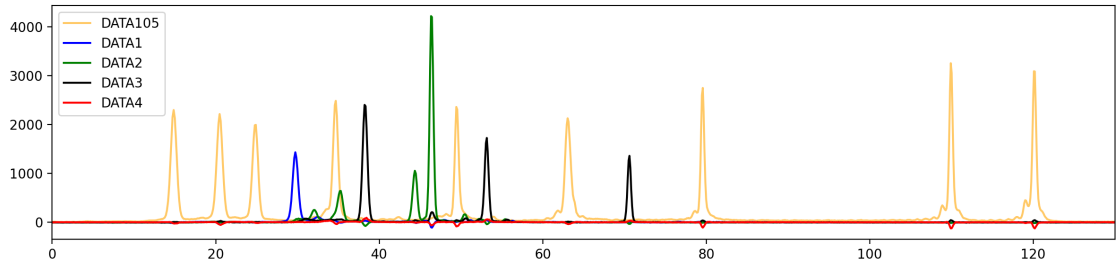


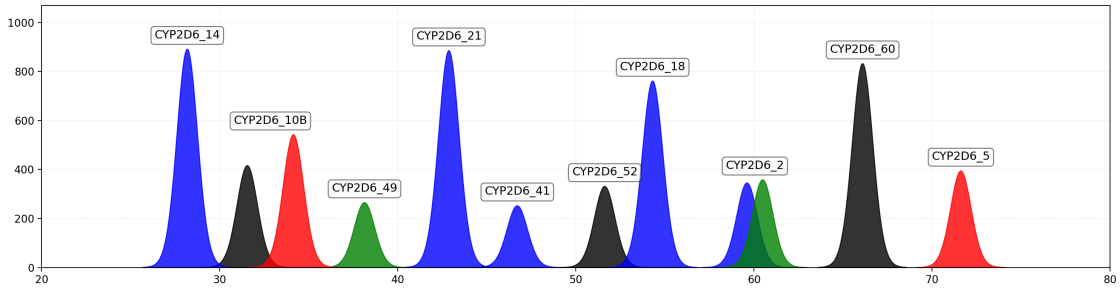
	sample	gene	marker	label	panel	direction	base
0	20181106-730829	CYP2D6	CYP2D6_001	CYP2D6_14	S1	Forward	G
1	20181106-730829	CYP2D6	CYP2D6_001	CYP2D6_14	S1	Forward	A
2	20181106-730829	CYP2D6	CYP2D6_002	CYP2D6_10B	S1	Forward	C
3	20181106-730829	CYP2D6	CYP2D6_002	CYP2D6_10B	S1	Forward	T

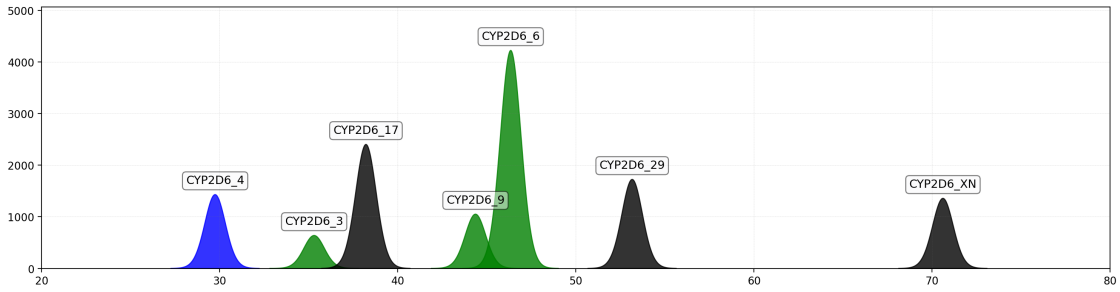
	basetype	min_bin	max_bin	min_height	is_forward	is_detected	peak
0	wildtype	25	35	500	1	True	25
1	mutant	27	36	500	1	False	
2	wildtype	28	38	500	1	False	
3	mutant	31	37	500	1	True	25

```
    size height status                                                 message  \
0  28.18  891.0     ok
1                      Peak(s) could not be detected. Please check pe…
2                      Peak(s) could not be detected. Please check pe…
3  34.14  542.0     ok

   color
0   blue
1  green
2  black
3    red
```








5000
4000
3000
2000
1000
0
20
30
40
50
60
70
80
CYP2D6_4
CYP2D6_3
CYP2D6_17
CYP2D6_9
CYP2D6_6
CYP2D6_29
CYP2D6_XN