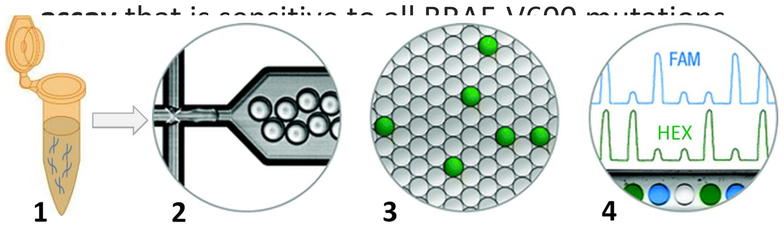
Automatic quantification of BRAF-V600 codon mutations

Dean Attali[1], **Roza Bidshahri**[2], **Charles Haynes**[2,3], **Jennifer Bryan**[3,4]

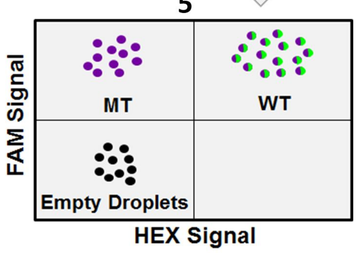
[1]Bioinformatics Training Program, University of British Columbia, Vancouver, BC, Canada; [2]Department of Chemical and Biological Engineering; [3]Michael Smith Laboratories; [4]Department of Statistics

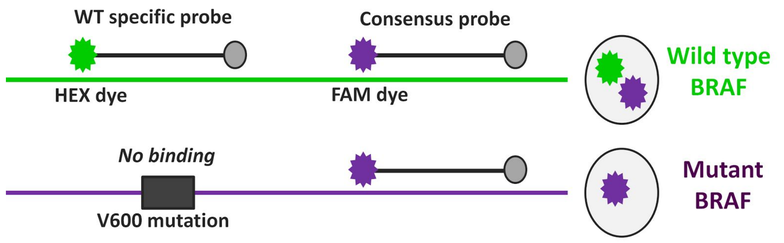
I. Introduction

- **B-Raf protein** (BRAF gene) is part of **MAPK pathway**
- MAPK initiated by growth factor, leads to **cell growth**
- BRAF **activating mutations** (V600 codon substitution) result in uncontrolled cell proliferation → implicated in **40% of melanomas** & **10% of colorectal cancers**
- **BRAF-V600** is **predictive biomarker** – mutation detection is **crucial for** proper **course of treatment**
- Existing BRAF-V600 mutation detection kits:
  - × Detect only a subset of possible mutations
  - × Require at least 5% mutation frequency
  - × Output yes/no; not actual mutation frequency
- Haynes lab developed **droplet digital PCR** (ddPCR) assay that is sensitive to all BRAF-V600 mutations



ddPCR protocol – 1. Sample DNA loaded. **2.** DNA partitioned into 20,000 droplets. **3.** Massively parallel qPCR in each droplet. **4.** Reader detects fluorescent in droplets in two wavelengths. **5.** Mutant frequency can be inferred by plotting signal in all droplets as 2D scatterplot.





BRAF ddPCR assay design – Two probes with fluorescent dyes are used. Probe with FAM dye binds to consensus region and will bind to any BRAF template. Probe with HEX dye binds across V600 codon and only binds to wild type (wt) BRAF.

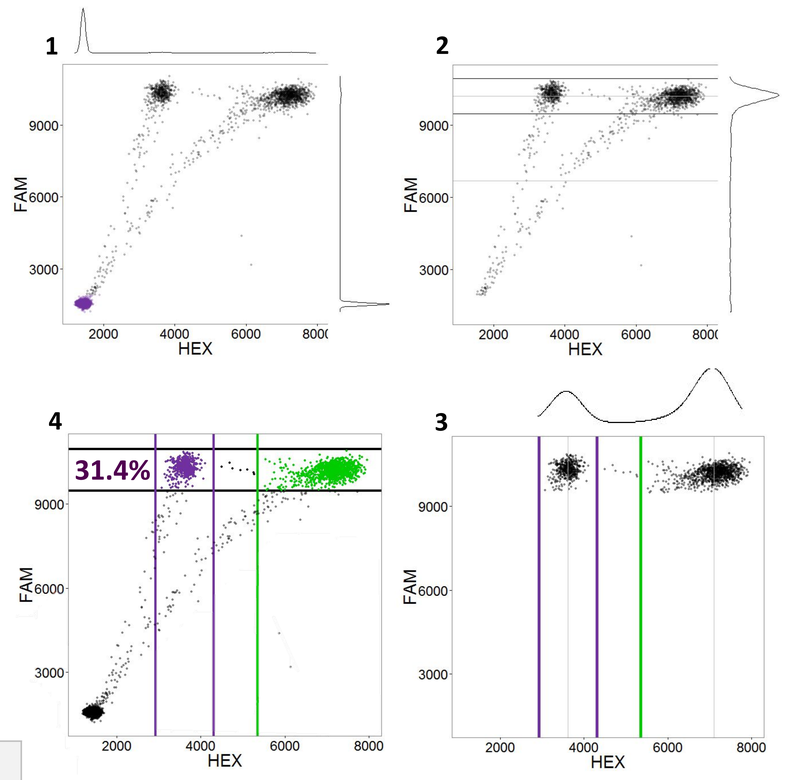
II. Objective

- We aim to replace manual analysis of ddPCR data with automated approach
- Given ddPCR data, estimate proportion of BRAF templates in sample that harbour V600 mutation

III. Methodology

Step	Description	Data
0	Initial ddPCR data	86 wells 16k drops/well
1	Remove outlier droplets	8 drops removed
2	Remove failed wells	9 wells removed (4 NTC)
3	Remove empty droplets	15k drops/well removed (95%)
4	Classify droplets as mutant vs wild type	32 wells wild type BRAF, 45 mutant

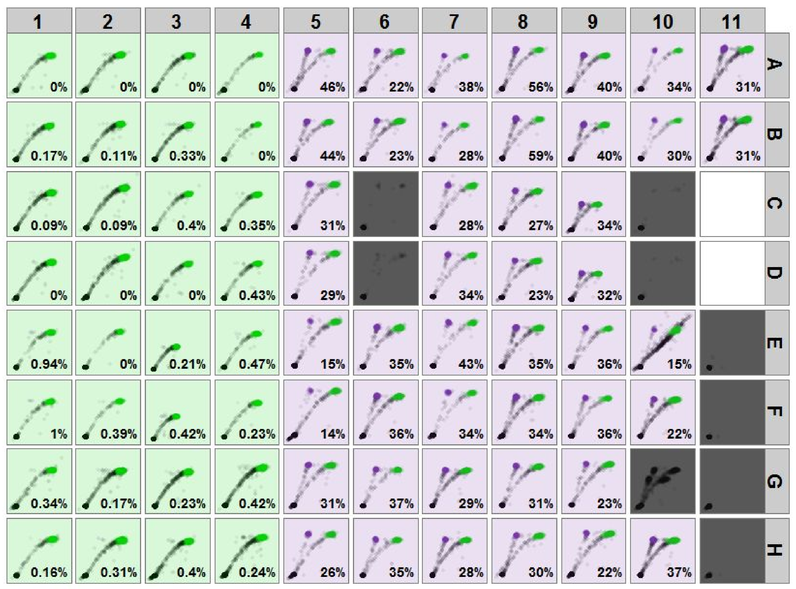
Pipeline steps overview – data shown for dataset consisting of formalin-fixed, paraffin-embedded samples from 41 colorectal cancer patients.

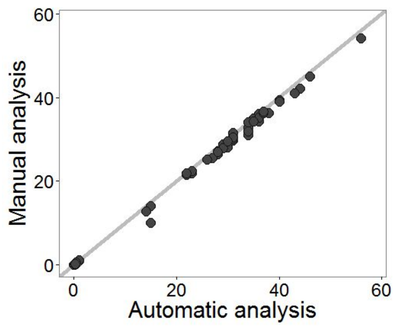


Steps 3-4 of pipeline: Using Gaussian mixture models to analyse droplets – 1. Most wells have ~95% empty droplets which are identified and removed. These drops are identified by fitting two Gaussian distributions (dist'n) along FAM and using the center & standard deviation (sd) of the lower dist'n to set an empty threshold. **2.** Filled droplets are identified similarly: two Gaussian dist'ns model FAM values, use center & sd of upper dist'n as filled threshold. **3.** After removing non-filled droplets, a similar approach is used to classify droplets as containing mutant or wild type BRAF-V600 templates. Mixture of two Gaussians are used to model HEX values, use center & sd of lower and upper dist'ns to define thresholds for mutant and wild type, respectively.
4. Calculate mutant frequency using droplet counts.

$$Mutant\ frequency = \frac{\#\ mutant\ drops}{\#\ mutant\ drops + \#\ wild\ type\ drops} = \frac{195}{195+426} = 0.314 = 31.4\%.$$

IV. Results





Analysis result – Running the full pipeline on colorectal cancer patients dataset took one minute on a personal laptop. Background colours: green = wild type BRAF, magenta = mutant BRAF, grey = failed well.

Algorithm accuracy – The automatically calculated mutation frequency vs the manually calculated mutation frequency of each sample. Grey line is y = x, points along the line signify perfect agreement between manual and automatic calculation.

V. Conclusions

- Automatic ddPCR analysis can attain similar results to human expert
- ddPCR assay more sensitive than existing BRAF-V600 mutation test kits – detects 1% mutation level
- Algorithm achieves high accuracy – successfully identified all samples deemed mutant by experts
- Fully automated analysis of ddPCR means objective and reproducible output
- More ddPCR assays can benefit from similar approaches to automate analysis