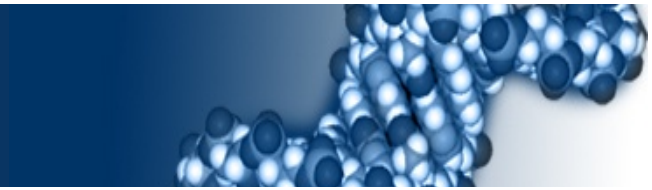
Parameterization of an *in-silico* DNA pipeline with & laboratory-specific experimental data allows for efficient & validation of the DNA analysis process &

Boston University School of Medicine
Program in Biomedical Forensic Sciences
72 E. Concord Street, Boston, MA 02118



Harish Swaminathan &

Forensic Science Error Management &
International Forensics Symposium &
NIST, Gaithersburg, MD &
July 27, 2017 &

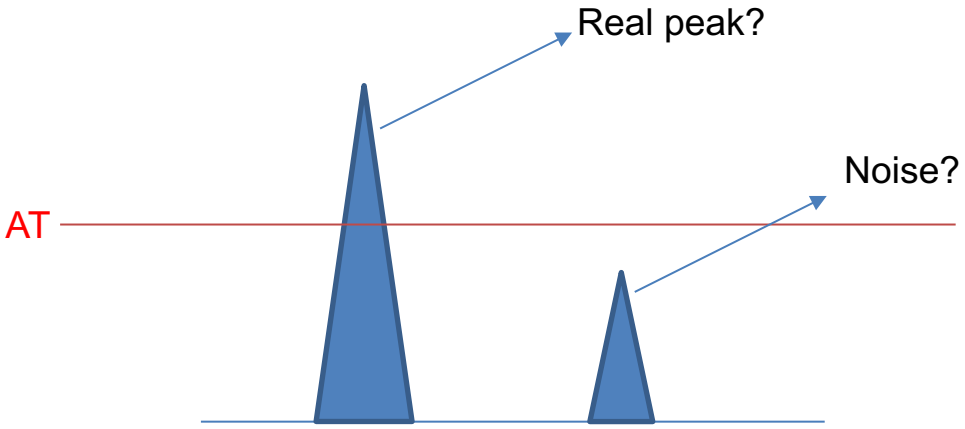
Quality of the electropherogram affects DNA mixture interpretation &

- Electropherogram (EPG): Allele signal + Background noise + Artifacts
- Interpretation can be challenging
- Mixtures with major and **minor**
- Low copy number samples typically exhibit signal loss
  - Sampling effects
  - **Detection effects**

$$Likelihood\ Ratio\ (LR) = \frac{\Pr(\boldsymbol{E}|H_1)}{\Pr(\boldsymbol{E}|H_2)}$$

- By improving the information content of E, one can expect a more informative LR
  - For e.g. a large LR for a true contributor and a small LR for a non-contributor
- Focus of the talk is on development of a validation scheme to **improve signal-to-noise resolution** and to **minimize detection error rates**

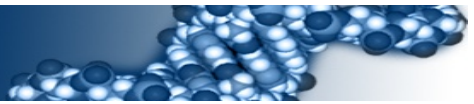
Optimal AT is necessary to minimize detection errors &



- Analytical Threshold (AT): the minimum height requirement at and above which detected peaks can be reliably distinguished from background noise*
- Errors can occur while applying an AT
- False Positive or Type I error: Noise peaks are mislabeled as real peaks
- False Negative or Type II error: Real peaks are not labelled (dropout)
- Ideally, the chosen AT minimizes both types of errors
- **AT impacts downstream interpretation process, including the match statistic**

*SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories – APPROVED 01/12/2017 &

Combined simulation + experimental approach &



- **Time** and **cost** are limiting factors in validation
  - For e.g. AT should be determined by large-scale in-house validation studies using negatives, dilution series, etc.

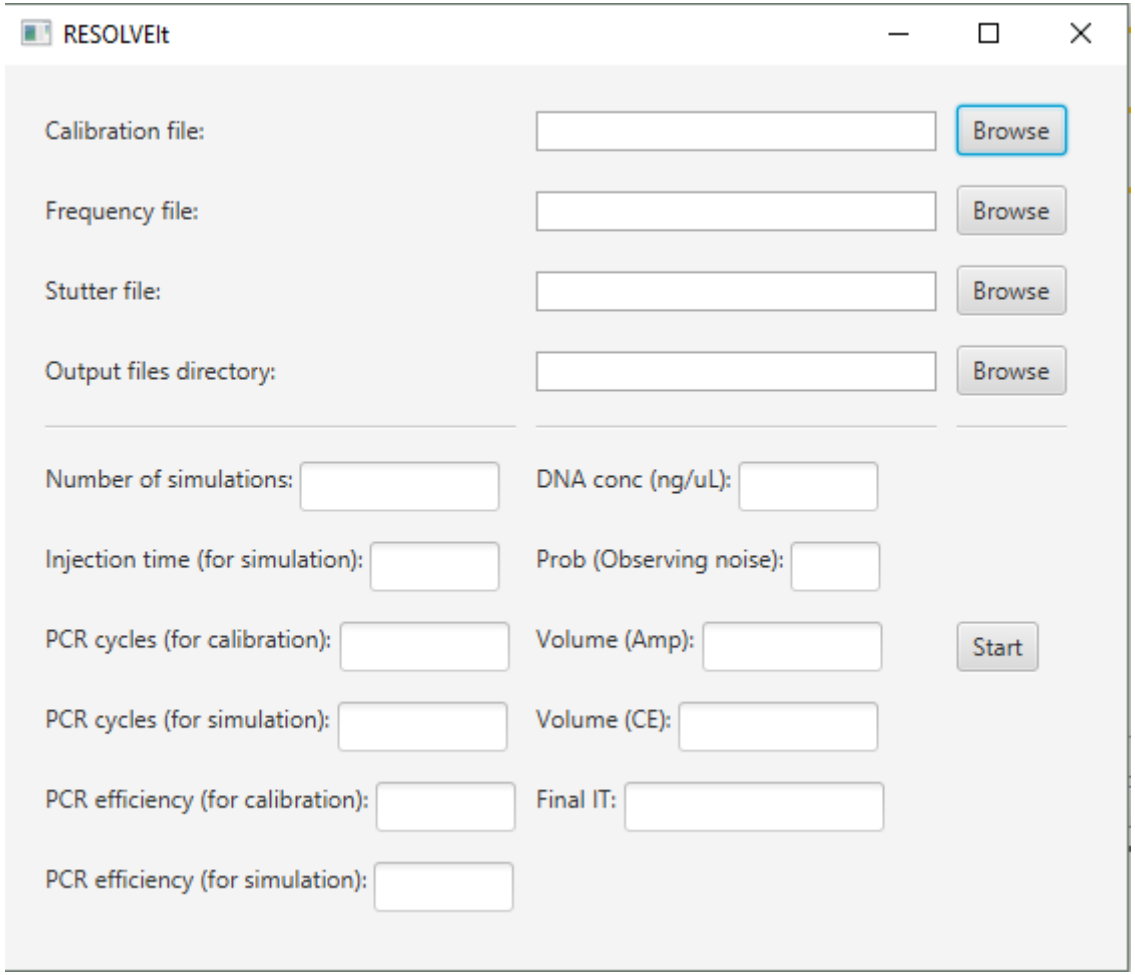
- *In-silico* execution of the forensic DNA analysis process allows for **fast, easy, inexpensive** generation of **representative** large-scale EPG data

- **Quickly evaluate optimal laboratory conditions under various scenarios** &

- Improve detection rates:
  - Determine optimal **AT** to minimize Type I and Type II error rates

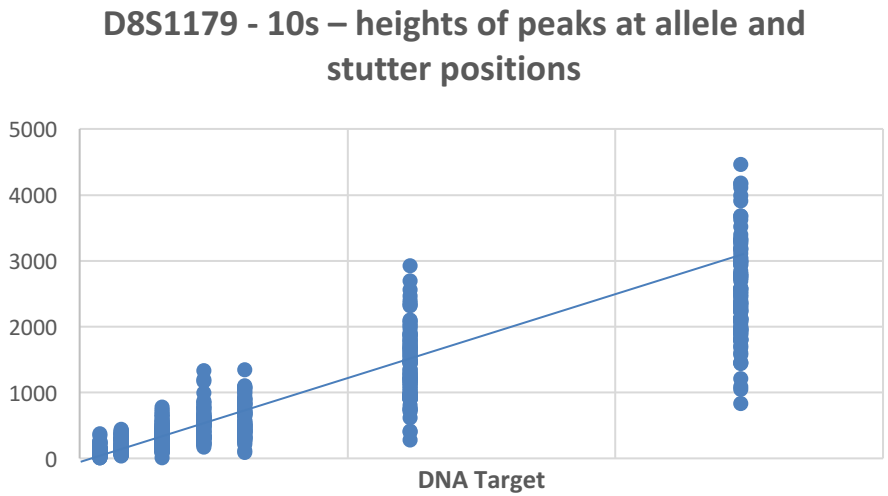
- Improve signal-to-noise resolution:
  - Explore optimal values for parameters such as **number of PCR cycles, time of injection**, etc.

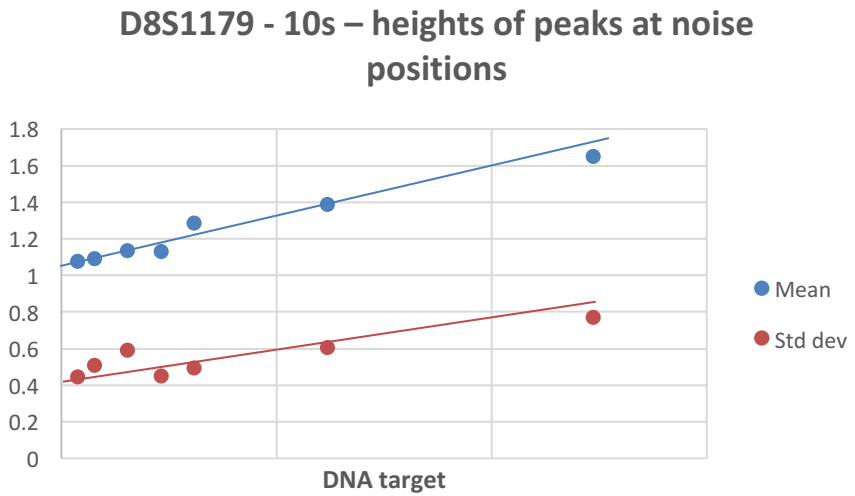
RESOLVEIt: Resolve Evidentiary Signal by Optimizing Laboratory's Validation &



Step I: Parametrization

- **Laboratory-specific data**: Large number of single source samples of known genotypes at different targets and injection times
- Calculate **CE sensitivity** α
  - Describes increase in signal wrt target concentration of DNA
- Calculate **noise parameters - mean** μ **and std dev** σ - at each target concentration, assuming a lognormal distribution*





*Probabilistic characterisation of baseline noise in STR profiles, Monich et al, Forensic Science International Genetics 19 (2015) 107-122. &

Step II: Simulation of signal and noise* &

- Signal is simulated starting from a **single target copy number**
- Noise is simulated from a **specified target mass**

Module 1

Genotype sampling

- 2 alleles are sampled at each locus
- N_0 = 1 for each allele

Module 2

DNA amplification (PCR)

- At each of the $\mathbf{N_{PCR}}$ cycles:
  - **Amp efficiency:**
    Each True amplicon:
    a) Gets copied or
    b) Doesn't get copied.
  - **Stutter slippage probability:**
    Each copied amplicon:
    a) Causes stutter or
    b) Doesn't cause stutter

Module 3

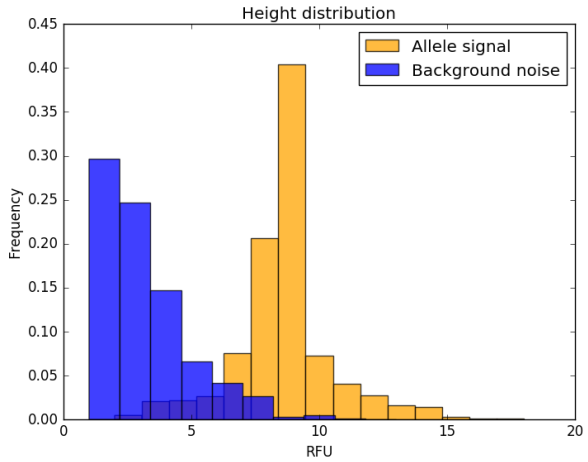
Capillary electrophoresis and fluorescence calculation

- $H_{allele} \sim B\left(N_{N_{PCR}, {}^{V_{CE}}/_{V_{AMP}}}\right) * \alpha$
- $H_{noise} \sim Lognormal(\mu, \sigma)$

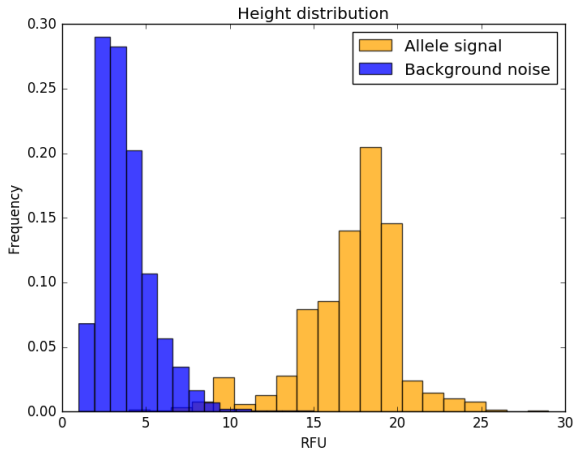
*Exploring STR signal in the single- and multicopy number regimes: Deductions from an in silico model of the entire DNA laboratory process. KR Duffy et al. Electrophoresis 38 (6), 855-868, 2017.

Results: D8S1179 – 1000 simulations &

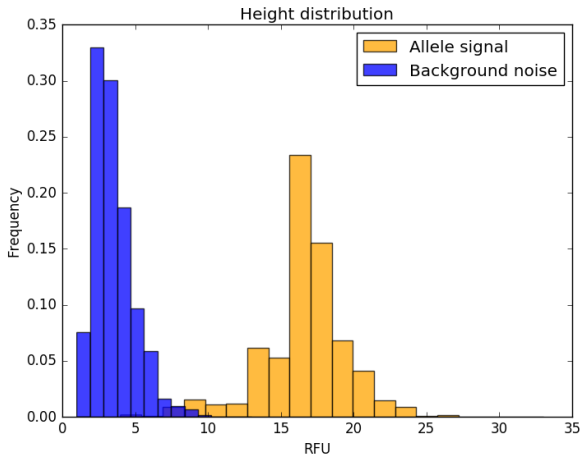
Sim 1 – 5s, 28 cycles &
Allele signal: 1 copy, Background noise: 0.008ng &



Sim 2 – 10s, 28 cycles
Allele signal: 1 copy, Background noise: 0.008ng



Sim 3 – 5s, 29 cycles
Allele signal: 1 copy, Background noise: 0.008ng



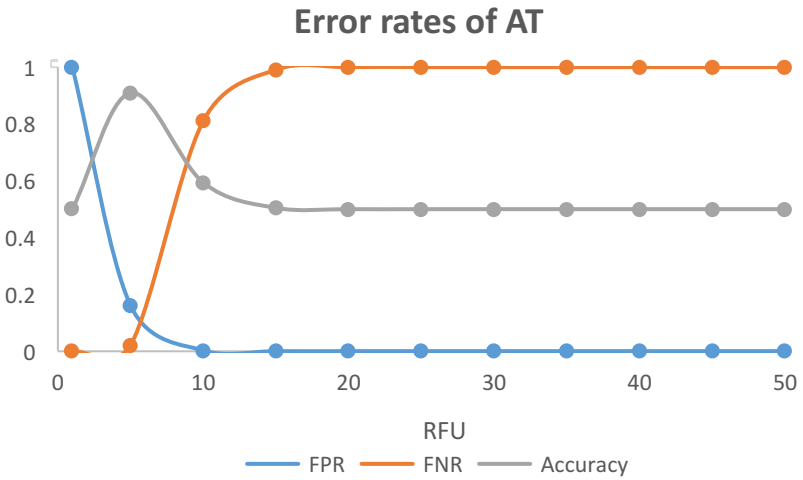
Results: D8S1179 – 1000 simulations &

$$FPR = \frac{FP}{FP + TN}$$

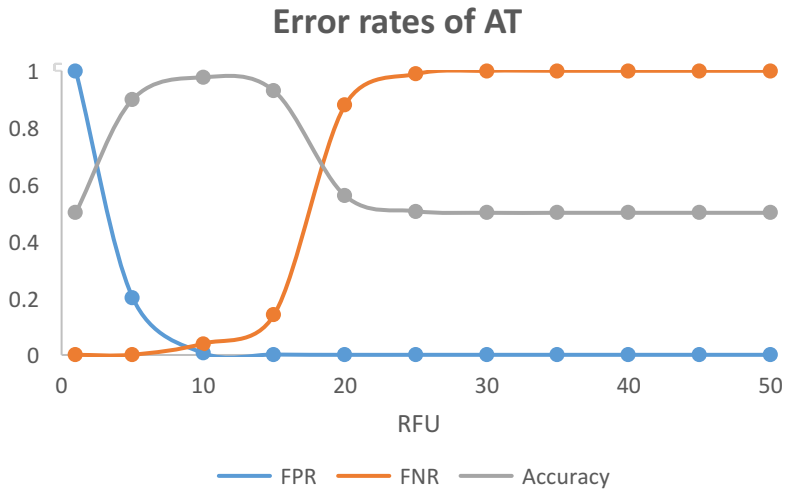
$$FNR = \frac{FN}{FN + TP}$$

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}$$

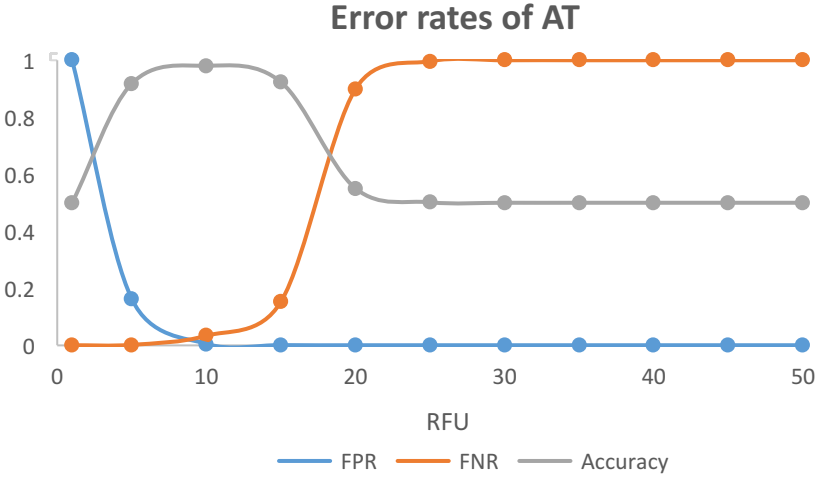
Sim 1 – 5s, 28 cycles &
Allele signal: 1 copy, Background noise: 0.008ng &



Sim 2 – 10s, 28 cycles
Allele signal: 1 copy, Background noise: 0.008ng

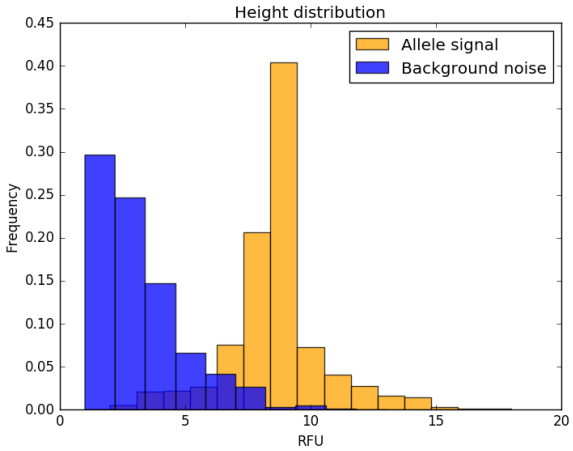


Sim 3 – 5s, 29 cycles
Allele signal: 1 copy, Background noise: 0.008ng

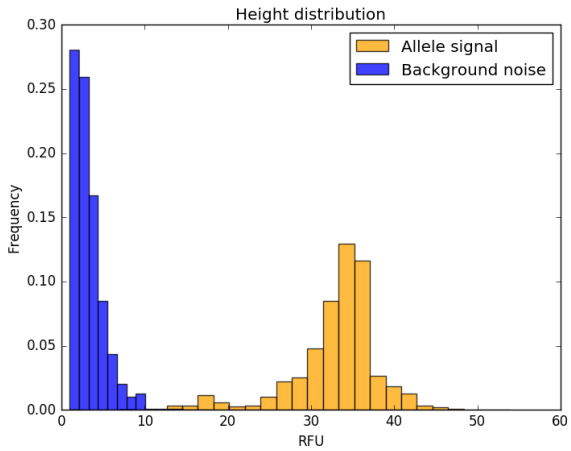


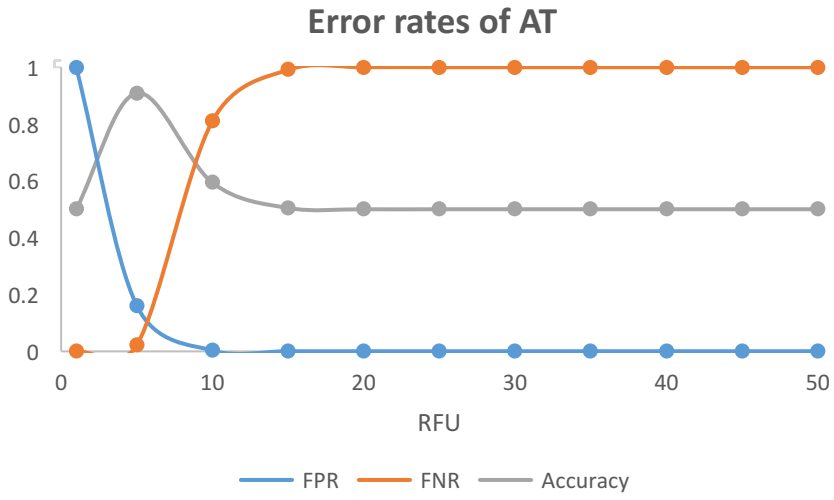
Results: D8S1179 – 1000 simulations &

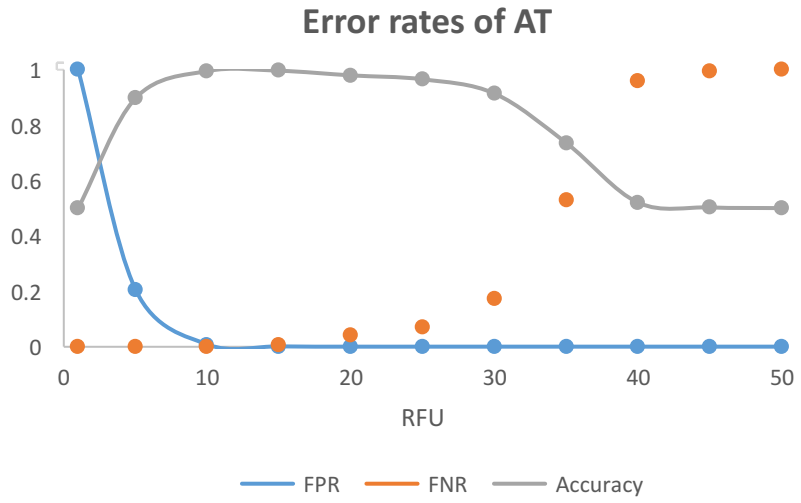
Sim 1 – 5s, 28 cycles
Allele signal: 1 copy, Background noise: 0.008ng



Sim 4 – 10s, 29 cycles &
Allele signal: 1 copy, Background noise: 0.008ng &

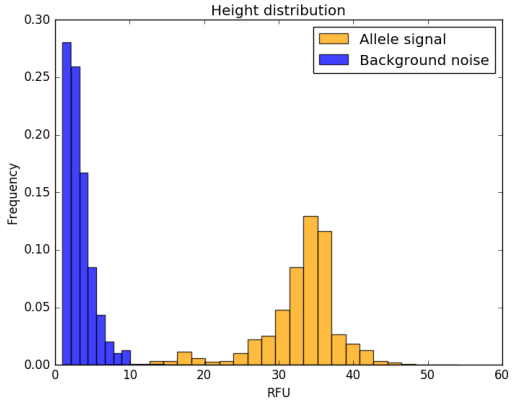




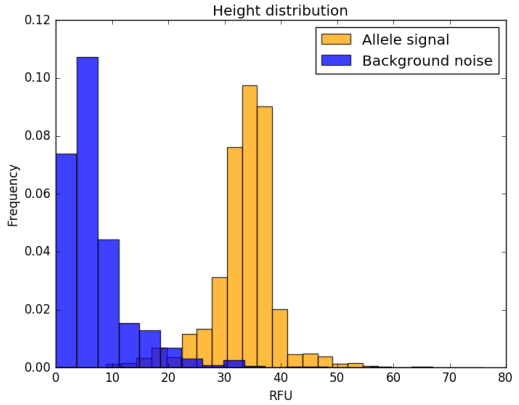


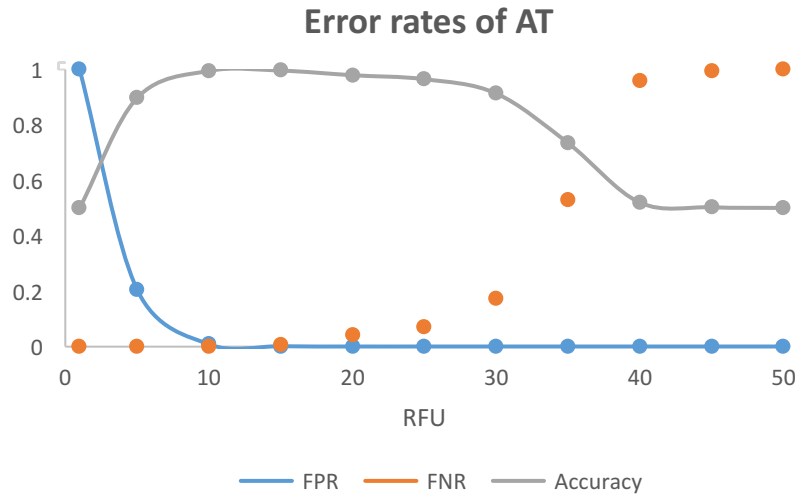
Results: D8S1179 – 1000 simulations &

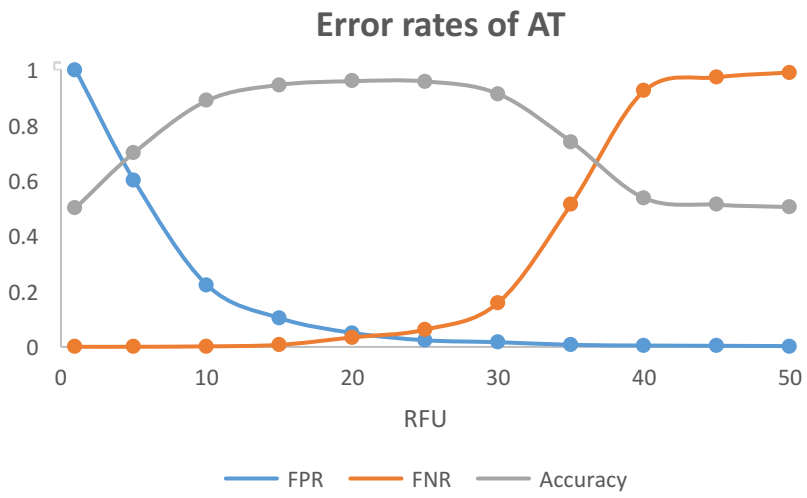
Sim 4 – 10s, 29 cycles
Allele signal: 1 copy, Background noise: 0.008ng



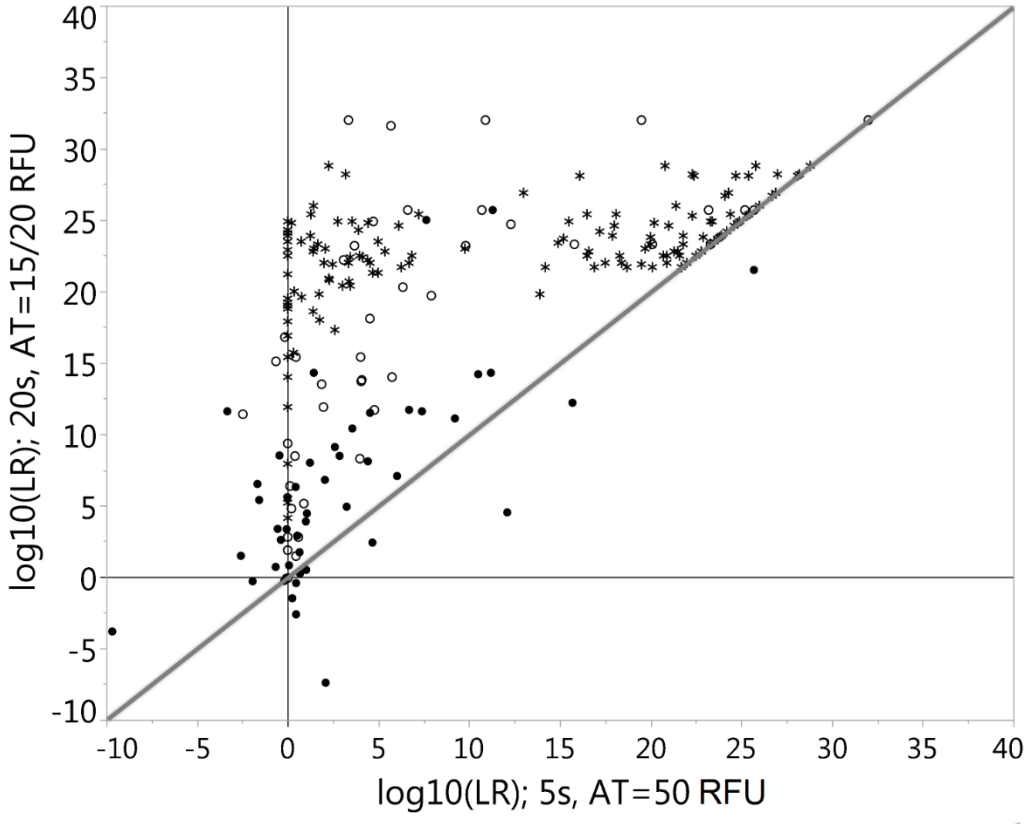
Sim 5 – 10s, 29 cycles &
Allele signal: 1 copy, Background noise: 0.25ng &





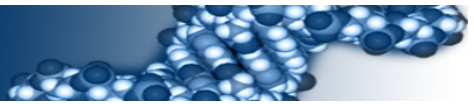


Impact of Information Content on Low-Template Probabilistic Interpretation* &



*Production of High-Fidelity Electropherograms Results in Improved and Consistent DNA Interpretation: Standardizing the Forensic Validation Process, Kelsey C Peters, et al. Forensic Sciences International: Genetics, Submitted.

Conclusions &



- Achieving signal$_{1\text{-copy}}$-to-noise resolution increases information content imported into LR calculation systems
- Choosing a condition-specific AT and laboratory parameters will maximize signal-to-noise resolution while simultaneously minimizing detection error rates
- A combined experimental & simulation-based approach makes the validation process fast and inexpensive