DNA Mixture Analysis:

Principles and Practice of Mixture Interpretation and Statistical Analysis Using the SWGDAM STR Interpretation Guidelines

Developing Thresholds, Protocols and Validation Studies Using the New SWGDAM Guidelines

Joanne B. Sgueglia

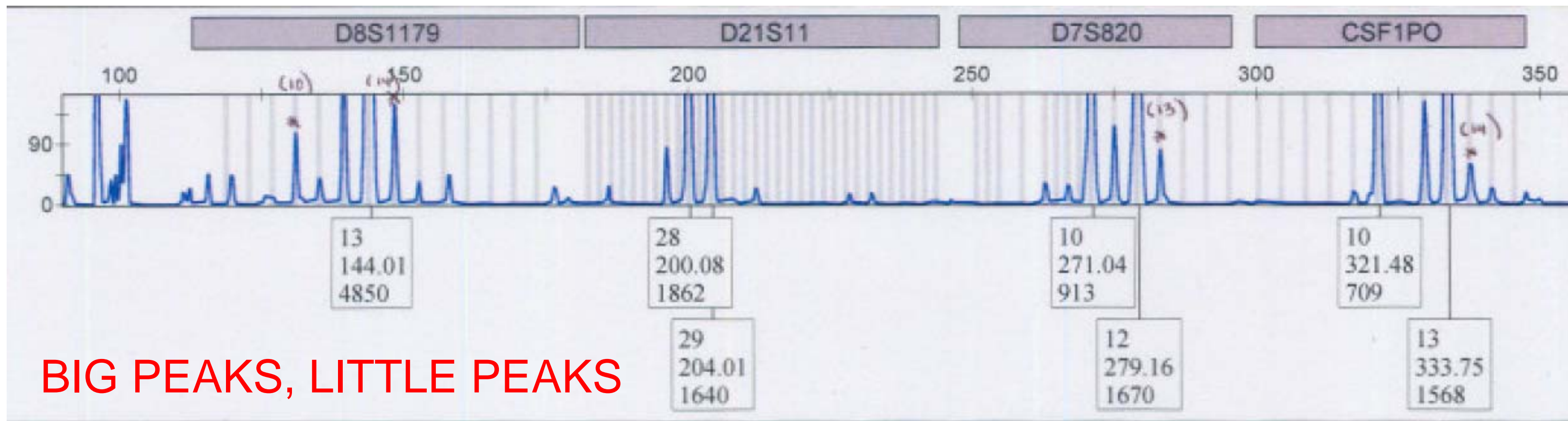
MA State Police Crime Laboratory

NIST

AAFS 2011 Workshop #17

Chicago, IL

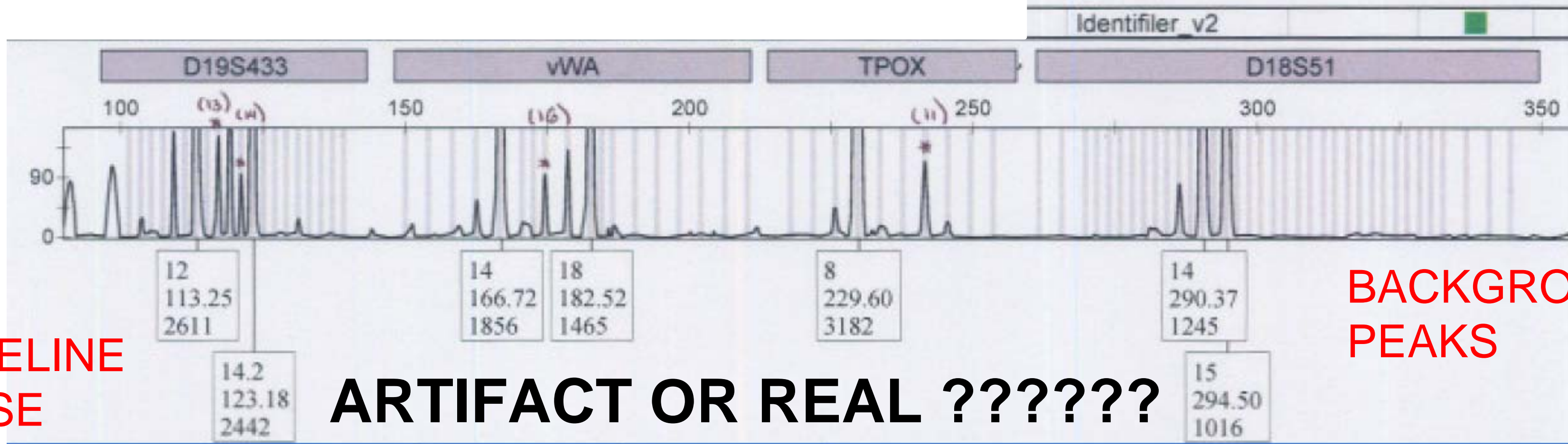
February 22, 2011



BIG PEAKS, LITTLE PEAKS

PULL UP PEAKS,
SPIKE PEAKS

STUTTER PEAKS,
DYE PEAKS



BACKGROUND
PEAKS

BASELINE
NOISE

ARTIFACT OR REAL ??????

SENSITIVITY (THRESHOLDS)

- Minimum Threshold (Section 3.1 Non-Allelic Peaks)
  - Must assess the low end of the scale to differentiate artifactual peaks from DNA peaks or detection of alleles (i.e. non-allelic data from alleles).
  - SWGDAM designates as an ANALYTICAL THRESHOLD.

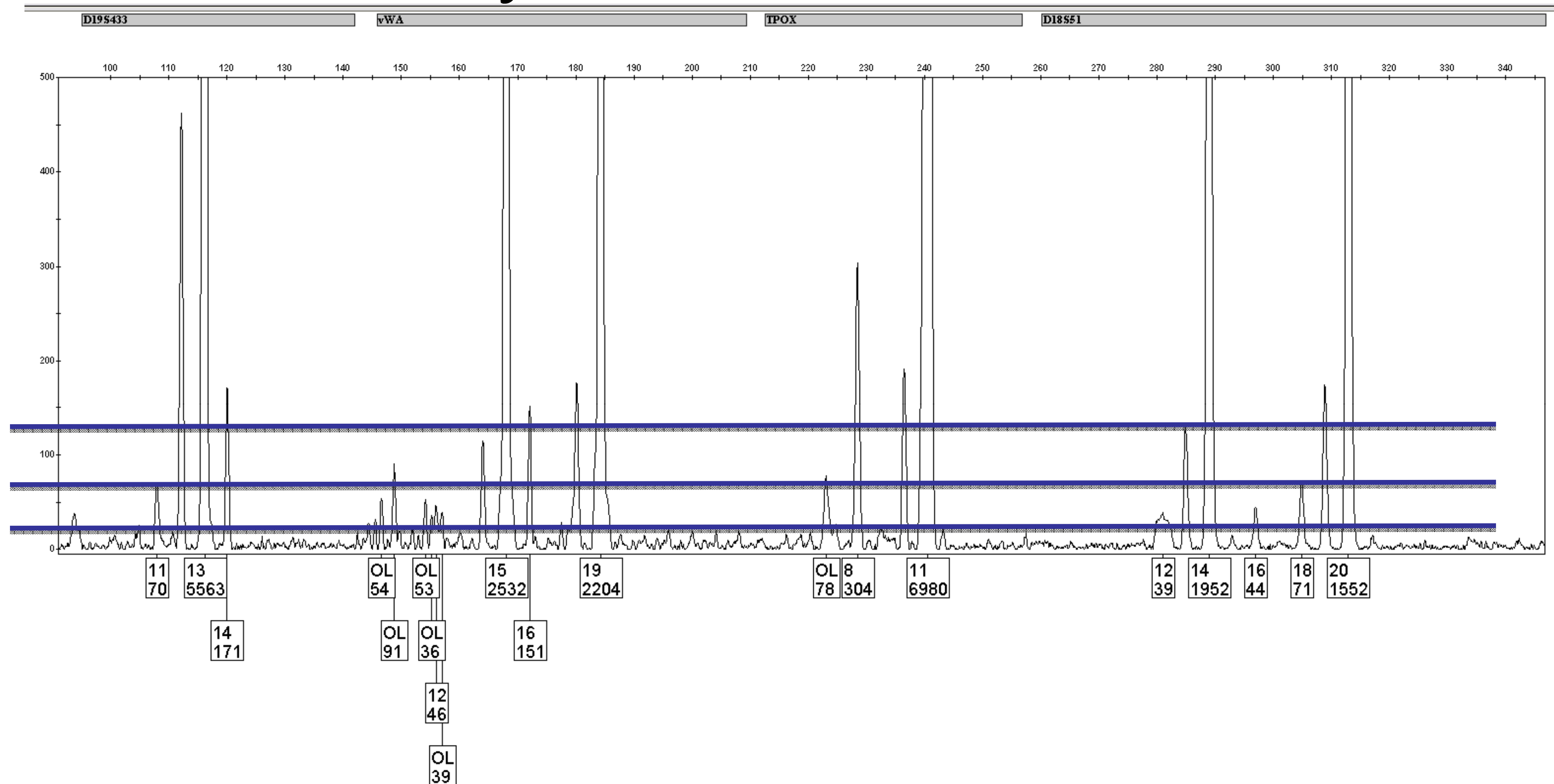
- Peak Height Thresholds for Allelic Peaks (Section 3.2)
  - Must consider stochastic effects and assess if all of the DNA typing information has been detected for a given sample.
  - SWGDAM designates as a STOCHASTIC THRESHOLD.

- Maximum Threshold (Section 3.1.1.3)
  - The laboratory establishes guidelines addressing off-scale data.
  - Saturated data should not be used in quantitative aspects of interpretation (e.g. stutter and peak height ratio assessments).

Maximum Threshold

- **Pull up, Nonspecific product, stutter, balance**
- **Saturation effects of overloading the CCD camera can interfere with evaluation of stutter and other artifacts.**
- **Usually at approximately 8100 rfu on CE instruments**
- **Note that pull up occurs prior to full saturation (dependent on matrix/spectral)**

Analytical Thresholds



- This is the lower threshold set when using dual thresholds and may be called different things in different laboratories (e.g. PAT-peak amplitude threshold, detection threshold or noise threshold).

Different Thresholds of Detection Influence Allele Calls

J Forensic Sci, January 2007, Vol. 52, No. 1
doi:10.1111/j.1556-4029.2006.00318.x
Available online at: www.blackwell-synergy.com

TECHNICAL NOTE

Jason R. Gilder,[1] M.S.; Travis E. Doom,[2] Ph.D.; Keith Inman,[3] M. Crim.; and Dan E. Krane,[4] Ph.D.

Run-Specific Limits of Detection and Quantitation for STR-based DNA Testing

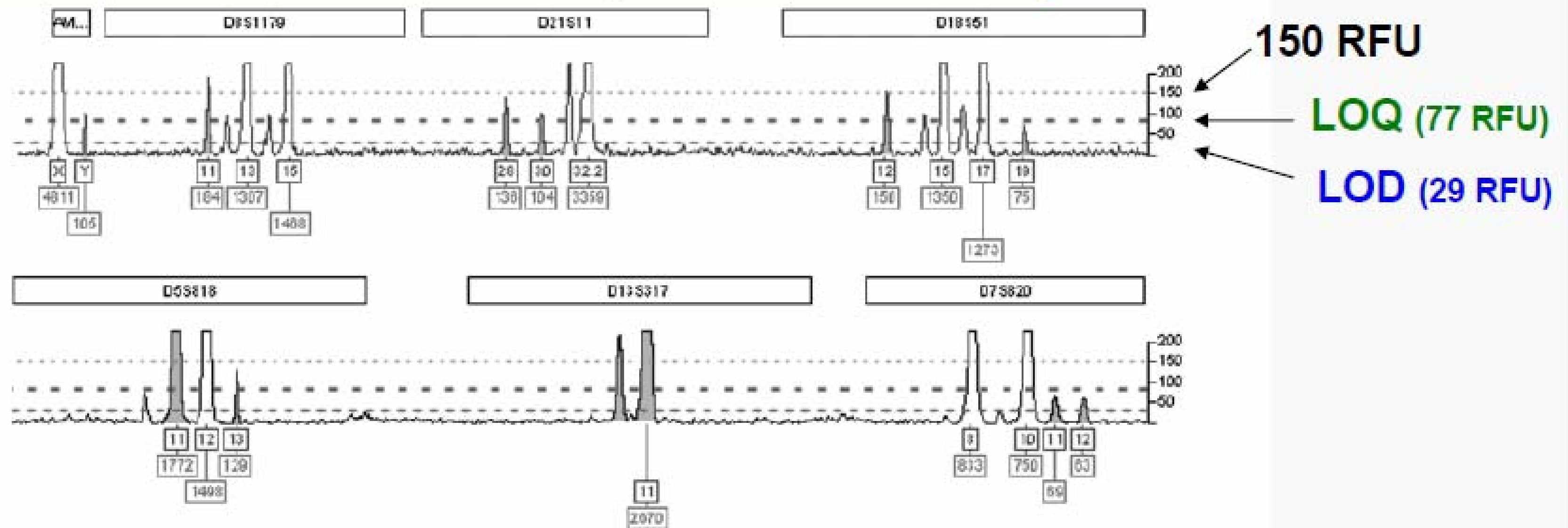


FIG. 2—*Electropherograms from an approximately 10:1 mixture of two reference samples. Three different thresholds are shown: a minimum peak height threshold at 150 relative fluorescent unit (RFU) (dotted line); a limit of quantitation (LOQ) threshold determined to be at 77 RFUs from the negative control for this electrophoresis run (dashed line); and a limit of detection (LOD) threshold determined to be at 29 RFUs for this electrophoresis run (small dashed line). Genotyper® assigned allele calls (with ABI stutter filters in place) are shown in boxes immediately below the electropherogram peaks while peak heights (in RFUs) are shown in boxes below those labels for all peaks with heights greater than the LOD. Peaks consistent with the known profile of the minor contributor are shaded.*

Gilder, J.R., Doom, T.E., Inman, K., Krane, D.E. (2007) Run-specific limits of detection and quantitation for STR-based DNA testing. *J. Forensic Sci.* 52(1): 97-101.

Section 1. Preliminary Evaluation of Data

- An analytical threshold defines the minimum height requirement at and above which detected peaks can be reliably distinguished from background noise.
- 1.1. Analytical threshold: The Laboratory should establish an analytical threshold based on signal-to-noise analyses of internally derived empirical data.
- 3.1.1.2. While the application of an analytical threshold may serve to filter out some non-allelic peaks, the analytical threshold should be established based on signal-to-noise considerations (i.e., distinguishing potential allelic peaks from background). The analytical threshold should not be established for purposes of avoiding artifact labeling as such may result in the potential loss of allelic data.

ARTIFACTS

- PCR
  - Stutter
  - Non-specific product
  - Split peaks
  - Y primer site mutation
- STR
  - Pull Up (Bleed through)
  - Spikes
  - Dye Blobs
- STR runs
  - Bubbles
  - Waves
  - Mis-shapen peaks

Artifacts

- Recognize, define, identify, characterize and edit out accordingly
  - Pull Up/Bleed through
  - Raised Baseline
  - Stutter Peaks
  - Split Peaks (-A)
  - Spikes
  - Shoulders
  - Dye Blobs

How to set an analytical threshold (AT)?

- Various Approaches
  1. **SWGDAM**: e.g. based on two times the intensity difference between the highest peak and lowest trough within the instrumental noise data.
  2. **AB**: Look at all negatives and create a limit of detection (LOD using average + 3 SD) and limit of quantitation (LOQ using average + 10 SD) for instrument noise. May then need to consider threefold the LOQ to set the AT.
  3. **Promega workshop**—5 methods presented with 3 using negatives and 2 using positives/titration series (see STRbase).
  4. **MA State Police**—set to be above all baseline noise from the instrument and other defined and undefined artifacts that present as background peaks (i.e. pull up, spikes, non-specific product) using an amplification positive control.

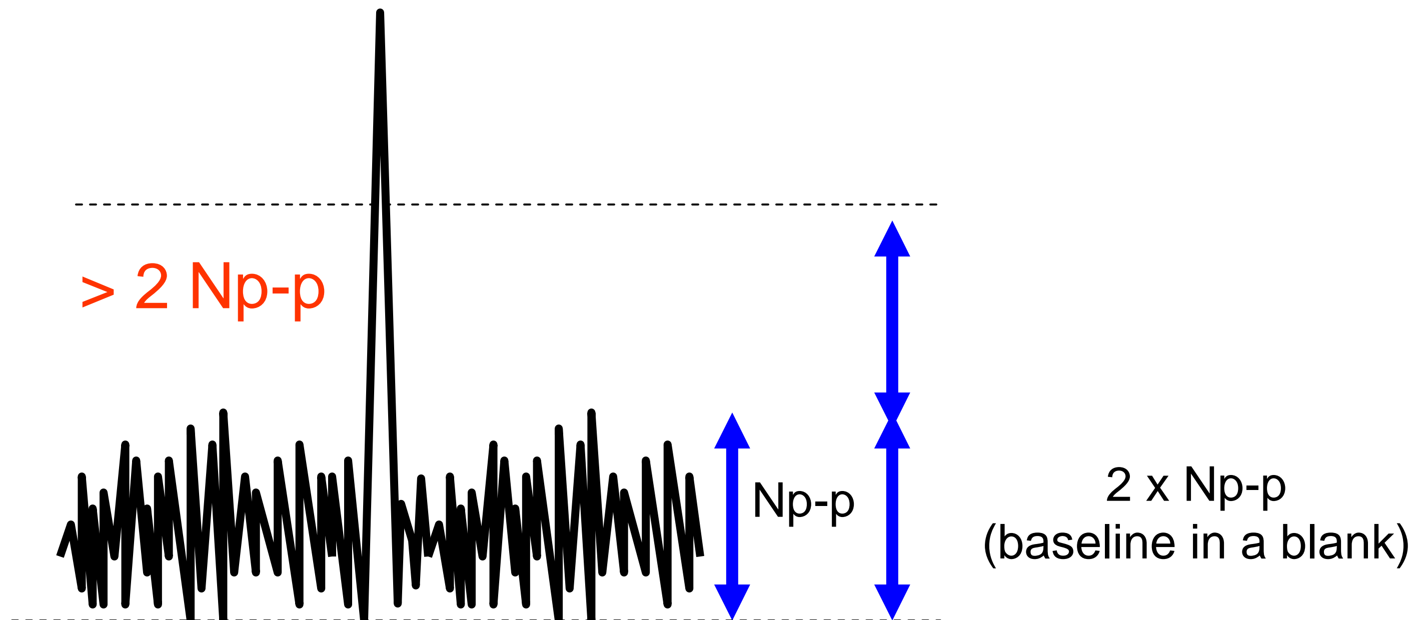
1. **SWGDAM**

Limit of Detection (LOD)

- Typically 3 times the signal-to-noise (based on standard deviation of the noise) or 2x Np-p

Is this peak real?

Yes, it is a peak but you cannot rely on it for concentration determinations as it is not >10 S/N



2. Applied Biosystems

ID Direct

Table 1. Identifiler Direct Minimum Threshold Data (26 amplification cycles)

Dye	Maximum RFU	Average RFU	Standard Deviation (SD)	Average RFU + 3 SD (LOD)[1]	Average RFU + 10 SD (LOQ)[2]
Blue (6-FAM™)	13	1.51	0.72	3.66	8.69
Green (VIC®)	12	1.76	0.76	4.05	9.41
Yellow (NED™)	16	3.18	1.07	6.41	13.92
Red (Pet®)	22	3.70	1.25	7.45	16.19
Orange (LIZ®)	9	1.77	0.84	4.29	10.15

[1] LOD = Limit of detection
[2] LOQ = Limit of quantification

Table 2. Identifiler Direct Minimum Threshold Data (27 amplification cycles)

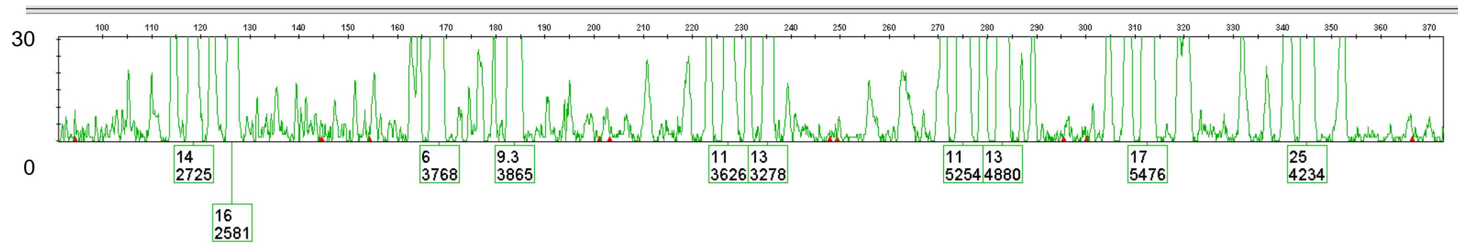
Dye	Maximum RFU	Average RFU	Standard Deviation (SD)	Average RFU + 3 SD (LOD)[1]	Average RFU + 10 SD (LOQ)[2]
Blue (6-FAM™)	15	1.78	1.11	5.11	12.88
Green (VIC®)	14	1.94	0.86	4.53	10.56
Yellow (NED™)	15	3.39	1.16	6.86	14.95
Red (Pet®)	14	3.92	1.35	7.97	17.43
Orange (LIZ®)	8	1.56	0.74	3.78	8.98

[1] LOD = Limit of detection
[2] LOQ = Limit of quantification

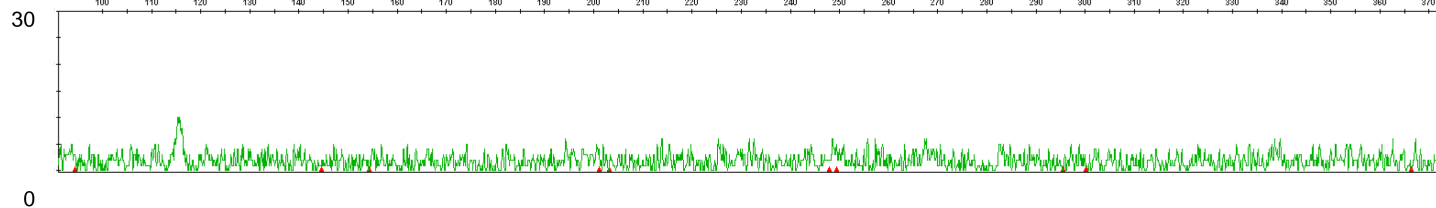
3. Promega Workshop

Method	Origin	Analytical Threshold for green 5s injection example
1	Negatives	7
2	Negatives	4
3	Negatives	20
4	DNA Series	31
5	DNA Series	39

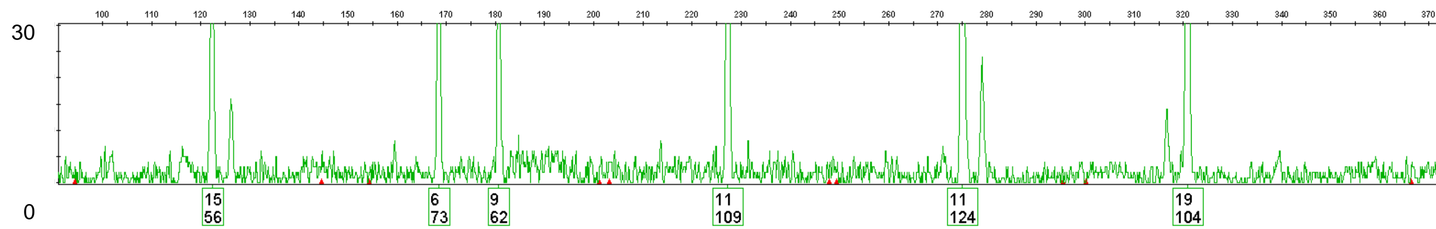
High input of DNA

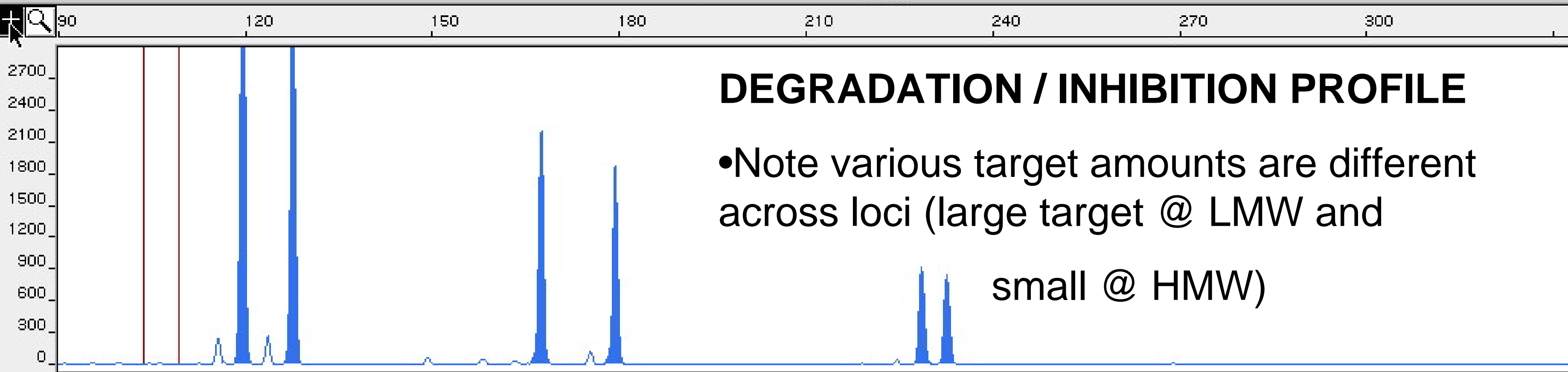


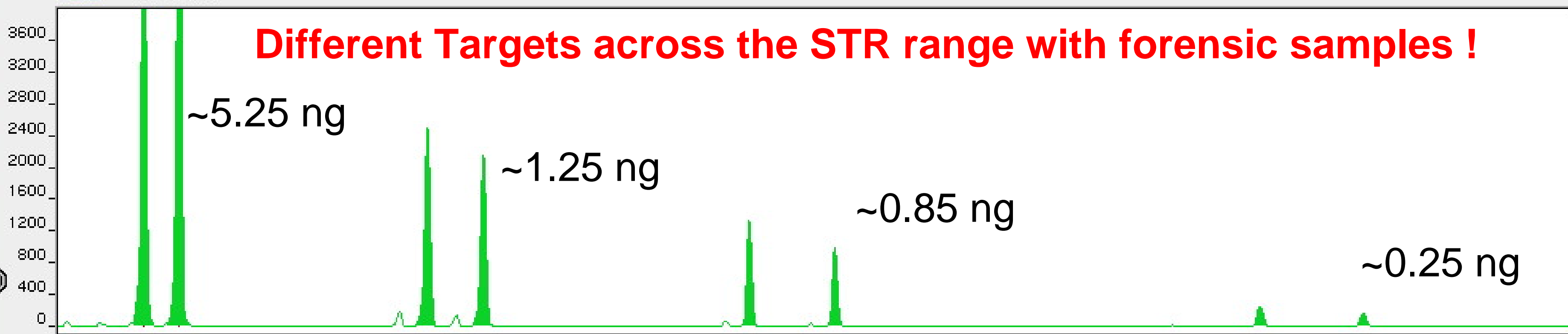
Neg amp control

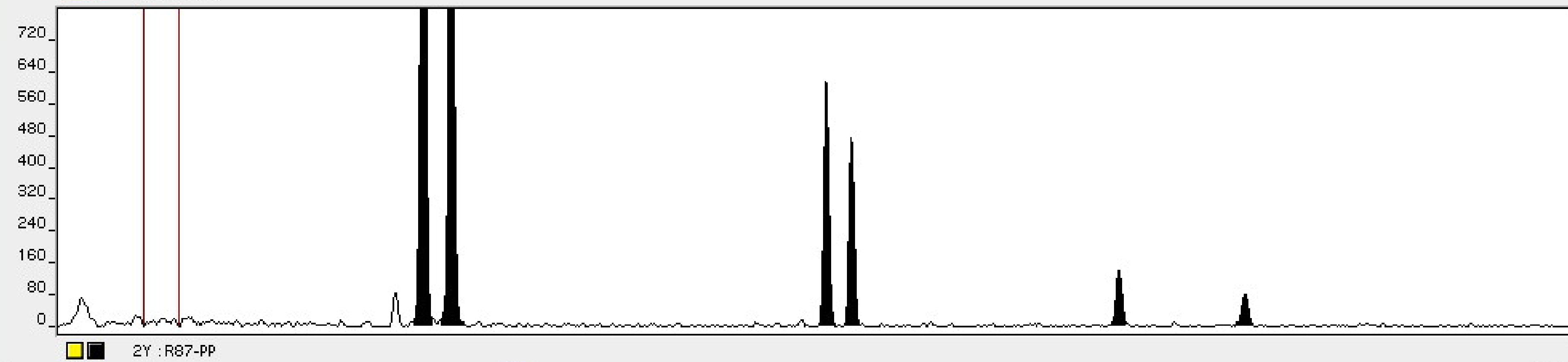


Low input of DNA




File Edit Project Sample Settings View Windows Help
Sat 4:47 PM
310 GeneScan® 3.1
degraded/inhib Display-1
90
120
150
180
210
240
270
300
2700
2400
2100
1800
1500
1200
900
600
300
0
DEGRADATION / INHIBITION PROFILE
•Note various target amounts are different across loci (large target @ LMW and small @ HMW)
2B : R87-PP


3600
3200
2800
2400
2000
1600
1200
800
400
0
Different Targets across the STR range with forensic samples !
~5.25 ng
~1.25 ng
~0.85 ng
~0.25 ng
2G : R87-PP

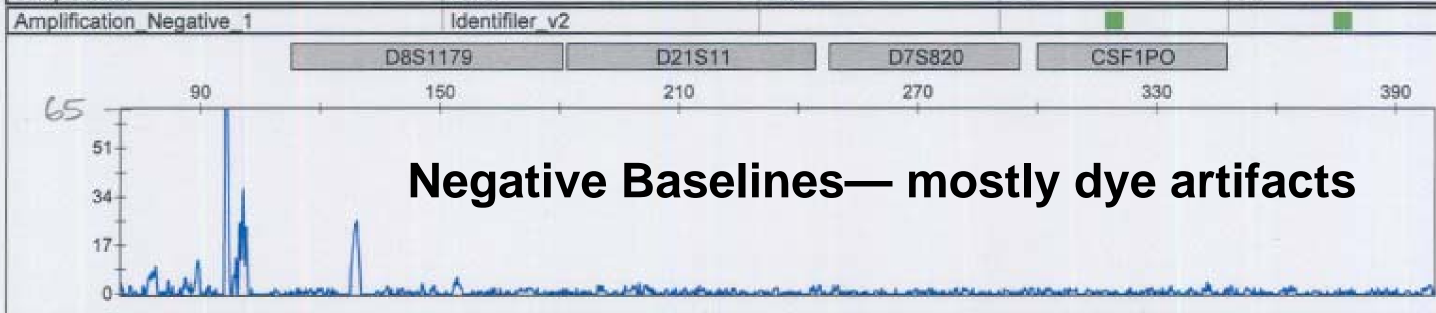

720
640
560
480
400
320
240
160
80
0
2Y : R87-PP
X:
Y:

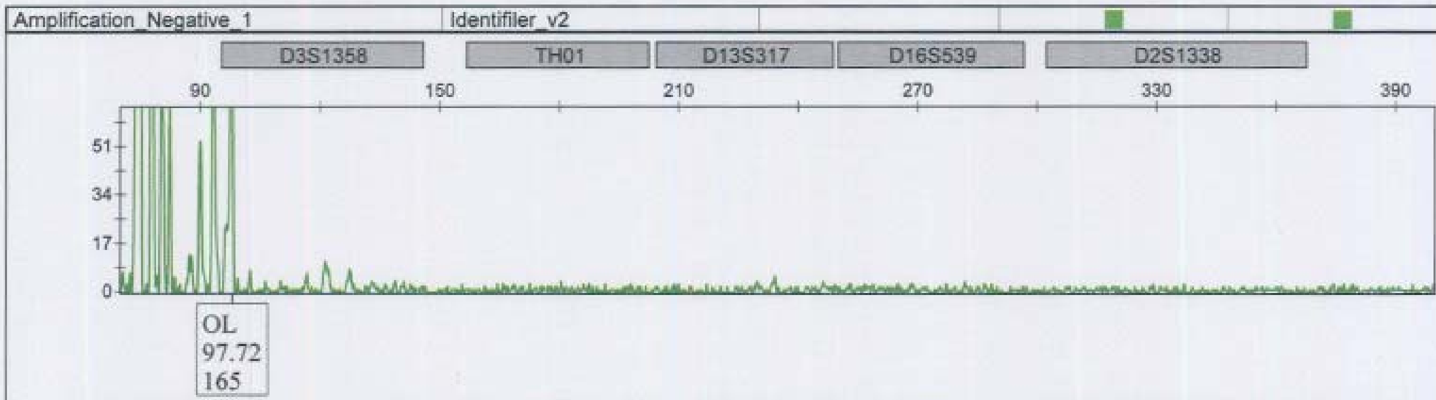
4. Massachusetts State Police

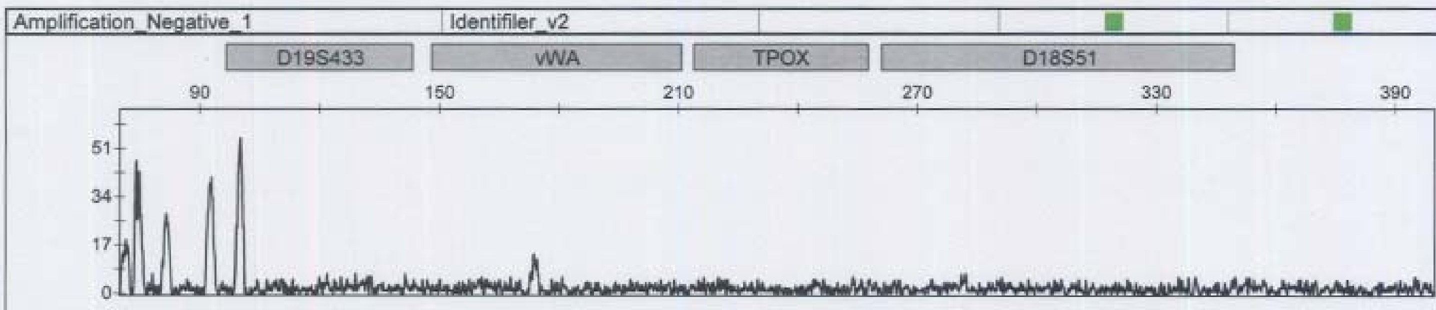
Recommend using Positives

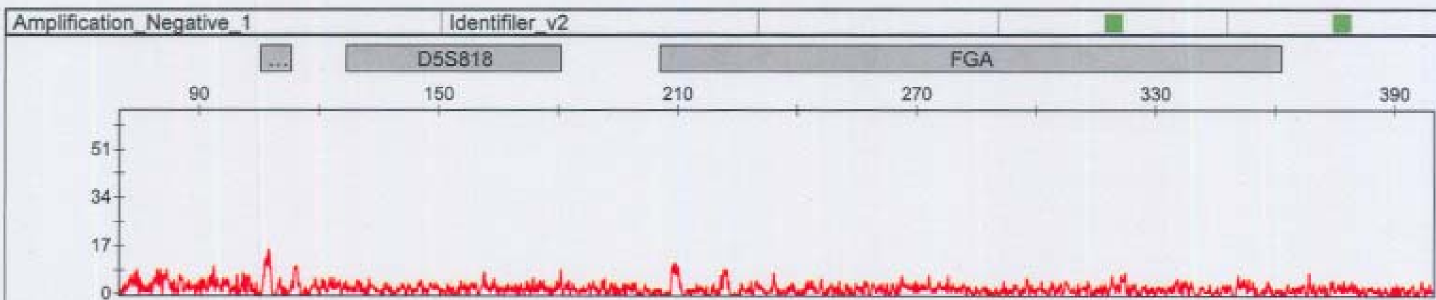
- MSP—used amplification positives in the past to represent the approximate target DNA used for samples.
- Note that this DNA would normally have peak heights in the ‘sweet zone’ or at about 1000 to 3000 rfu on a CE instrument.
- Assess all peaks at low end and try to define the artifacts or mark those that are undefined (i.e. non-specific product).
- Visually set the AT to compensate for the low level artifacts, especially for those that are undefined and could not be distinguished as artifact or real DNA.

Negative Baselines— mostly dye artifacts









Sample Name	Panel	SQO	OS	SQ
Amplification_Positive_1	Identifiler_v2			

D8S1179 D21S11 D7S820 CSF1PO

90 150 210 270 330 390

65 51 34 17 0

13
143.81
3751

30
208.01
3082

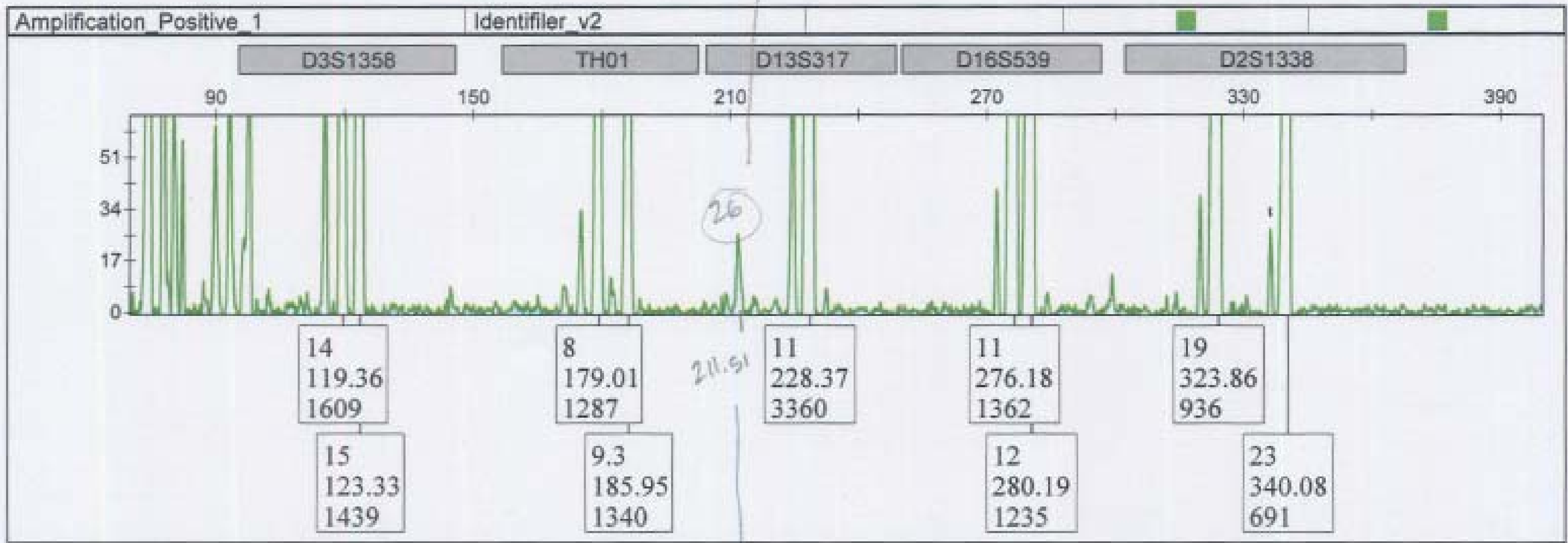
10
271.04
1029

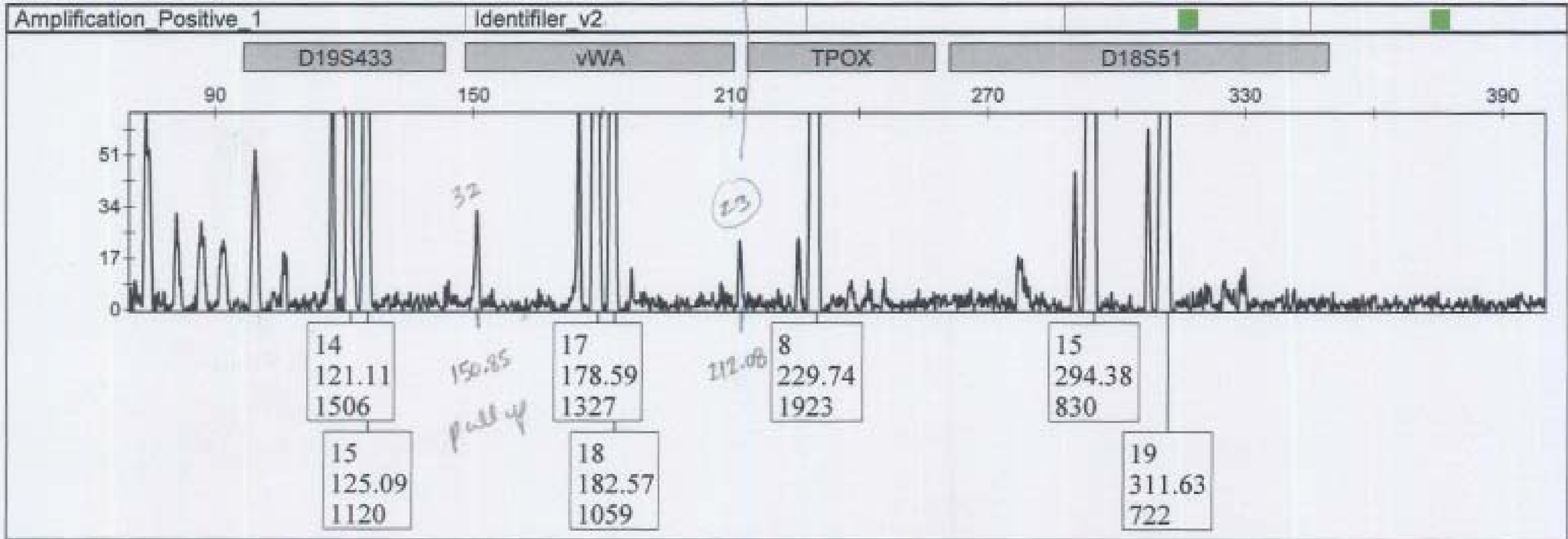
11
275.12
823

10
321.43
1384

12
329.59
1141

Positives- more artifacts observed





Sample Name	Panel	SQO	OS	SQ
Amplification_Positive_1	Identifiler_v2			

...
D5S818
FGA

90 150 210 270 330 390

51
34
17
0

X
106.34
1990

11
151.05
2538

23
238.39
876

24
242.44
836

Analytical Thresholds @ MSP

- MSP has set analytical thresholds using Identifiler based on visual inspection from data on six 3130xl instruments.
- Values range from 40 rfu on two older instruments, 50 rfu on two purchased in 2007 and 55 rfu on two purchased thereafter.
- Need to assess for each kit chemistry at each condition (e.g. injection time, increased cycles, etc.) for each instrument.

Stochastic Thresholds (ST)

- SWGDAM Definition: the value above which it is reasonable to assume that allelic dropout has not occurred within a single-source sample.

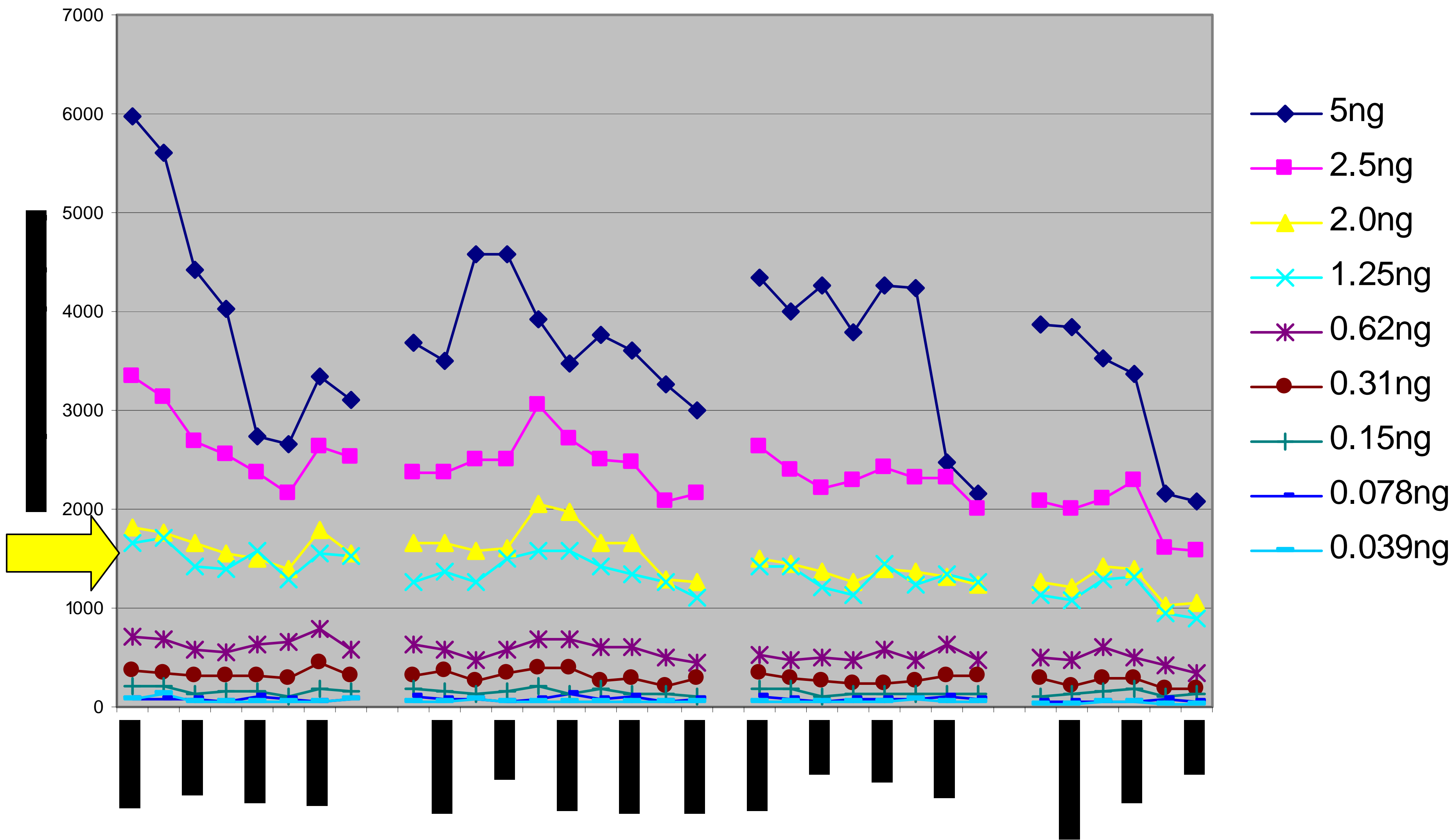
- 3.2.1. The laboratory establishes a stochastic threshold based on empirical data derived within the laboratory and specific to the quantitation and amplification systems (e.g., kits) and the detection instrumentation used. It is noted that a stochastic threshold may be established by assessing peak height ratios across multiple loci in dilution series of DNA amplified in replicate. The RFU value above which it is reasonable to assume that, at a given locus, allelic dropout of a sister allele has not occurred constitutes a stochastic threshold.

MSP Approach-Identifiler

- Titration sets of at least two sources (have used 5 sources for some validations).
- Target DNA using two fold dilution series from 5 ng, 2.5 ng, 1.25 ng …….39 pg.
- Establish AT prior to ST and use as a guide to have ST satisfy two criteria:
  - 1. Three fold the AT
  - 2. Obtain a partial profile at 150pg. This ensures no data is interpreted with statistical weight in the stochastic or low copy number region (gray zone).

IDENTIFILER (Optimal Target of 2 ng)

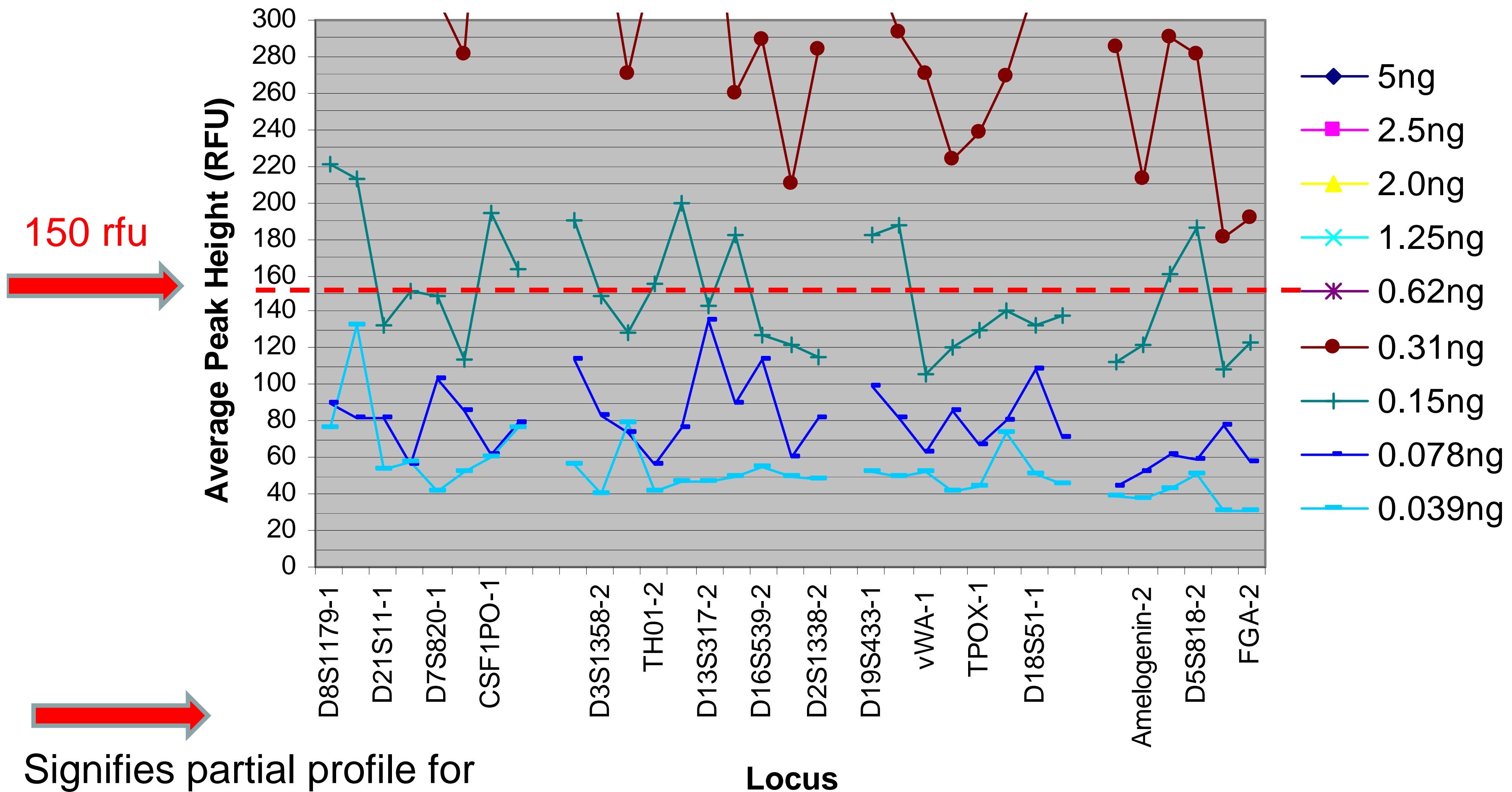
CE011 Titration Sets: Average of All Sets and Replicates



IDENTIFILER

Reporting Threshold set at Red Line for partial profile @ 0.15ng & 3X AT (baseline + background)

CE011 Titration Sets: Average of All Sets and Replicates



Signifies partial profile for 150 pg of DNA

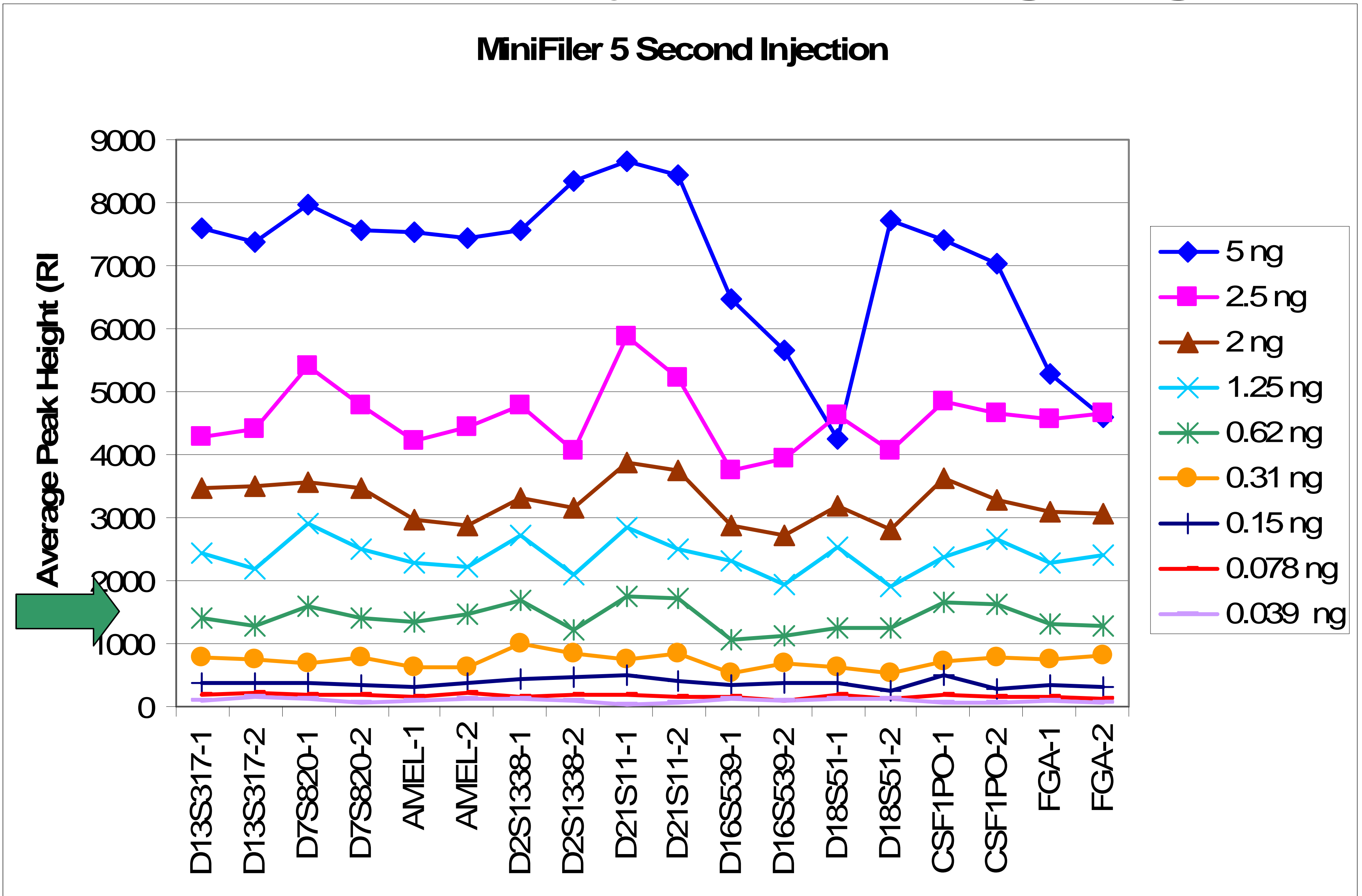
Other Kits and Conditions: MiniFiler and ID Direct

3.2.1.1. If measures are used to enhance detection sensitivity (i.e., allelic height), the laboratory should perform additional studies to establish independent criteria for application of a separate stochastic threshold(s). Such measures may include but not be limited to increased amplification cycle number, increased injection time, and post-amplification purification/concentration of amplified products.

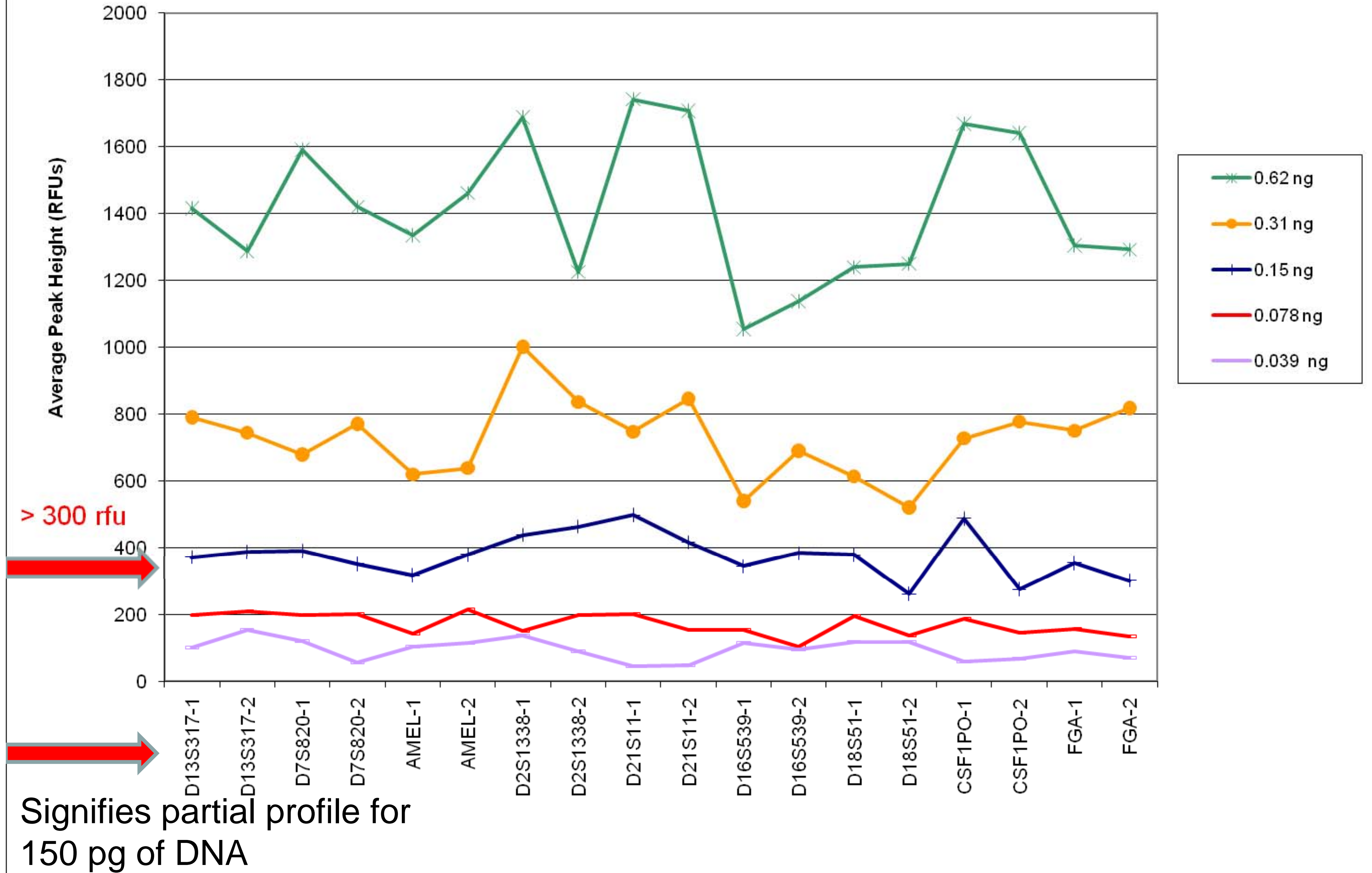
MiniFiler

- Amplification using 30 cycles
- Enhanced Buffer
- 8 loci + Amelogenin
- Much higher peak heights for same target of DNA than produced with ID
- Note that going from ID @28 to MF @30 cycles-- would expect a four fold increase. The optimal target changed from 2ng for ID to 0.5ng for MF.

MiniFiler 5 sec Injection (0.5ng target)

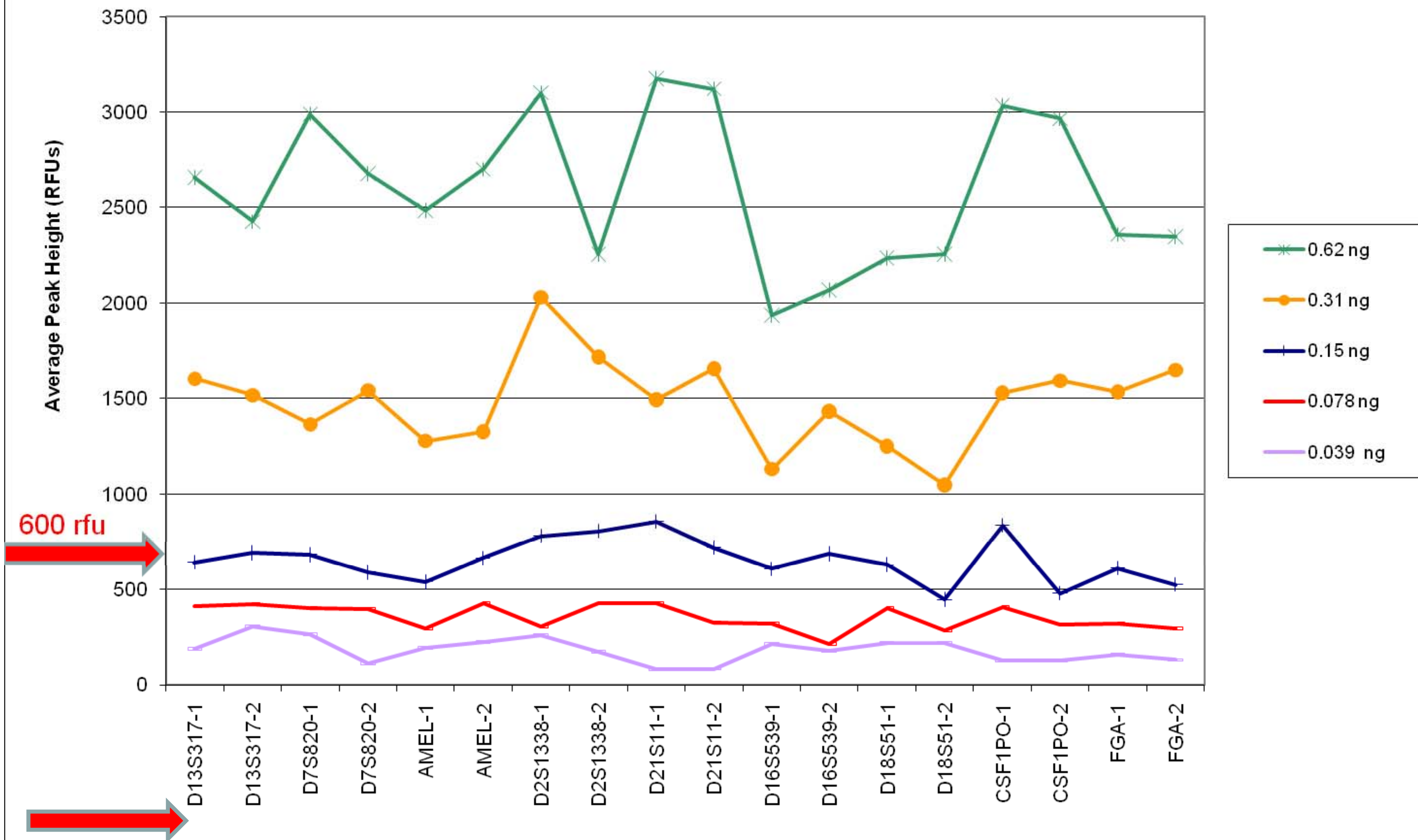


MiniFiler 5 Second Injection



Signifies partial profile for
150 pg of DNA

MiniFiler 10 Second Injection

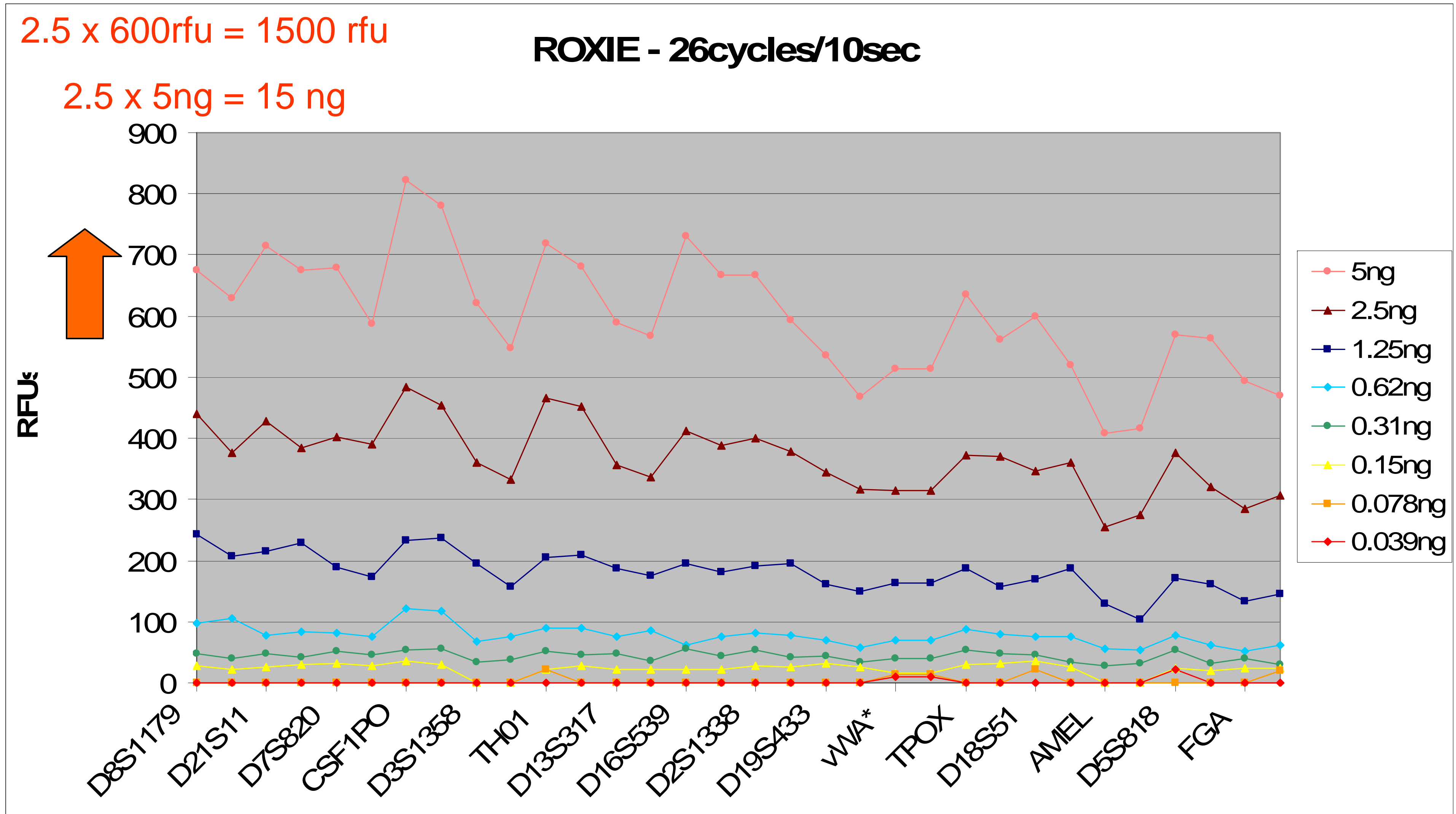


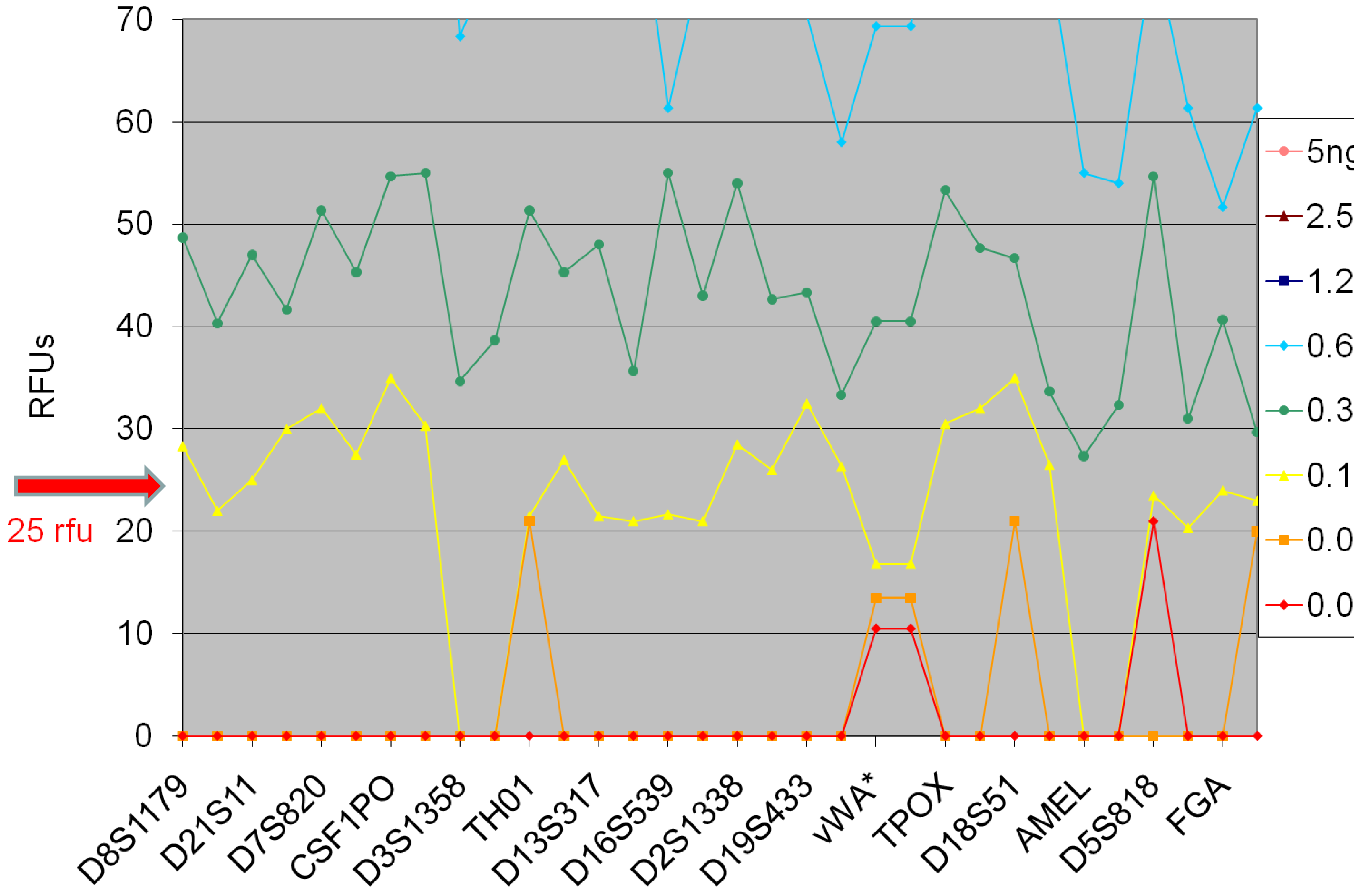
Signifies partial profile for
150 pg of DNA

Identifiler Direct

- Used for direct amplification of blood or buccal on FTA paper.
- Validation encompasses various cycle numbers and injection times to obtain optimal settings for given biological samples.
- Used a titration set to compare sensitivity of ID to IDD.
  - IDD determined to be four fold less sensitive (as expected as comparison of 28 cycles for ID and 26 cycles for IDD).
- Compare low end target of approximately 150 pg with IDD at various parameters to investigate changes in peak heights and considerations for AT and ST.

ID Direct—Optimal Target approximately 15 ng (1.2mm blood punch)





ID Direct
ROXIE - 26cycles/10sec
RFUs
70
60
50
40
30
20
10
0
25 rfu
D8S1179
D21S11
D7S820
CSF1PO
D3S1358
TH01
D13S317
D16S539
D2S1338
D19S433
vWA*
TPOX
D18S51
AMEL
D5S818
FGA
5ng
2.5
1.2
0.6
0.3
0.1
0.0
0.0

ID Direct

ROXIE - 26cycles/20sec

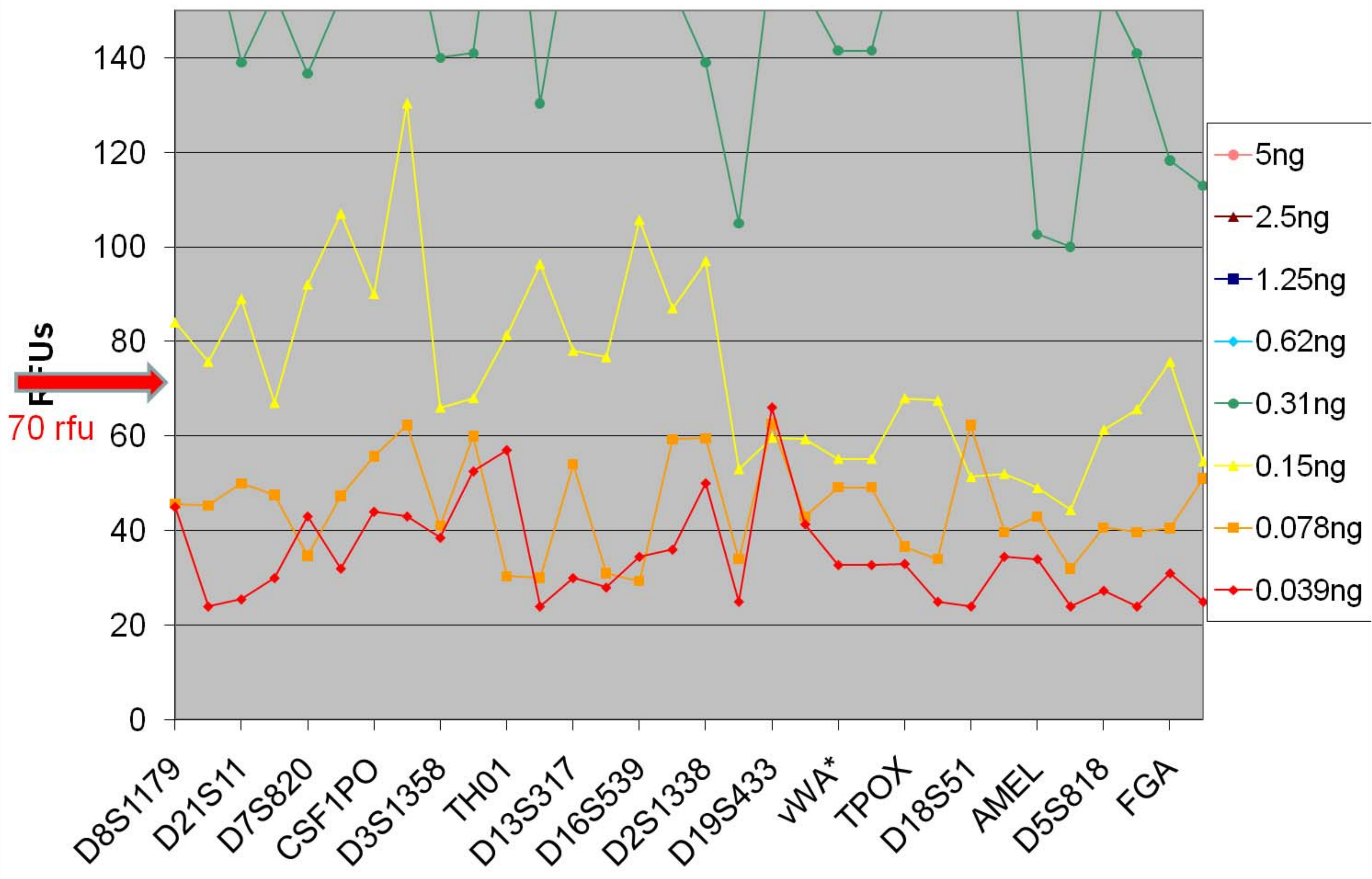
RFUs

38 rfu

70
60
50
40
30
20
10
0

D8S1179
D21S11
D7S820
CSF1PO
D3S1358
TH01
D13S317
D16S539
D2S1338
D19S433
vWA*
TPOX
D18S51
AMEL
D5S818
FGA

5ng
2.5ng
1.25ng
0.62ng
0.31ng
0.15ng
0.078ng
0.039ng



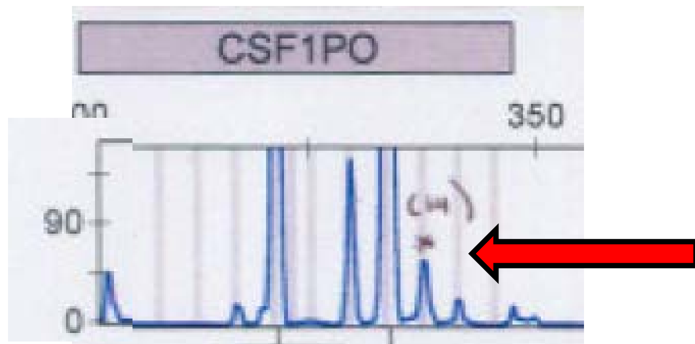
ID Direct
ROXIE - 27cycles/20sec
RFUs
70 rfu
140
120
100
80
60
40
20
0
D8S1179
D21S11
D7S820
CSF1PO
D3S1358
TH01
D13S317
D16S539
D2S1338
D19S433
vWA*
TPOX
D18S51
AMEL
D5S818
FGA
5ng
2.5ng
1.25ng
0.62ng
0.31ng
0.15ng
0.078ng
0.039ng

Sensitivity Studies and Thresholds

- Validation
  - Titration series of range of target DNA in ng (5, 2.5 ------- 0.039) -use @ least two sources
  - Evaluation of rfu values for all targets (both at low end and for high end saturation)
  - Evaluation of signal-to-noise
  - Noise comprised of baseline and background effects

- Guideline
  - Establish criteria for an analytical threshold (AT) and stochastic threshold (ST).
  - ST set to obtain a partial profile @ 0.15ng and to be at 3X of established noise
  - Peaks above ST have met criteria to be given an allele designation and accompanying statistic (to type 0.25 ng of DNA)

Gray Zone Peaks-Alleles between the AT and ST?



- How to utilize data contained in the zone between the AT and ST?
  - AT set to exclude any spurious artifacts from baseline/background and peaks below the AT are not assessed
  - Gray zone-ensure all peaks are assessed and edit out artifacts and mark only if DNA (i.e. exclude pull up, spikes, dyes, etc.)

- May have partial data and drop out of major and/or minor
- Data used for comprehensive view of questioned profile(s) and interpretation of exclusion or inclusion
- No statistic performed at a locus if alleles are in gray zone--Unless assumptions of number of donors and/or quantitative PH information is used.

Stochastic Effects

- Note that stochastic effects are expected with lower targets of DNA (e.g. partial profiles, imbalance, drop out) and need to be addressed.
- These effects can be accounted for by using all of the information available for a comprehensive view and comparative analyses of exclusion or inclusion.
- Conservatively, no statistic would be reported for data below the ST. Hence no weight or significance in a court of law.
- No need to consider this information inconclusive. Know your system and how DNA behaves to not have false exclusions or inclusions.

Interpretation Guideline Excerpt -use all information—COMPREHENSIVE PROFILE

Each locus must be analyzed individually to differentiate and quantify peaks in order to distinguish the profiles observed in the mixture. However, the combined profile can be utilized to provide a comprehensive view of the mixture possibilities and ratios of major to minor components.

Acknowledgements

- AAFS
- John Butler and Mike Coble and co-instructors.
- Dr. Catherine Grgicak (Promega-AT methods)
- MA State Police DNA Unit
- Graphs- Jennifer Rogean, Leanna Farnam & Dr. Sandra Haddad
- All DNA analysts working with difficult mixtures.