21st International Symposium on Human Identification
Mixture Interpretation Workshop:
Principles, Protocols, and Practice
October 11, 2010 – San Antonio, TX

Analytical Thresholds: Determination of Minimum Distinguishable Signals

Catherine M. Grgicak

Outline for Analytical Thresholds

- **PRINCIPLES**
  - SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories
  - IUPAC (International Union of Pure & Applied Chemists) definitions
  - Other definitions
- **PROTOCOLS**
  - Review 5 approaches used to calculate the Analytical Signal (a.k.a Minimum Distinguishable Signal or RFU Threshold)
- **PRACTICE**
  - Example Calculations
  - Available web, literature and software resources

21st International Symposium on Human Identification
Mixture Interpretation Workshop:
Principles, Protocols, and Practice
October 11, 2010 – San Antonio, TX

PRINCIPLES

PRINCIPLES

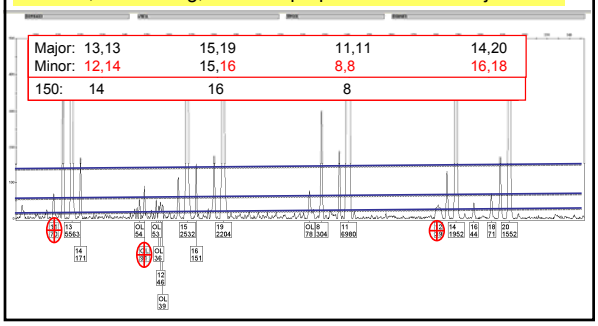
SWGDAM Interpretation Guidelines

- 1.1. "*Analytical threshold: The Laboratory should establish an analytical threshold based on signal-to-noise analyses of internally derived empirical data. As an example, an analytical threshold may be based on two times the intensity difference between the highest peak and lowest trough within the instrumental noise data. Other scientific methods may be used"*
  - What are these 'Other scientific methods'?
- 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"*
  - How does one determine analytical threshold?

PRINCIPLES

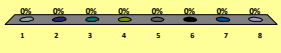
Implications Different Thresholds Have on Analysis of Mixtures

2 males, 1:19 at 2ng, 1ul 3130 prep volume and 5s injection



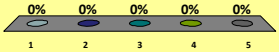
Question: What is your laboratory's analytical threshold?

1. <49 RFU
2. 50 RFU
3. 51-75 RFU
4. 76-100 RFU
5. 101-150 RFU
6. >150 RFU
7. I do not work in a lab
8. I have absolutely no idea! I just press the little green arrow and magic happens!



Question: Do you think your laboratory's analytical (RFU) threshold is:

1. Too high
2. Too low
3. Just right!
4. What is an RFU?
5. Need to think about it.



PRINCIPLES

Summary of Protocols/Methods That will be Discussed Today

Use data from negatives (i.e. samples with no DNA)

- Method 1.
  - Kaiser (IUPAC 1976)
    - Winefordner 1983 and Krane 2007
- Method 2.
  - Currie (IUPAC 1995)
    - Winefordner 1983
- Method 3.
  - Example in SWGDAM Guidelines

Use data from DNA dilution series

- Method 4.
  - Miller & Miller. *Statistics for Analytical Chemistry (Ellis Horwood & Prentice Hall)*
    - IUPAC 1997 ElectroAnalytical Committee
- Method 5.
  - 1997 IUPAC ElectroAnalytical Committee Recommendations

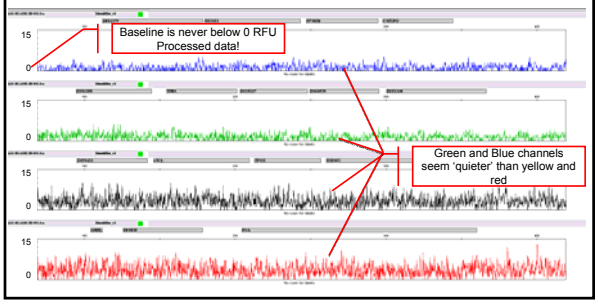
PRINCIPLES

Method 0. Qualitative Assessment - Negative

-Negative sample run with an internal size standard (not shown) using manufacturer's recommended protocol

Negative = extraction or amplification negative or run blank (more on this later!)

- Analyzed with GeneMapper v.3.1 using a threshold of 1 RFU



PRINCIPLES

Method 1. IUPAC (1976) – Negatives

- *The Limit of Detection, expressed as a concentration or quantity, is derived from the smallest measure that can be detected with reasonable certainty for a given analytical procedure*
- We must determine at what signal we can no longer RELIABLY separate signal from noise.
  - We do this by determining an Analytical Threshold (AT)

$$AT_{M1} = \overline{Y}_{bl} + ks_{bl}$$

$\overline{Y}_{bl}$ *is the average blank RFU signal*
s_{bl} *the std deviation of the blank signal*
AT_{M1} *the analytical signal calculated using Method 1*

- Kaiser argued a value of k = 3 will result in an AT whereby we are at least 89% confident (if the noise is not normally distributed) and at most 99.86% confident (when noise is normally distributed) noise will be below this value.

PRINCIPLES

Method 2. IUPAC (1995) – Negatives

- *The Critical Value (i.e. detected vs. not detected) is the minimum significant value of an estimated net signal or concentration, applied as a discriminator against background noise. This corresponds to a 1-sided significance test.*

$$AT_{M2} = \overline{Y}_{bl} + t_{1-\alpha,v}\frac{s_{bl}}{\sqrt{n}}$$

$\frac{s_{bl}}{\sqrt{n}}$ *is the estimated standard deviation of* the net signal when $x = 0$ *(i.e. blanks)*

$t_{1-\alpha,v}$ *from student t-table (i.e. 30 measurements, and 99% confidence,* $t_{1-\alpha,v} = 2.46$*)*

AT_{M2} *the analytical threshold calculated using Method 2*

PRINCIPLES

Method 3. Example in SWGDAM Guidelines – Negatives

- 1.1. ……*example, an analytical threshold may be based on two times the intensity difference between the highest peak and lowest trough within the instrumental noise data*

$$AT_{M3} = 2(Y_{\max} - Y_{\min})$$

Y_{max} *is the highest peak within instrumental noise data*

Y_{min} *is the signal of the lowest trough*

AT_{M3} *the analytical threshold calculated using Method 3*

PRINCIPLES

Method 4. Miller - DNA Dilution Series

- Determine the AT by considering the relationship between RFU and input DNA
- Regression of positive samples (i.e. samples with DNA in them)
- Run a sensitivity/dilution series
  - Does not assume blanks act as samples

$$AT_{M4} = b + 3S_y$$

b is the y-intercept

S_y *the standard error of regression*

AT_{M4} *the analytical threshold calculated using Method 4*

PRINCIPLES

Method 5. IUPAC CEAC (1997) – DNA Dilutions Series

- Determine the AT by considering the relationship between RFU and input DNA
- Regression of set(s) of dilution series
  - Uses confidence intervals to determine the analytical threshold

$$AT_{M5} = b + t_{n-1,\alpha} S_y$$

b is the y-intercept

S_y *the standard error of regression*

$t_{1-\alpha,v}$ *from student t-table (i.e. 5 dilutions, and 99% confidence,* $t_{1-\alpha,v} = 3.75$*)*

AT_{M5} *the analytical threshold calculated using Method 5*

21st International Symposium on Human Identification

Mixture Interpretation Workshop:

Principles, Protocols, and Practice

October 11, 2010 – San Antonio, TX

PROTOCOLS

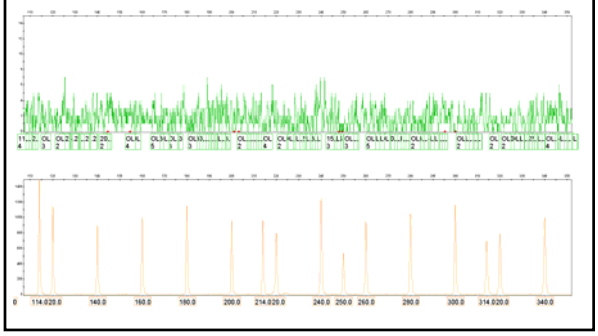
PROTOCOLS

Method 1, 2 and 3. – Negatives

- Run >20 negatives (more is always better, recommendation typically ranges from 20-30)
  - Negatives are those samples which have been through amplification and run with an Internal Size Standard
    - Although, our PRELIMINARY data show AT's derived from run blanks are similar to those determined from negatives
    - Suggest using amplification negatives as they are more representative
- Analyze with GeneMapper ID at an RFU of 1 for blue, green yellow, (red) and remove labels within +/-2 bases of the Internal Size Standard
- Export your files from GeneMapper into Excel (or other software) and begin determining your ATs!
- Analysis of raw data may also be performed

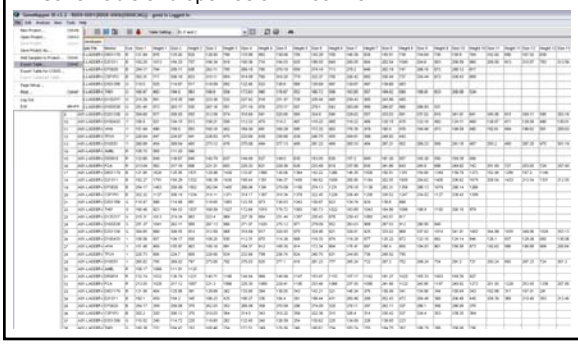
PROTOCOLS

- EXAMPLE
  - Blank run with LIZ®600 analyzed at 1 RFU and all labels/calls within 2 bases of the LIZ deleted



PROTOCOLS

- Ensure Genotypes tab is on screen and you can see the table
- File ► Export Table…
- Save Table and open as an Excel file

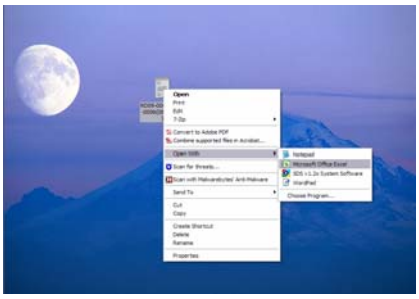


PROTOCOLS

- Calculate the
  - Average RFU of everything called/labeled
  - Standard Deviation of RFU
  - Minimum signal
  - Maximum signal
- The next 3 slides will show you one way to do this

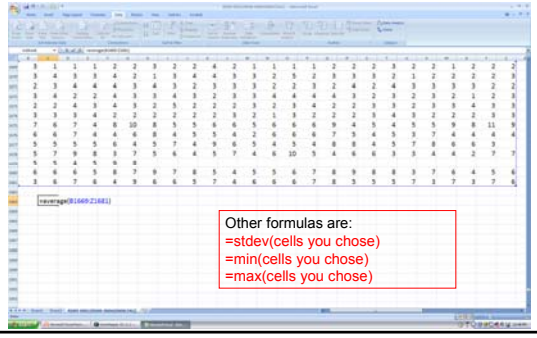
PROTOCOLS

- To open
  - Right click on the file
  - Choose Open With - Microsoft Office Excel



PROTOCOLS

- Delete the colors you do not want to analyze from the sheet
- Click on empty cell and type =average(
- Highlight everything you want to analyze
- Close the brackets so the formula looks like this now =average(cells you chose)
- and press Enter

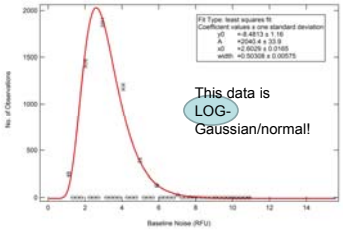


Other formulas are:
=stdev(cells you chose)
=min(cells you chose)
=max(cells you chose)

PROTOCOLS

Method 1. - Negatives

- These are values we obtained for 30 negatives in the green channel using 5s injections
- Results
  - Average Height (RFU)
    - 3.11
  - Standard Deviation
    - 1.14



$$AT_{M1} = \overline{Y}_{bl} + ks_{bl}$$

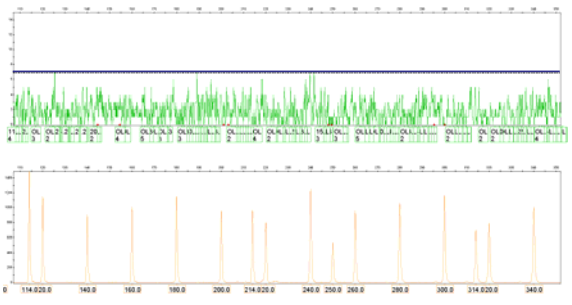
$$AT_{M1} = 3.11 + (3*1.14) = 6.53$$

$$AT_{M1} = 7$$

Wow! That is low! Can I use a larger k like 6 or 9 or 10? Yes, but be aware of what each means.

PROTOCOLS

If a threshold of 7 is used, bleed through from all colors (not just adjacent colors) needs to be considered!
Also, positive stutter needs to be considered!



PROTOCOLS

Method 2. – Negatives

- Average
  - 3.11
- Standard Deviation
  - 1.14
- t statistic (one-tailed) for 30 samples at 99% confidence
  - ?

$$AT_{M2} = \overline{Y}_{bl} + t_{1-\alpha,v}\frac{s_{bl}}{\sqrt{n}}$$

AT = analytical threshold

PROTOCOLS

Method 2. - Negatives

t

Confidence	z	Degrees of freedom (df) 100	50	40	30	25	20	15	12	10	9	8	7	6	5	4	3	2	1	p value 2-tailed	1-tailed
0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	.500
10%	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.16	.90	.450
20%	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.27	0.27	0.28	0.29	0.32	.80	.400
30%	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.40	0.40	0.40	0.40	0.40	0.41	0.41	0.42	0.44	0.51	.70	.350
40%	0.52	0.53	0.53	0.53	0.53	0.53	0.53	0.54	0.54	0.54	0.54	0.55	0.55	0.55	0.56	0.57	0.58	0.62	0.73	.60	.300
50%	0.67	0.68	0.68	0.68	0.68	0.68	0.69	0.69	0.70	0.70	0.70	0.71	0.71	0.72	0.73	0.74	0.76	0.82	1.00	.50	.250
60%	0.84	0.85	0.85	0.85	0.85	0.86	0.86	0.87	0.87	0.88	0.88	0.89	0.90	0.91	0.92	0.94	0.98	1.06	1.38	.40	.200
70%	1.04	1.04	1.05	1.05	1.05	1.06	1.06	1.07	1.08	1.09	1.10	1.11	1.12	1.13	1.16	1.19	1.25	1.39	1.96	.30	.150
80%	1.28	1.29	1.30	1.30	1.31	1.32	1.33	1.34	1.36	1.37	1.38	1.40	1.41	1.44	1.48	1.53	1.64	1.89	3.08	.20	.100
85%	1.44	1.45	1.46	1.47	1.48	1.49	1.50	1.52	1.54	1.56	1.57	1.59	1.62	1.65	1.70	1.78	1.92	2.28	4.17	.15	.075
90%	1.64	1.66	1.68	1.68	1.70	1.71	1.72	1.75	1.78	1.81	1.83	1.86	1.89	1.94	2.02	2.13	2.35	2.92	6.31	.10	.050
91%	1.70	1.71	1.73	1.74	1.75	1.76	1.78	1.81	1.84	1.86	1.90	1.93	1.97	2.02	2.10	2.23	2.47	3.10	7.03	.09	.045
92%	1.75	1.77	1.79	1.80	1.81	1.82	1.84	1.88	1.91	1.95	1.97	2.00	2.05	2.10	2.19	2.33	2.61	3.32	7.92	.08	.040
93%	1.81	1.83	1.85	1.86	1.88	1.89	1.91	1.95	1.99	2.03	2.06	2.09	2.14	2.20	2.30	2.46	2.76	3.58	9.06	.07	.035
94%	1.88	1.90	1.92	1.94	1.95	1.97	1.99	2.03	2.06	2.12	2.15	2.19	2.24	2.31	2.42	2.60	2.95	3.90	10.58	.06	.030
95%	1.96	1.98	2.01	2.02	2.04	2.06	2.09	2.13	2.18	2.23	2.26	2.31	2.36	2.45	2.57	2.78	3.18	4.30	12.71	.05	.025
96%	2.05	2.08	2.11	2.12	2.15	2.17	2.20	2.25	2.30	2.36	2.40	2.45	2.52	2.61	2.76	3.00	3.48	4.85	15.89	.04	.020
97%	2.17	2.20	2.23	2.25	2.28	2.30	2.34	2.40	2.46	2.53	2.57	2.63	2.71	2.83	3.00	3.30	3.90	5.64	21.21	.03	.015
98%	2.33	2.36	2.40	2.42	2.46	2.49	2.53	2.60	2.68	2.76	2.82	2.90	3.00	3.14	3.36	3.75	4.54	6.96	31.82	.02	.010
99%	2.58	2.63	2.68	2.70	2.75	2.79	2.85	2.95	3.05	3.17	3.25	3.36	3.50	3.71	4.03	4.60	5.84	9.92	63.66	.01	.005
99.9%	3.29	3.39	3.50	3.55	3.65	3.73	3.85	4.07	4.32	4.59	4.78	5.04	5.41	5.96	6.87	8.61	12.92	31.60	636.6	.001	.0005

Note: Round your df to the nearest tabled value. For t with df>100, use z.

Adapted from Dawson, R.J.M. (1997). "Turning the tables: A t table for today." *Journal of Statistics Education* v.5, n.2.
Generated using Microsoft Excel.

PROTOCOLS

Method 2. - Negatives

- Average
  - 3.11
- Standard Deviation
  - 1.14
- t statistic (one-tailed) for 30 samples at 99% confidence – assumes Gaussian distribution!
  - 2.46

$$AT_{M2} = \overline{Y}_{bl} + t_{1-\alpha,v}\frac{s_{bl}}{\sqrt{n}}$$

$$AT_{M2} = 3.11 + \left(2.46 * \frac{1.14}{\sqrt{30}}\right) = 3.68$$

AT = analytical threshold

$$AT_{M2} = 4$$

Can I use a larger confidence interval?

PROTOCOLS

Method 3. – Negatives

- Max 10
- Min 0

$$AT_{M3} = 2(Y_{\max} - Y_{\min})$$

$$AT_{M3} = 2(10 - 0) = 20$$

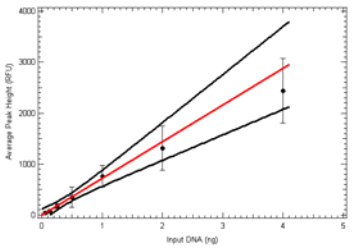
$$AT_{M3} = 20$$

NOTE: Because we are NOT using raw data (but analyzed GeneMapper data), data below 0 RFU is not 'observed' and therefore, the number calculated is smaller than expected!!! HOWEVER, the calculated AT is still larger than either Method 1 or 2!

PROTOCOLS

Method 4 and 5. – DNA Dilution Series

- Regression of positive samples (i.e. single source samples)
- Amplified 0.0625-4ng dilution series, injected 5s using manufacturer's recommended protocol
  - Will help determine the point at which you begin to see offscale data – SWGDAM 3.1.1.3
- Plot of Input DNA (ng) versus average peak height (per color)
  - If a peak was homozygous, the RFU was divided by 2



- The points at 2 and 4 ng fall off the line (PCR efficiency approaching a plateau)!
- The error bars become larger with increased DNA input!

PROTOCOLS

Method 4. – DNA Dilution Series

- Because of the plateau of PCR efficiency and increasing error bars
  - We need to ensure the regression is within a linear range (i.e. 0.0625 – 1 ng)
  - A weighted regression will be necessary to determine the y-intercept and other linear parameters

• HOW?

- Excel based templates or macros tools are freely available for your use/practice on the web
  - Template
    - www.bumc.bu.edu/biomedforensic/faculty-and-staff/faculty/catherine-grgicak/tools
  - Macros 'Excellaneous" website at
    - www.bowdoin.edu/~rdelevie/excellaneous by Robert de Levie (www.bowdoin.edu/~rdelevie/)

PROTOCOLS

Method 4 and 5. – DNA Dilution Series

- Amplify a dilution/sensitivity series of single source DNA from ~0.0625 to 1 ng (you will have to confirm this range is linear for you)
  - RFUs of homozygotes are divided by 2
- Run it using your platform and run protocol
  - NOTE: Different injection times will result in different ATs!
- Export the genotypes table with allele heights and open it in Excel (see previous slides)
- Calculate the average and standard deviations (per color) of the RFUs obtained for samples run at that target

PROTOCOLS

Method 4 and 5. – DNA Dilution Series

- Open the template on the www.bumc.bu.edu (full website address above) and plug the values into the grey areas

Weighted regression — 5

x_i	y_i	s_i	$1/s_i2$	w_i	w_ix_i	w_iy_i	$w_ix_iy_i$	$w_ix_i^2$	$w_iy_i^2$
0.0625	48.99375	22.1127	0.002045	2.708906	0.169307	132.7195	8.294967	0.010582	6502.425
0.125	77.80625	27.84388	0.00129	1.708512	0.213564	132.9329	16.61661	0.026695	10343.01
0.25	177.00625	50.50294	0.000392	0.519331	0.129833	91.9248	22.9812	0.032458	16271.26
0.5	352.21875	194.4713	2.64E-05	0.035024	0.017512	12.33613	6.168067	0.008756	4345.018
1	768.2625	216.6237	2.13E-05	0.028227	0.028227	21.68579	21.68579	0.028227	16660.38
		sums	0.003775	5	0.558442	391.5991	75.74664	0.106718	54122.1
		means			0.111688	78.31982			

Output	
slope(m)	721.7997156
y intercept (b)	-2.296895975
standard error(S_y)	10.76503242

Input your values into the grey cells
x_i is the amount of DNA (ng) input into PCR
y_i is the average RFU
s_i is the standard deviation of RFU
Don't forget to put in the number of values in grey cell at the top (above s_i)!

PROTOCOLS

Method 4.– DNA Dilution Series

- b (y-intercept) = -2.30
- S_y (standard error of regression) = 10.77

$$AT_{M4} = b + 3S_y$$

$$AT_{M4} = -2.30 + (3*10.77) = 30.01$$

$$\boxed{AT_{M4} = 31}$$

AT = analytical threshold

PROTOCOLS

Method 5. – DNA Dilution Series

- b (y-intercept) = -2.30
- Sy (standard error of regression) = 10.77
- t-stat (n-1=4) and alpha of 99% t=3.75 (see Method 2.)

$$\mathrm{AT_{M5}} = \mathrm{b} + t_{n-1,\alpha} S_y$$

$$AT_{M5} = -2.30 + (3.75*10.77) = 38.09$$

$$\boxed{AT_{M5} = 39}$$

AT = analytical threshold

PROTOCOLS

Which Method Does One Choose?

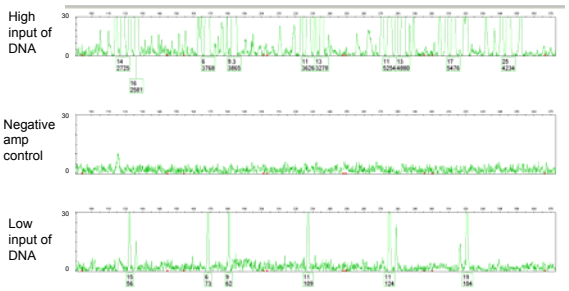
- Summary of Results

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

- Before you choose, consider the following slide…

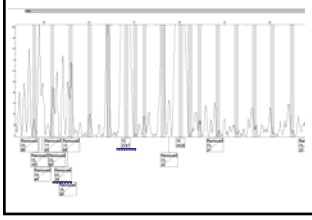
PROTOCOLS

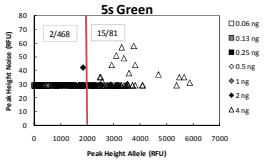
Do samples with a significant amount of template result in baselines different from those observed with blanks?

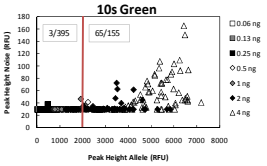


PROTOCOLS

- Single source DNA data amplified from 0.0625 – 4ng
  - Differentiated 'noise' from artifact
    - -A, pull-up, stutter (+ or -), spikes, dye artifacts
- Plotted RFU of the known/expected peak versus the highest 'noise' peak







21st International Symposium on Human Identification
Mixture Interpretation Workshop:
Principles, Protocols, and Practice
October 11, 2010 – San Antonio, TX

PRACTICE

PRACTICE

Which Method Does One Choose?

- Remember: The most prudent method is not always the easiest one
- Become familiar with the literature
  - Five of a multitude of methods were reviewed
- Each of the methods can be used to describe the analytical threshold in SWGDAM Guidelines 1.1 and 3.1.1.2
  - Data suggest 'noise' does not remain constant between negatives and samples with a significant amount of DNA (i.e. RFUs >2000).
  - There may be amplification 'noise' that cannot be characterized as known artifact (i.e. bleed-through, spike, stutter, etc)

PRACTICE

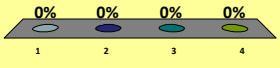
Which Method Does One Choose?

- Methods 4 and 5 require that each run protocol (i.e. different injections times) be analyzed separately
- Different color channels behave differently – if possible, determine ATs for each color
- ATs derived from methods based on negative sample analysis (i.e. Method 1) may not be optimal for medium-high template samples – but reasonable for low-template ones

AT = analytical threshold

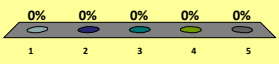
If you were developing your protocol from scratch, what data would you use?

1. Negative controls
2. DNA dilution series (with peaks removed)
3. None (use manufacturer's recommendation)
4. A dart board!



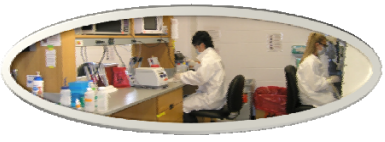
After hearing this presentation, do you think your laboratory's analytical (RFU) threshold is:

1. Too high
2. Too low
3. Just right!
4. What is an RFU?
5. Need to think about it.



Thank-you!

- Thanks to the BU BMFS students who did all of the work!
  - Joli Bregu
  - Danielle Conklin
  - Elisse Coronado



- Thank-you to you for your attention!
- Any questions/comments
  - cgrgicak@bu.edu