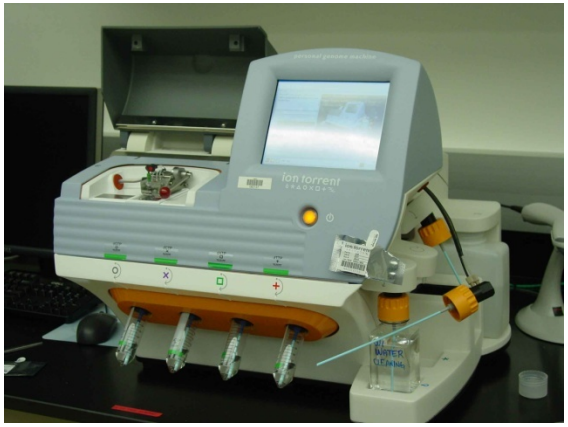
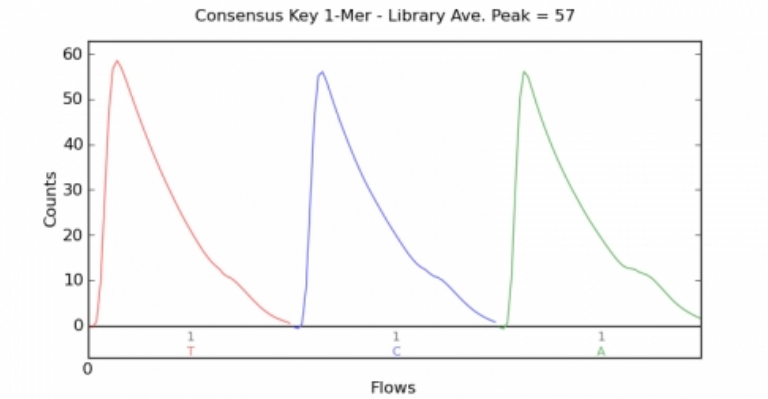
BINF 6350 ITSC 8350
Fall 2011
Biotechnology & Genomics Lab
Creating DNA libraries

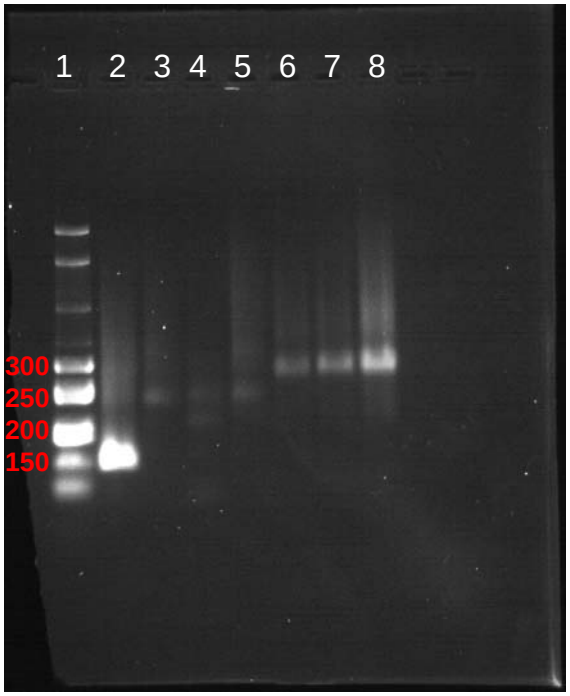
http://webpages.uncc.edu/~jweller2





Library Quantification

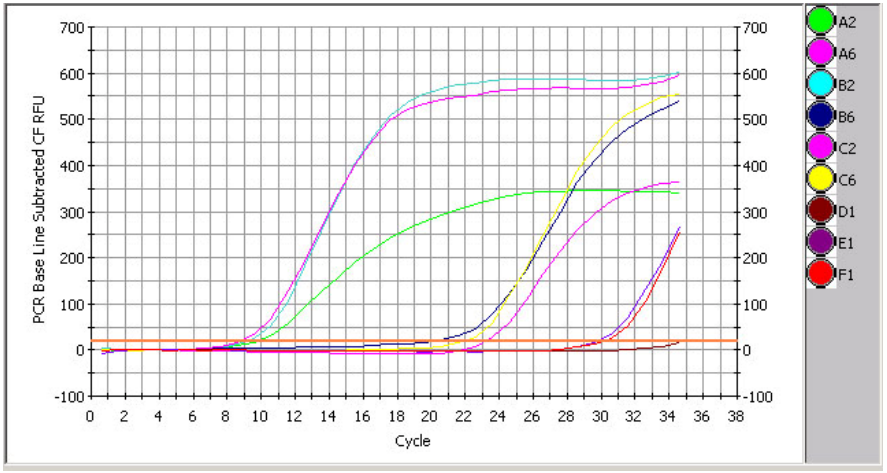
- Examine some of the fragments from our qPCR experiment

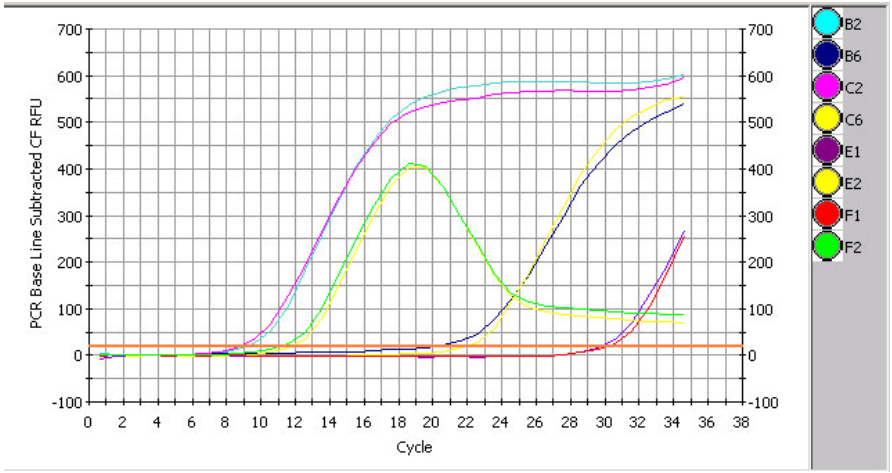


Lane 1: Amresco DNA ladder
Lane 2: Kapa 153-bp standard
Lane 3: Library 1 – mean ~250
Lane 4: Library 2 – mean ~250
Lane 5: Library 3 – mean ~250
Lane 6: Library 4 – mean ~300
Lane 7: Library 5 – mean ~300
Lane 8: Library 6 – mean ~300

4ul of each sample, 2% gel, 50V for 3 hours in 1X TBE

Real-time results



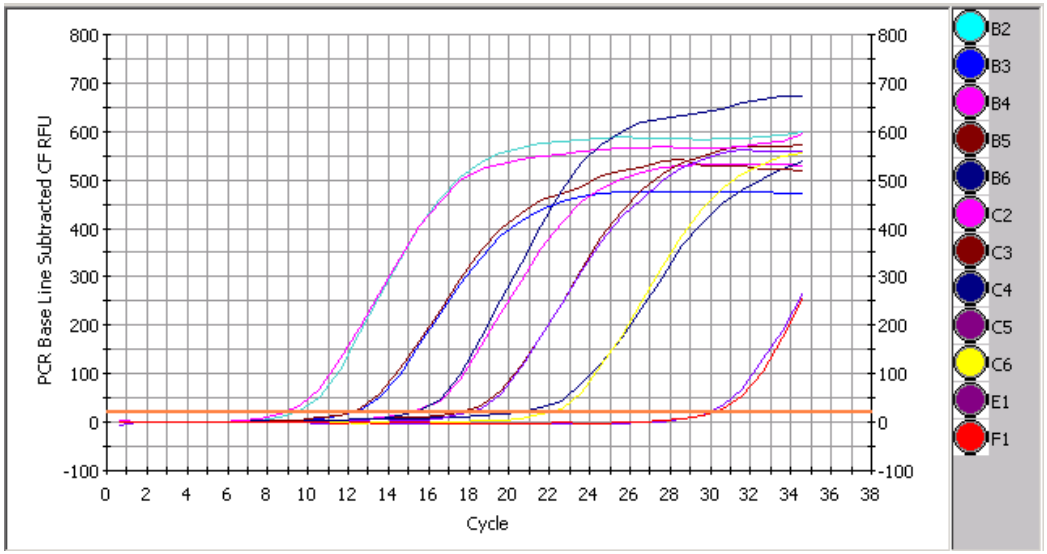


Grouping are
Standard #2: A2, B2, C2
Standard #6: A6, B6, C6
Neg ctrl: D1, E1, F1

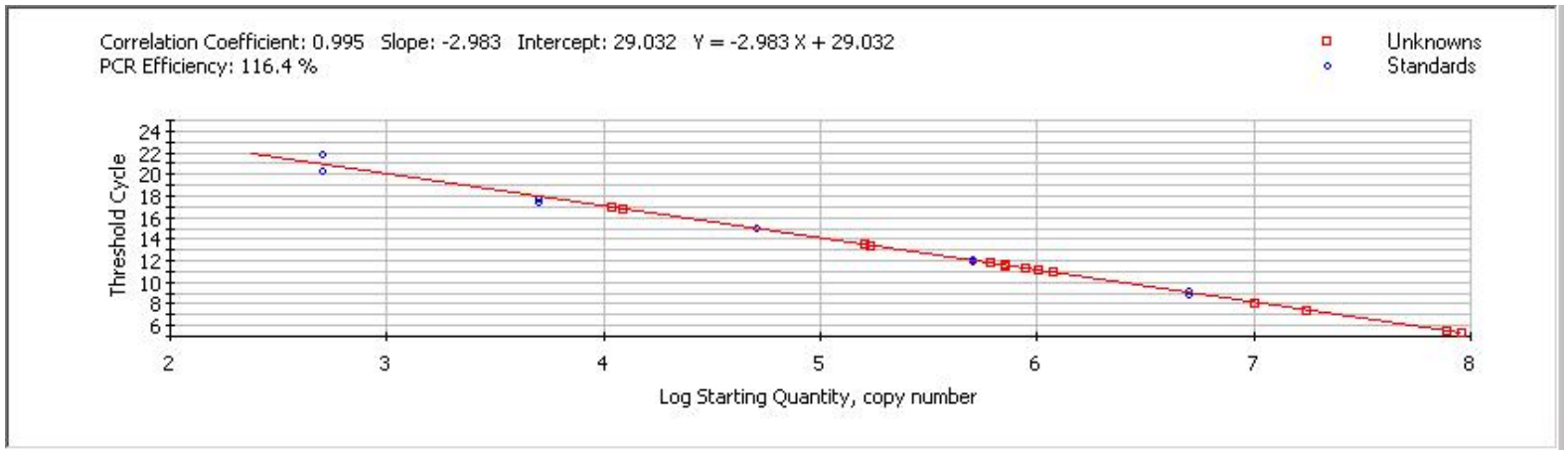
Why is Row 1 so far off?
Inspect plate

Leaving out the row A, D:
Standard #2: B2, C2
Standard #6: B6, C6
NegCtrl: E1, E2
Library 3 dil: F1, F2

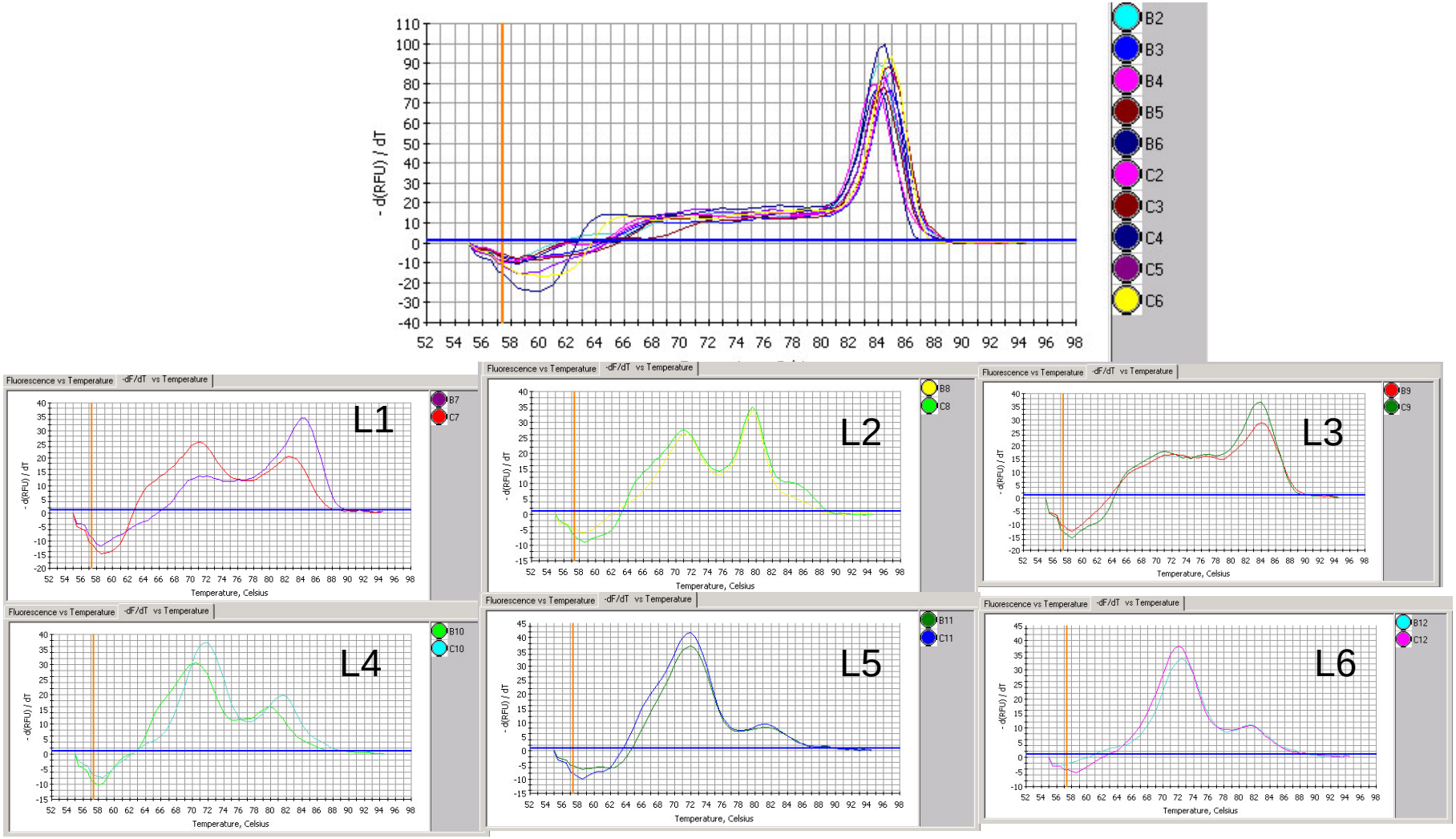
qPCR check



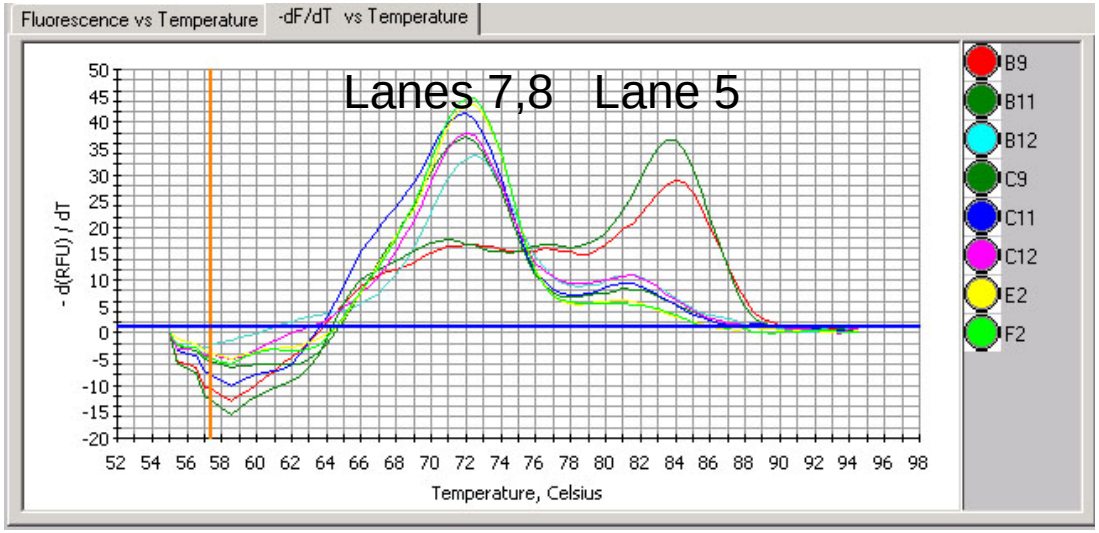
If all three points are used, the slope is -3.06 and the intercept is 29.75.

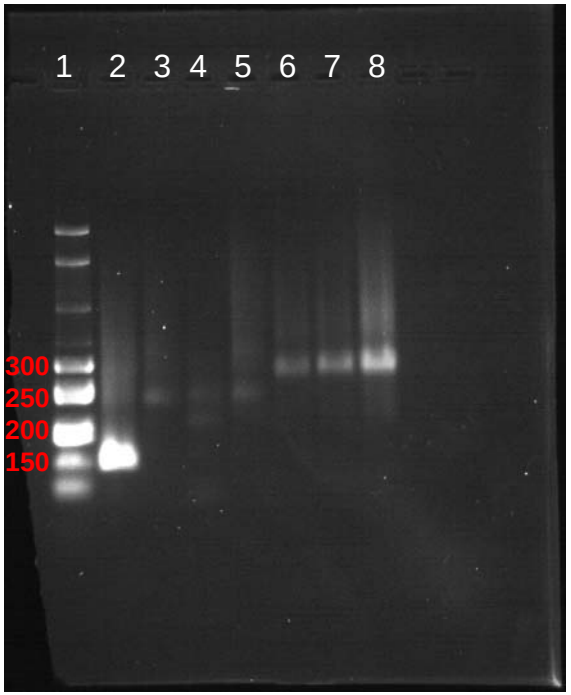


Melt curves


- d(RFU) / dT
Temperature, Celsius
B2
B3
B4
B5
B6
C2
C3
C4
C5
C6
L1
B7
C7
L2
B8
C8
L3
B9
C9
L4
B10
C10
L5
B11
C11
L6
B12
C12

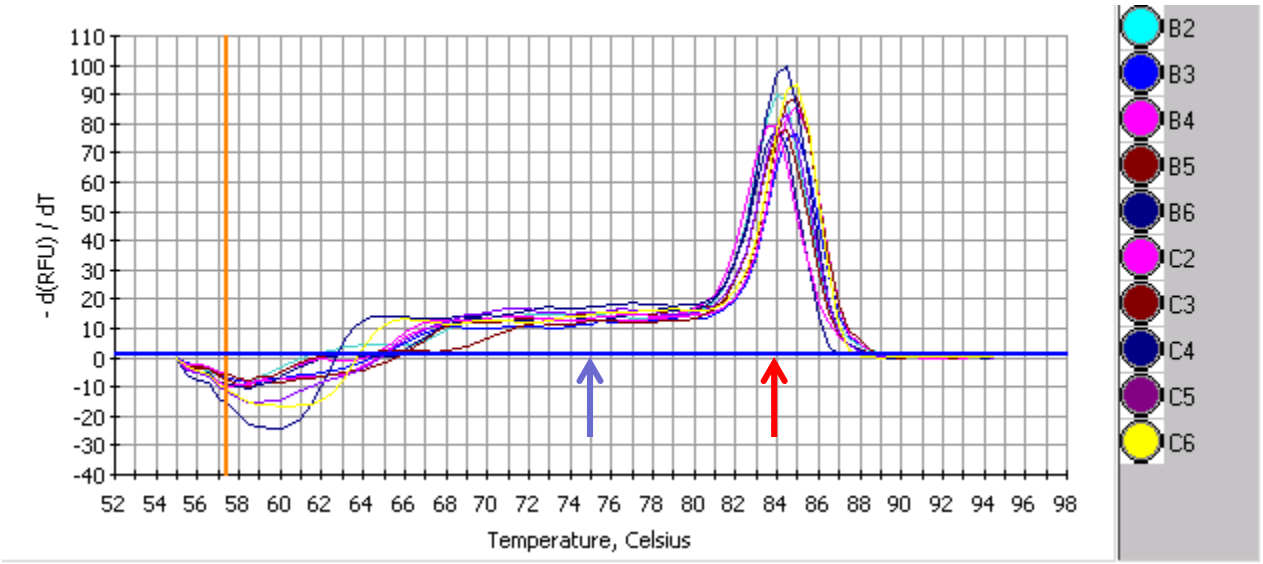
Multiple population?





Can I correlate Ct to gel band intensity?
(I combined wells so not reliable in this case)

Analyzing the qPCR data



Melt Curve Analysis Spreadsheet Data for SYBR-490

Well	Well Identifier Peak Descriptor	Peak ID	Melt Temp	Beg. Temp	End Temp
B2		B2.1	84.0 ←	78.5	94.0
		B2.2	75.0 ←	74.5	78.0
		B2.3	73.5	66.0	74.0
B3		B3.1	85.0	77.5	90.5
		B3.2	75.5	73.5	76.5
		B3.3	72.0	70.5	73.0
		B3.4	68.5	64.0	70.0

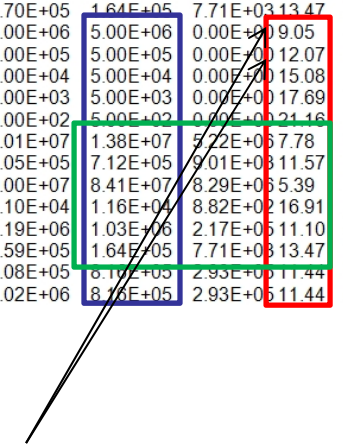
PCR Quantification Spreadsheet Data for SYBR-490

Well	Identifier	Ct	Setpoint
B02		9.24	
B03		12.15	
B04		15.08	
B05		17.54	
B06		20.38	
B07		7.43	
B08		11.56	
B09		5.48	
B10		16.84	
B11		11.30	
B12		13.43	
C02		8.86	
C03		11.99	
C04		15.09	
C05		17.85	
C06		21.95	
C07		8.14	
C08		11.58	
C09		5.30	
C10		16.98	
C11		10.91	
C12		13.52	
E01		29.87	
E02		11.78	
F01		30.14	
F02		11.10	

Standard Curve Spreadsheet Data for SYBR-490

Units: copy number

	Type	Identifier	Rep	Ct	Log SQ	SQ	SQ Mean	SQ SD	Ct Mean	Ct SD
B02	Standard		2	9.24	6.699	5.00E+06	5.00E+06	0.00E+00	9.05	2.69E-01
B03	Standard		3	12.15	5.699	5.00E+05	5.00E+05	0.00E+00	12.07	1.15E-01
B04	Standard		4	15.08	4.699	5.00E+04	5.00E+04	0.00E+00	15.08	7.76E-03
B05	Standard		5	17.54	3.699	5.00E+03	5.00E+03	0.00E+00	17.69	2.22E-01
B06	Standard		6	20.38	2.699	5.00E+02	5.00E+02	0.00E+00	21.16	1.11E+00
B07	Unknown		1	7.43	7.242	1.74E+07	1.38E+07	5.22E+06	7.78	5.04E-01
B08	Unknown		2	11.56	5.856	7.18E+05	7.12E+05	9.01E+03	11.57	1.64E-02
B09	Unknown		3	5.48	7.893	7.82E+07	8.41E+07	8.29E+06	5.39	1.28E-01
B10	Unknown		4	16.84	4.087	1.22E+04	1.16E+04	8.82E+02	16.91	9.87E-02
B11	Unknown		5	11.30	5.944	8.80E+05	1.03E+06	2.17E+05	11.10	2.74E-01
B12	Unknown		6	13.43	5.229	1.70E+05	1.64E+05	7.71E+03	13.47	6.09E-02
C02	Standard		2	8.86	6.699	5.00E+06	5.00E+06	0.00E+00	9.05	2.69E-01
C03	Standard		3	11.99	5.699	5.00E+05	5.00E+05	0.00E+00	12.07	1.15E-01
C04	Standard		4	15.09	4.699	5.00E+04	5.00E+04	0.00E+00	15.08	7.76E-03
C05	Standard		5	17.85	3.699	5.00E+03	5.00E+03	0.00E+00	17.69	2.22E-01
C06	Standard		6	21.95	2.699	5.00E+02	5.00E+02	0.00E+00	21.16	1.11E+00
C07	Unknown		1	8.14	7.003	1.01E+07	1.38E+07	5.22E+06	7.78	5.04E-01
C08	Unknown		2	11.58	5.848	7.05E+05	7.12E+05	9.01E+03	11.57	1.64E-02
C09	Unknown		3	5.30	7.954	9.00E+07	8.41E+07	8.29E+06	5.39	1.28E-01
C10	Unknown		4	16.98	4.040	1.10E+04	1.16E+04	8.82E+02	16.91	9.87E-02
C11	Unknown		5	10.91	6.074	1.19E+06	1.03E+06	2.17E+05	11.10	2.74E-01
C12	Unknown		6	13.52	5.200	1.59E+05	1.64E+05	7.71E+03	13.47	6.09E-02
E02	Unknown		7	11.78	5.784	6.08E+05	8.16E+05	2.93E+05	11.44	4.77E-01
F02	Unknown		7	11.10	6.010	1.02E+06	8.16E+05	2.93E+05	11.44	4.77E-01



About 3 cycles apart

Modifying the Concentration

- Confirm the reaction efficiency to be between 90-110% (our is 112-116%, which is not bad considering how many people were pipetting and the loss of Row A).
- Our library dilution factors were 1:500. The average fragment lengths also need to be incorporated.

Library	Ct	Unmod Molecules	Avg fragment (from Gel)	Efficiency Factor*	Dil factor	Converted Conc (Unmod*EF * DF)
(1)	7.78	1.38E7	250	0.61	500	4.21E+09
(2)	11.57	7.12E5	250	0.61	500	2.17E+08
(3)	5.39	8.41E7	250	0.61	None	5.13E+07
(4)	16.91	1.16E4	300	0.51	500	2.96E+06
(5)	11.10	1.03E6	300	0.51	500	2.63E+08
(6)	13.47	1.64E5	300	0.51	500	4.18E+07

* From length comparison with 153bp standard fragment.

Target dilutions

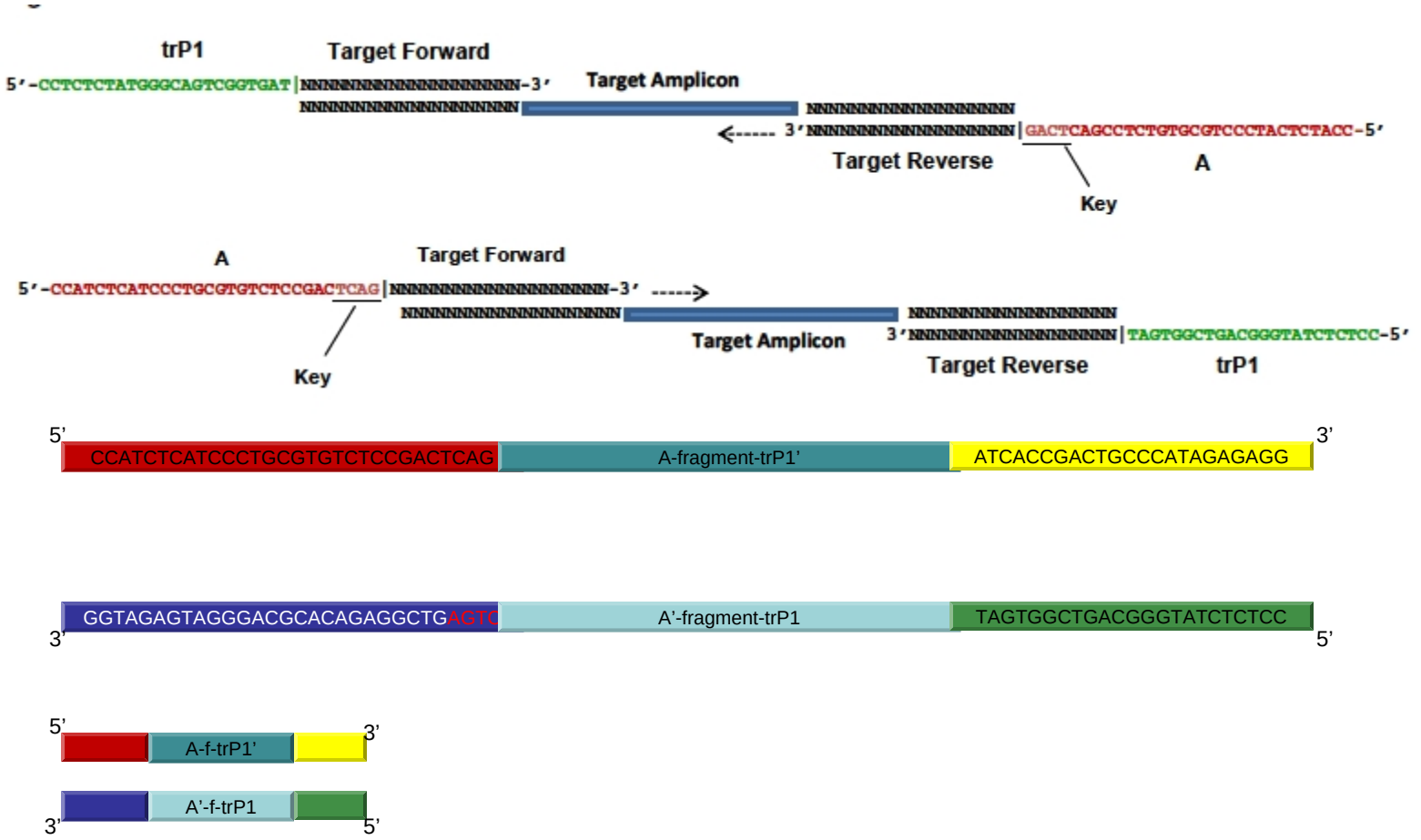
- The Ion Torrent protocol for emPCR requires that you make a solution that has 1.56 *10^7 molecules per ul (this changes with the specific chip and target length so don't take this as an eternal verity.
- The calculations for making 500ul follow

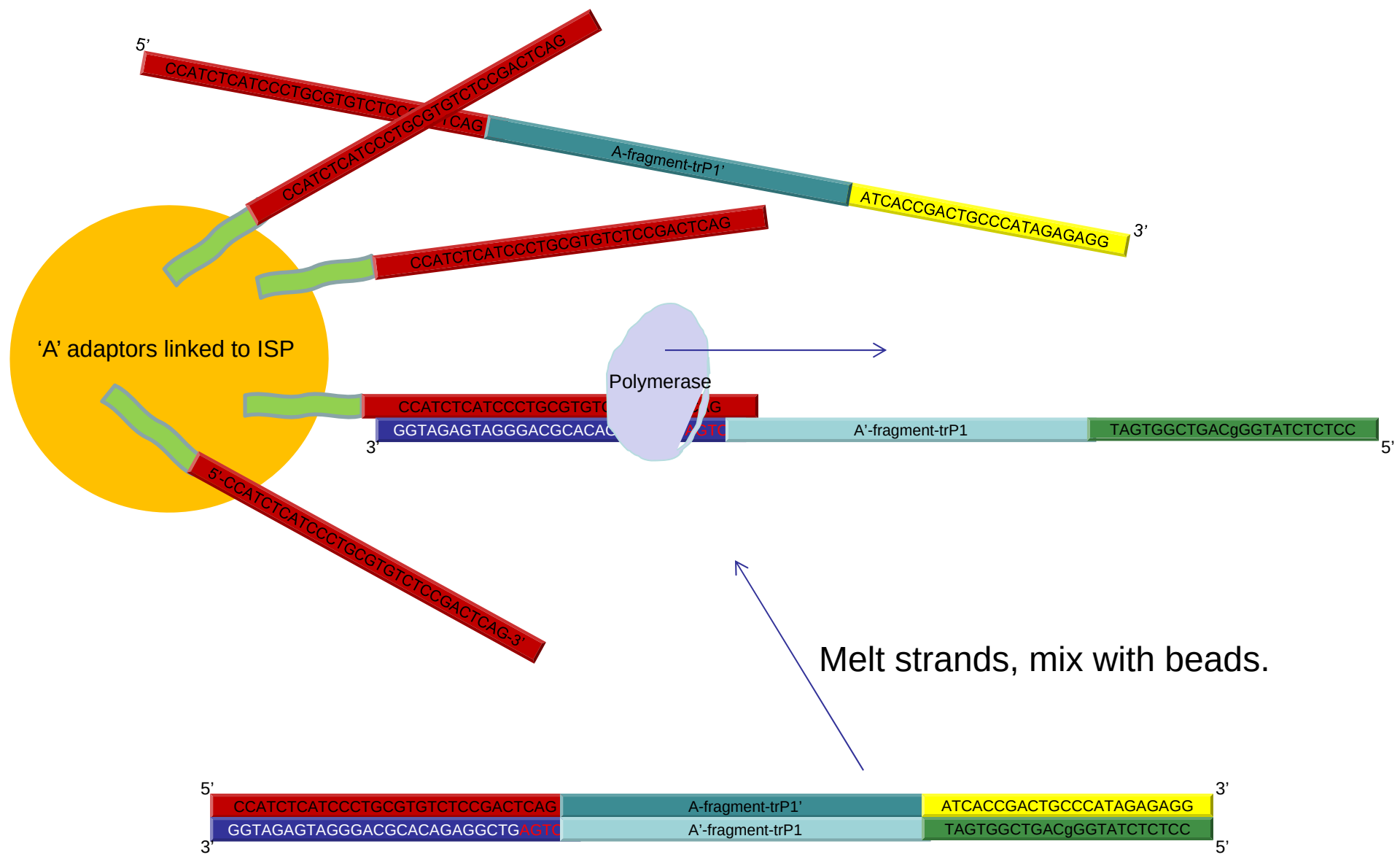
Library	Stock Conc (in Molecules per ul)	ul stock for $7.78*10^9$	Volume buffer	
1	4.21E+09	1.8484	498.1516	
2	2.17E+08	35.8261	464.1739	
3	5.13E+07	151.6540	348.3460	
4	2.96E+06	2630.1555	-2130.1555	Note: don't use
5	2.63E+08	29.6212	470.3788	
6	4.18E+07	186.0354	313.9646	

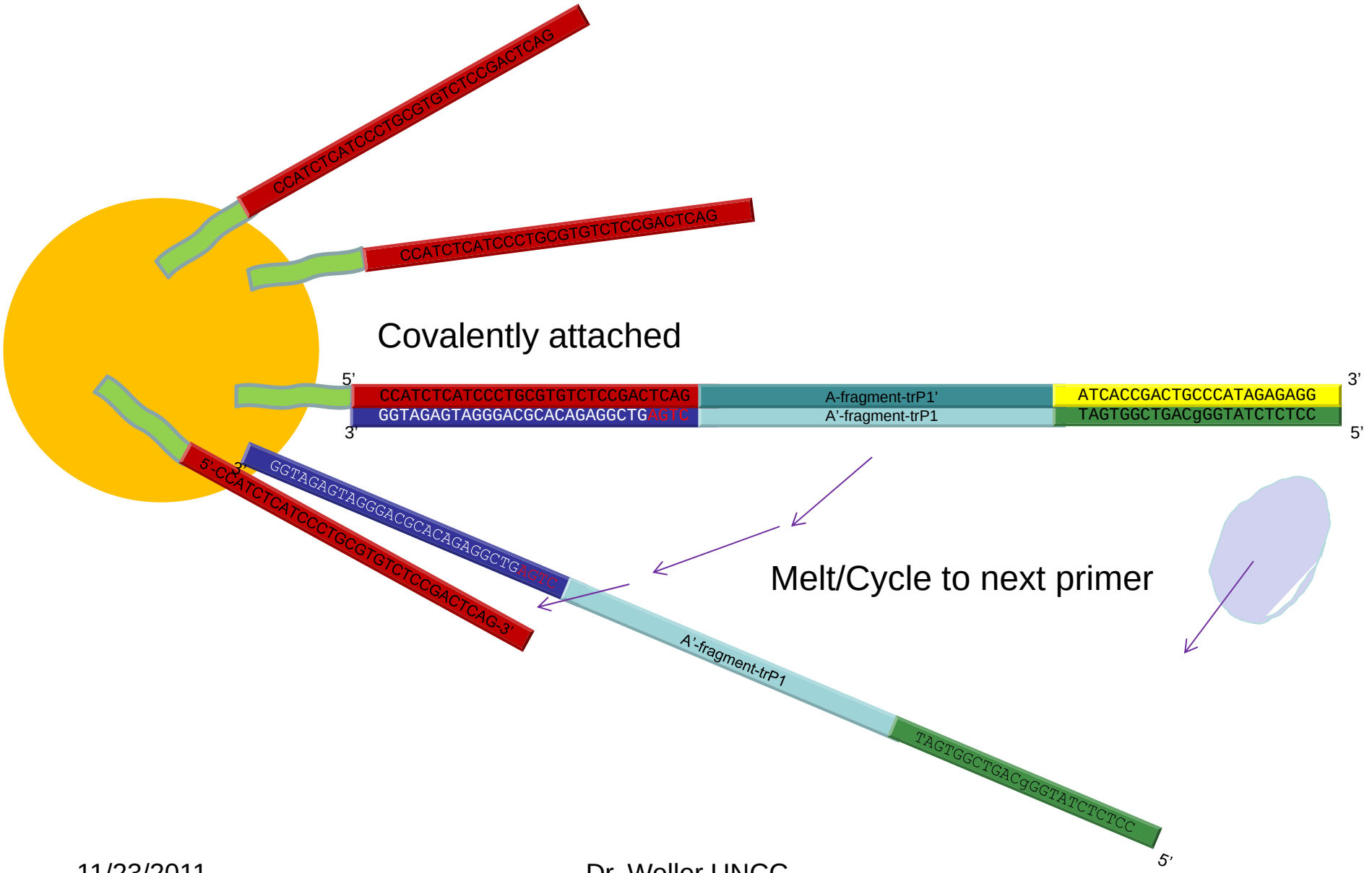
Setting up emulsion PCR

- Set up the emulsions by
  - IKA Turrax and then dispensing to the plate
    - The solution is very viscous – it is hard to deliver the same volume repeatedly
  - OneTouch – streams solutions together into the plate
    - Materials are resistant?
    - The plates may leak (a recent warning)
      - Decontamination/cross-contamination issues?

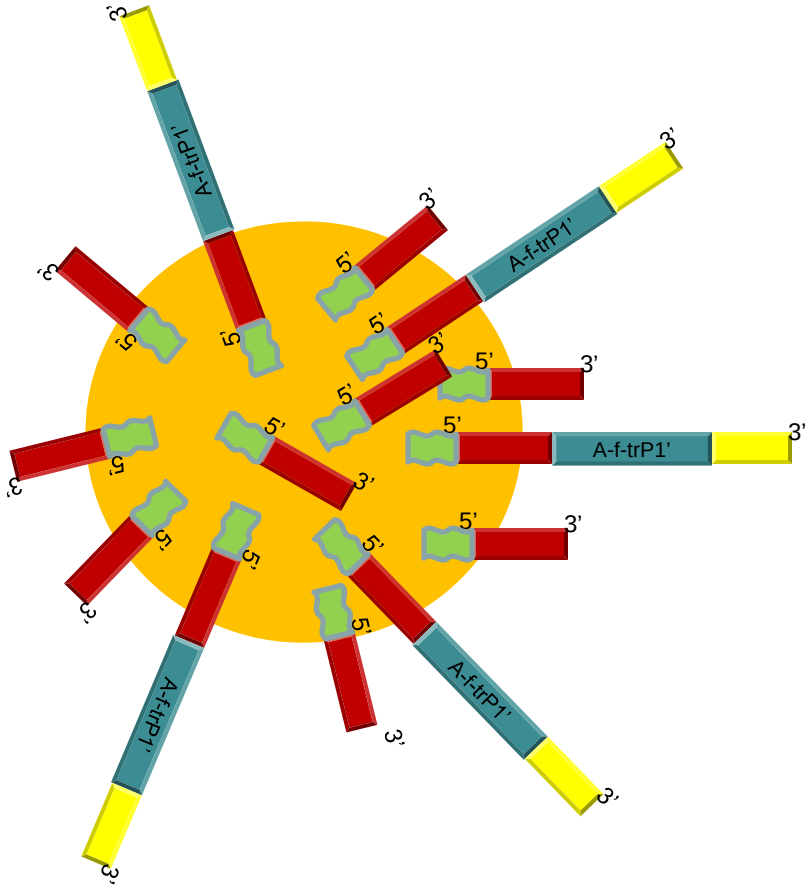
Bead work : Library representations

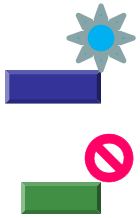



5’
CCATCTCATCCCTGCGTGTCTCCGACTCAG
A-fragment-trP1’
ATCACCGACTGCCCATAGAGAGG
3’
CCATCTCATCCCTGCGTGTCTCCGACTCAG
‘A’ adaptors linked to ISP
Polymerase
CCATCTCATCCCTGCGTGTC
GGTAGAGTAGGGACGCACAC
A’-fragment-trP1
TAGTGGCTGACgGGTATCTCTCC
3’
5’
5’-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3’
Melt strands, mix with beads.
5’
CCATCTCATCCCTGCGTGTCTCCGACTCAG
A-fragment-trP1’
ATCACCGACTGCCCATAGAGAGG
3’
GGTAGAGTAGGGACGCACAGAGGCTGAGTC
A’-fragment-trP1
TAGTGGCTGACgGGTATCTCTCC
3’
5’


CCATCTCATCCCTGCGTGTCTCCGACTCAG
CCATCTCATCCCTGCGTGTCTCCGACTCAG
Covalently attached
5’
CCATCTCATCCCTGCGTGTCTCCGACTCAG
A-fragment-trP1’
ATCACCGACTGCCCATAGAGAGG
3’
GGTAGAGTAGGGACGCACAGAGGCTGAGTC
A’-fragment-trP1
TAGTGGCTGACgGGTATCTCTCC
3’
5’
3’
5’-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3’
GGTAGAGTAGGGACGCACAGAGGCTGAGTC
Melt/Cycle to next primer
A’-fragment-trP1
TAGTGGCTGACgGGTATCTCTCC
5’

Qubit Logic

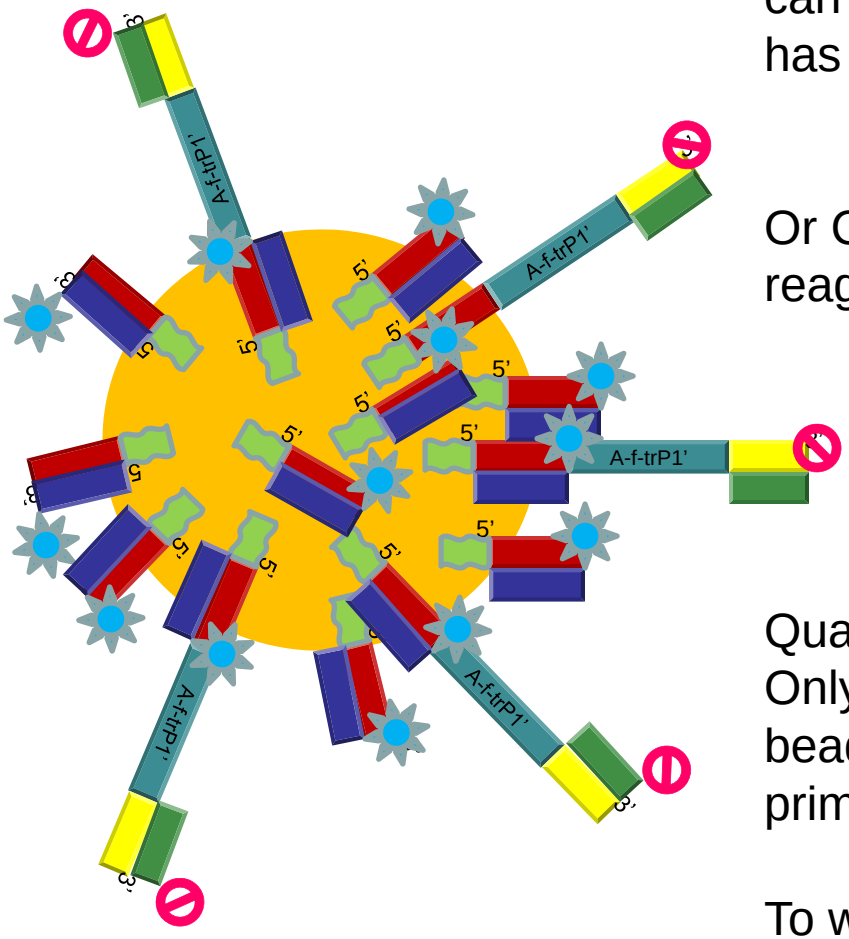




Oligo 1 is [A' + Dye 1]

Oligo 2 is [trP1 + Dye 2]

Melt, Hybridize and wash



The Ratio of dyes gives an estimate of total A primers (synthesized on the ISP) to the number of trP1 primers, which can only have a complement if a target has been added. Ratio of ✸ to ⊘.

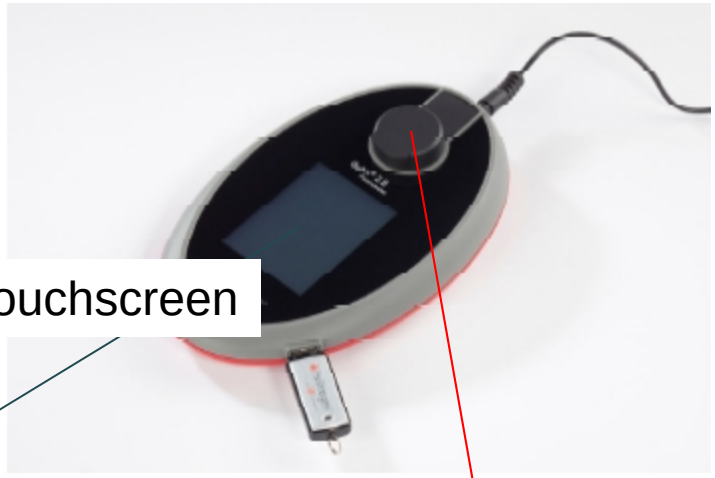
Or Cy5 to FAM, in the case of our reagents.

Quantitative?
Only if you could accurately pipette beads or could calibrate amount of A primer per bead very accurately.

To which primer do you think the biotin is attached, A' or trP1?

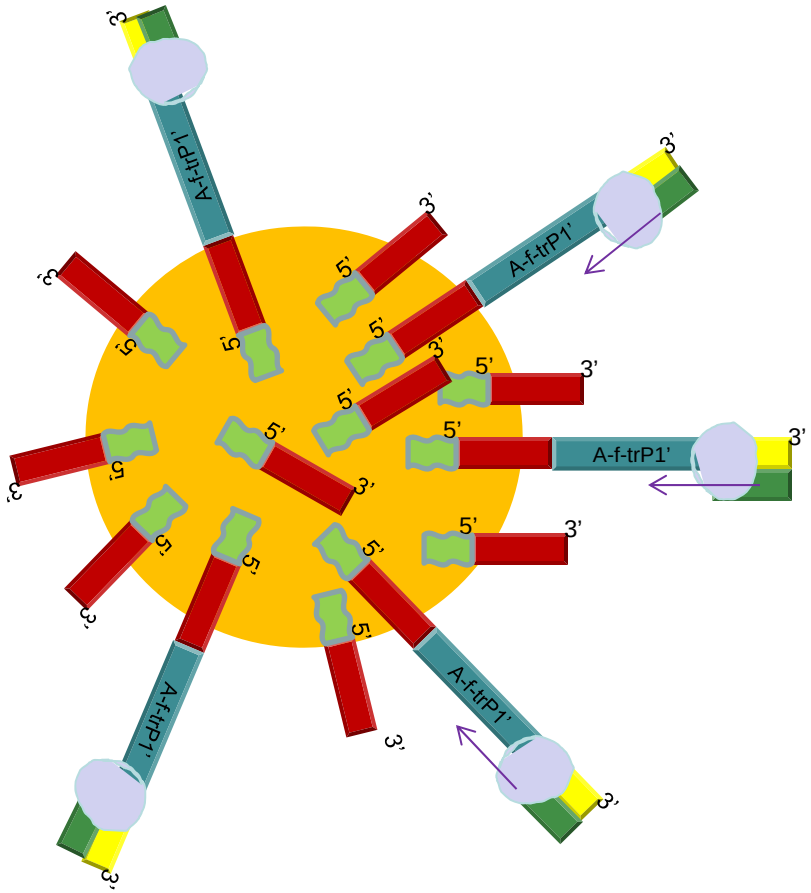
Qubit Fluorometer

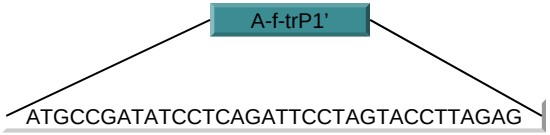
- Note that this is a fluorometer – calibration standards are required for interpretation of signal (RFUs).
- Data is stored on the USB in .csv format so you can transfer the output to a modeling environment.



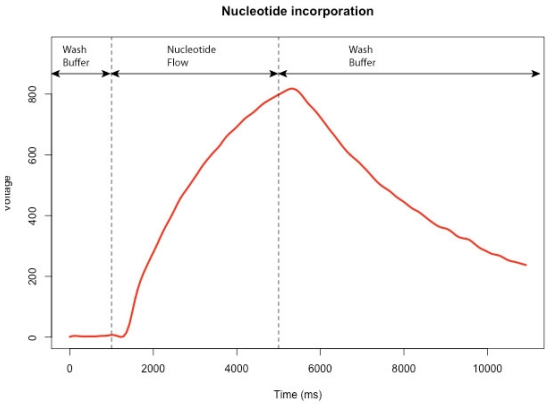
Lid flips up and tube sits in here

Signal and Ionograms



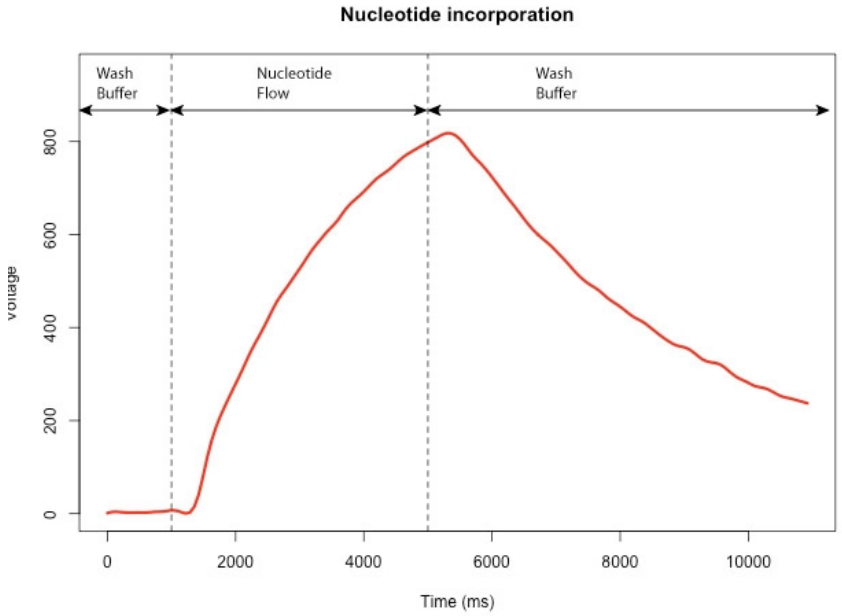


Flow 1: add G
Flow 2: add C
Flow 3: add A
Flow 4: add T



Engineering notes

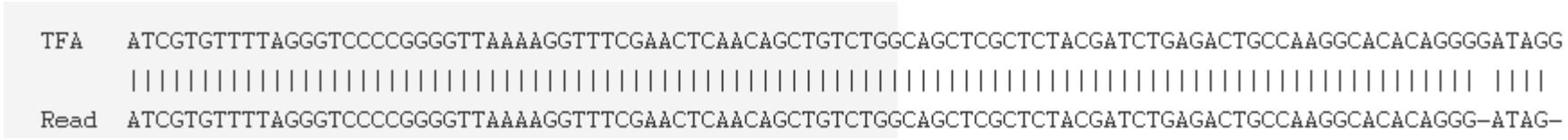
- Measurement frequency is asymmetric
  - More frequent in the rise portion (every 4ms) and less frequent in the decay
  - You could adjust flow times and sampling intervals to optimize accuracy.
- Background signal – random fluctuations in sensors from ambient solution
  - Wells are different distances from the inlets/outlets
  - Flows vary in type
  - Empty wells are used to estimate background (can be iterative)
- Foreground subtract the background – you have to pick the time interval
  - For homopolymers there is a (very short) delay



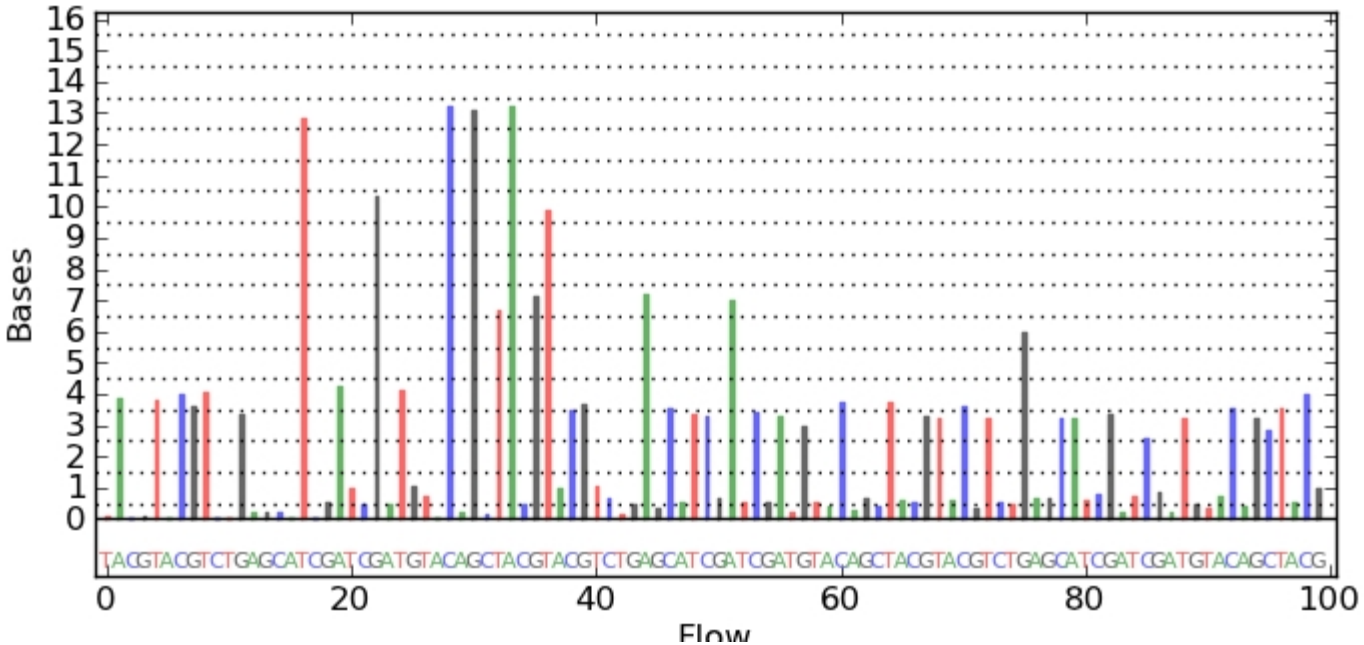
Nice discussion of signal processing and Grand Challenges at http://biolektures.wordpress.com/2011/08/28/

Read normalization

- We go from ion → voltage → ionogram → base call/number
- The test fragment (spiked in) is used to normalize the signal
  - The key sequence is used to initiate the process.

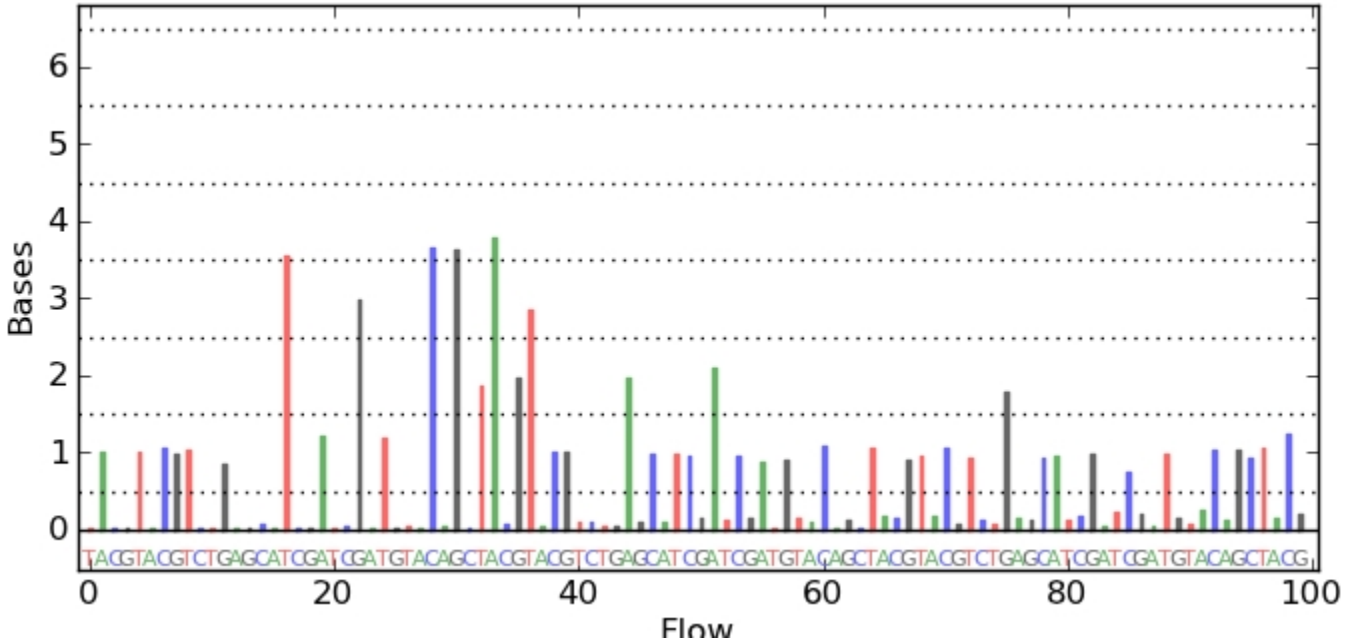


Registration: the key is TACG. Every sequence has this. The first 7 flows are in the order: T,A,C,G,T,A,C . That is values for flows 1,4 and 6 are for 1-nt incorporation and values for flows 0,2,3,5 should mimic background. (why is G not mentioned here?)
After subtracting the background, divide each flow value by its 1-mer signal (assuming a linear relationship)



Flow 0, 2, 3, 5 values (i.e. background signal)
= 0.11, 0.00, 0.07, 0.00

Average background signal
= 0.0450

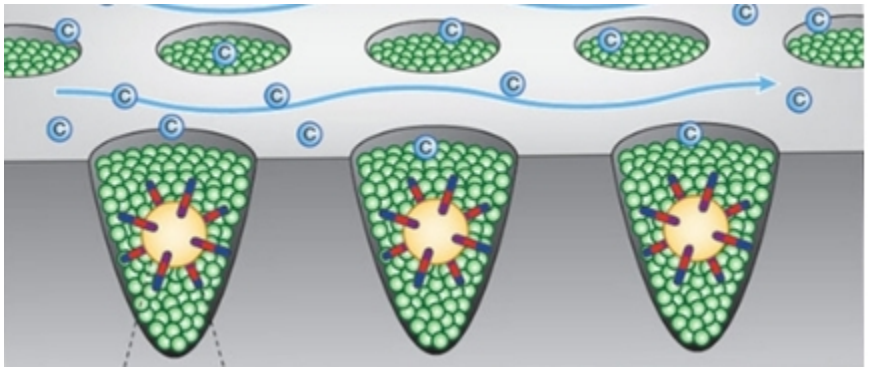


Flow 1, 4, 6 values (i.e. 1mer signal)
= 3.88, 3.83, 3.98

Average 1mer signal (after background subtraction) = 3.8588

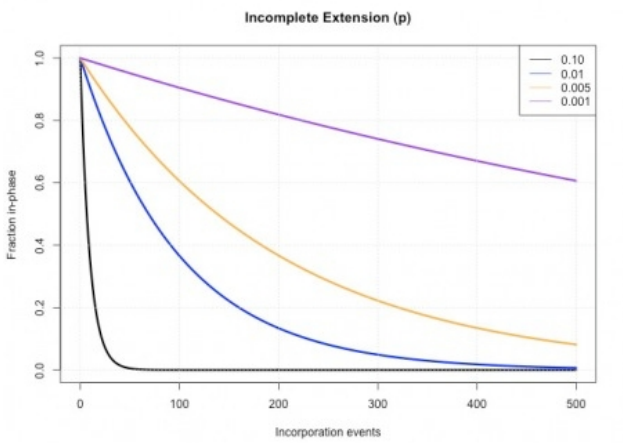
Sources of sequencing errors

- Chemical reactions (even catalyzed reactions) do not go to 100% completion.
- Enzymes have a rate of progression.
- Nucleotides flow across the surface (uneven, different distances).
  - Incorporation of a nt is a probabilistic event.
- What happens if I have a homopolymer string?
  - The [H+] is ~ double
  - Rate of release is fast
  - Persistence of H+ is large (solution is not buffered)
- What happens if I did not QUITE wash away all of the previous nucleotide?



Incomplete Extension

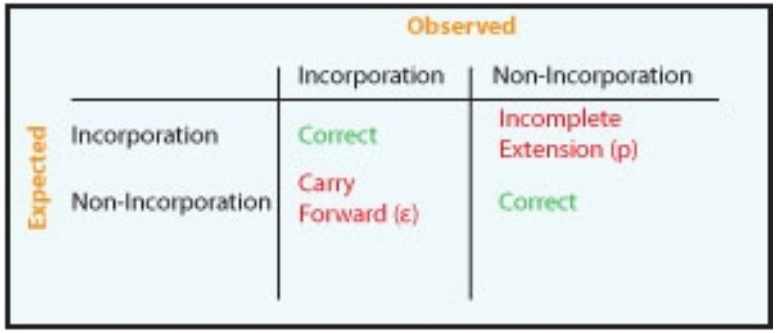
- If the sequence is AATTCGGG
  - There are 8 bases
  - It would only require 4 positive flows to process through them
- Incomplete extension in the context of incorporation events



Test fragment A requires 72 incorporations, but has an error spike at the 59th step (p = 0.012)

Advanced phase problems also exist

- Washing is also incomplete, scavenged nucleotides will extend the sequence but won't be correctly accounted.
  - Response patterns may vary (non-synchronicity).
  - A delayed phase is a kind of echo, something that already happened. Advanced phase Is more complicated to model so iterative passes through the data are performed.



Phase correction

- Model the probability events, for a flow (n)
- Determine the normalized signal incorporation value y_n
- You know the nucleotide that corresponds to a given flow
- You have the ideal signal (a) for the flow of that nucleotide (there are always empty wells to provide the background response, and test fragments that are spiked in to provide a known response)
- There is some probability of incomplete extension (p).
- The value of p give the fraction of strands that are in-phase, from which the subpopulations that are out of phase (w_{n-4}, w_{n-8}…) can be determined.
- $u_n = (y_n – w_{n-4}*a_{n-4} – w_{n-8}*a_{n-8} – w_{n-12}*a_{n-12}….)/w_n$

Signal Decay

- Signal decays over time
  - Polymerase degrades, sensors overload, etc.
  - On the PGM the decay has been modeled at 0.05% decay per flow step
- The phase solver (CAFIE) – compensates for the slippage since chemistry is not perfect