Purification with Millipore MultiScreen $_{384}$SEQ Plates on Beckman-Coulter's Biomek FX for High Quality Direct Sequencing in High-Throughput Genomics

Wei-Hsien Yang[1], Nathan McDonald[1,3], Jack Leonard[2], Christopher Barbagallo[2], and Hans Fuernkranz[1,3]

[1]DNA Sciences, Inc., [2]Millipore Corporation, [3]Applied Biosystems

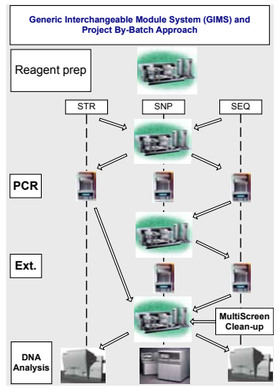
Overview

The authors have developed a post-extension clean-up protocol for sequencing using *Millipore's* MultiScreen$_{384}$ SEQ plates that is high-throughput, cost effective, amenable to automation and results with high consistency in outstanding sequencing quality scores (phred >= 35 for approx. 90% of samples).

DNA Sciences, Inc.

a genetics discovery company focused on identifying the genetic basis of disease susceptibility and response to drug treatment.

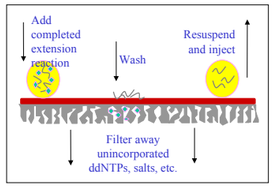
- Sequencing
  - High-Quality Sequencing for polymorphism discovery, incl.
  - Heterozygote detection (DYEnamic ET Dye Terminator on MegaBACE)
- SNP Genotyping
  - High-Throughput multiplexed SNP genotyping (Single Base Extension, SNaPshot on ABI3700)
- STR Genotyping
  - Linkage analysis studies, (LMS HD5 [ABI] on MegaBACE)



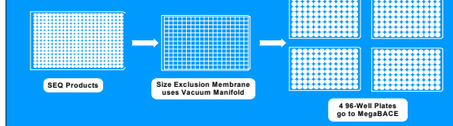
DNA Sciences, Inc. utilizes a 'Generic Interchangeable Module System (GIMS)' that allows for rapid changes in project type (sequencing or genotyping) without production delays.

Millipore's MultiScreen$_{384}$SEQ Plates setup on *Beckman-Coulter's* Biomek FX is an integral part of *DNA Sciences'* 'GIMS' that has produced sequencing quality scores of phred 50.

- The size-exclusion method efficiently removes salts and unincorporated dye terminators through vacuum filtration.
- No sephadex gel.
- No centrifugation.
- MultiScreen$_{384}$SEQ Plates easily adaptable to liquid-handling instruments.
- Consistent high quality results.



MultiScreen$_{384}$SEQ Plate Cleanup Process:



- Throughput: ~3 384-well plates / hour
- Currently using 2 Vacuum Manifolds on one FX Deck
  - *Throughput can be significantly increased by using 4 or more manifolds.*
- 96-Well Plates get run on MegaBACE 1000

Comparison Study of Ethanol and MultiScreen$_{384}$ SEQ Clean-up

Size of the study: Forward and Reverse sequences of 18 regions on 96 genomic samples → ~ 3600 samples.

- Prepared 3 sets of the above samples:
  - **Set 1** = Half reaction sequences for Ethanol Clean-up
  - **Set 2** = Quarter reaction sequences for Ethanol Clean-up
  - **Set 3** = Quarter reaction sequences for MultiScreen$_{384}$ SEQ Clean-up

Compared the results of Average Phred, > Phred 30 Read length, Pass Rate, and Number of resequenced plates.

Set 1

ROI ID	Half Rxn Ethanol Cleanup Avg Phred	Avg Phred 30 RL	Pass Rate
fnar1_4-A yq	37.92	367.09	97%
fnar1_4-A zq	39.35	400.37	98%
fnar1_6 yq	39.47	377.26	93%
fnar1_6 zq	35.87	302.61	93%
fnar1_9 yq	37.21	338.9	95%
fnar1_9 zq	35.83	306.95	88%
fnar2_2 yq	37.26	374.83	99%
fnar2_2 zq	38.51	376.88	98%
fnar2_3 yq	41.86	410.79	97%
fnar2_3 zq	39.38	383.49	95%
fngr2_2 yq	39.24	332.17	95%
fngr2_2 zq	40.2	349.38	86%
ft12a_1 yq	36.89	336.82	98%
ft12a_1 zq	39.78	400.66	98%
ft12a_7-A yq r1	15.65	0	0%
ft12a_7-A yq r2	0	0	0%
ft12a_7-A yq r3	41.49	394.22	94%
ft12a_7-A zq r1	38.97	383.58	76%
ft12a_7-A zq r2	42.17	436.31	76%
ft12a_7-A zq r3	42.22	423.2	75%
ip10_4a yq	38.64	333.96	93%
ip10_4a zq	43.31	386.95	91%
mag_10-B yq	35.44	329.11	91%
mag_10-B zq	39.33	385.29	93%
mag_6 yq	38.52	364.95	98%
mag_6 zq	40.85	375.01	97%
mcp1_2 yq	41.38	350.53	86%
mcp1_2 zq	41.49	358.03	83%
mcp1_pro yq	38.85	302.66	94%
mcp1_pro zq	42.53	334.03	98%
mpo_12-A yq r1	41.1	404.83	99%
mpo_12-A yq r2			
mpo_12-A zq r1	38.82	372.05	98%
mpo_2 yq	36.96	239.65	95%
mpo_2 zq	38.25	291.5	91%
tnf_pro_a2 yq	41.45	370.51	91%
tnf_pro_a2 zq	29.84	220.71	86%
tnf_pro_b2 yq	35.72	223.55	92%
tnf_pro_b2 zq	39.18	299.04	94%
tnf_pro_c2 yq r1	34.87	359.52	98%
tnf_pro_c2 yq r2	35.41	359.24	98%
tnf_pro_c2 zq r1	25.79	87.48	36%
tnf_pro_c2 zq r2	29.4	117.87	80%
tnf_pro_c2 zq r3	35.96	387.83	98%
Averages:	**37**	**323**	**87%**

Set 2

ROI ID	Quarter Rxn Ethanol Cleanup Avg Phred	Avg Phred 30 RL	Pass Rate
fnar1_4-A yq	40.53	386.33	93%
fnar1_4-A zq	41.45	419.2	98%
fnar1_6 yq	42.42	373.77	93%
fnar1_6 zq	39.44	339.58	91%
fnar1_9 yq	37.63	301.23	91%
fnar1_9 zq	39.15	323.14	93%
fnar2_2 yq	39.35	414.91	98%
fnar2_2 zq	40.94	398.09	94%
fnar2_3 yq	42.31	408.01	92%
fnar2_3 zq	35.12	298.71	94%
fngr2_2 yq	40.1	338.87	89%
fngr2_2 zq	38.3	295.86	78%
ft12a_1 yq	37.48	371.95	98%
ft12a_1 zq	41.14	383.16	88%
ft12a_7-A yq r1	0	0	0%
ft12a_7-A yq r2	0	0	0%
ft12a_7-A yq r3	40.01	382.05	97%
ft12a_7-A zq r1	40.55	400.58	76%
ft12a_7-A zq r2	34.63	293.07	68%
ft12a_7-A zq r3	40.56	409.43	75%
ip10_4a yq	41.28	359.4	95%
ip10_4a zq	43.09	372.77	80%
mag_10-B yq	33.29	301.25	92%
mag_10-B zq	38.64	378.15	97%
mag_6 yq	38.07	368.7	97%
mag_6 zq	40.35	395.19	98%
mcp1_2 yq	39.56	326.16	80%
mcp1_2 zq	41.38	342.22	80%
mcp1_pro yq	37.92	303.81	97%
mcp1_pro zq	43.03	333.97	97%
mpo_12-A yq r1	40.25	385.62	94%
mpo_12-A yq r2	39.49	408.77	94%
mpo_12-A zq r2	38.69	387.47	98%
mpo_2 yq	32.34	184.32	86%
mpo_2 zq	36.38	248.52	95%
tnf_pro-a-2 yq	42.97	413.47	97%
tnf_pro-a-2 zq	32.5	286.19	88%
tnf_pro-b-2 yq	37.88	283.87	97%
tnf_pro-b-2 zq	38.74	290.11	88%
tnf_pro-c-2 yq	35.37	353.19	98%
tnf_pro-c-2 zq	36.45	357.45	91%
tnf_pro-c-2 zq r1	25.47	59.73	10%
tnf_pro-c-2 zq r2	27.02	62.57	8%
tnf_pro-c-2 zq r3	36.28	346.12	92%
Averages:	**37**	**321**	**83%**

Set 3

ROI ID	Quarter Rxn MultiScreen384SEQ Cleanup Avg Phred	Avg Phred 30 RL	Pass Rate
fnar1_4-A yq	41.95	413.66	98%
fnar1_4-A zq	39.5	395.27	95%
fnar1_6 yq	42.5	373.61	91%
fnar1_6 zq	38.88	354.38	97%
fnar1_9 yq	39.36	379.63	97%
fnar1_9 zq	38.61	380.21	97%
fnar2_2 yq	37.95	385.32	98%
fnar2_2 zq	42.28	430.63	98%
fnar2_3 yq	43.09	445.2	98%
fnar2_3 zq	38.5	352.94	90%
fngr2_2 yq	34.96	270.93	88%
fngr2_2 zq	35.97	277.05	83%
ft12a_1 yq	37.44	365.39	93%
ft12a_1 zq	44.93	452.26	98%
ft12a_7-A yq r1	40.75	411.51	97%
ft12a_7-A yq r2	*no resequencing necessary*		
ft12a_7-A yq r3	*no resequencing necessary*		
ft12a_7-A zq r1	40.03	403.13	75%
ft12a_7-A zq r2	*no resequencing necessary*		
ft12a_7-A zq r3	*no resequencing necessary*		
ip10_4a yq	39.4	360.73	98%
ip10_4a zq	42.75	386.13	95%
mag_10-B yq	35.82	335.52	93%
mag_10-B zq	41.93	422.06	97%
mag_6 yq	40.43	391.16	93%
mag_6 zq	37.93	357.41	97%
mcp1_2 yq	41.11	366.75	88%
mcp1_2 zq	42.69	393.69	89%
mcp1_pro yq	39.44	314.03	91%
mcp1_pro zq	42.31	327.21	92%
mpo_12-A yq r1	37.44	372.12	93%
mpo_12-A yq r2	*no resequencing necessary*		
mpo_12-A zq r1	39.77	335.67	95%
mpo_2 yq	31.25	181.96	93%
mpo_2 zq	32.52	234.39	98%
tnf_pro-a-2 yq	39.82	364.73	98%
tnf_pro-a-2 zq	32.04	292.7	93%
tnf_pro-b-2 yq	40.47	322.09	97%
tnf_pro-b-2 zq	35.7	271.96	97%
tnf_pro-c-2 yq	39.46	336.16	95%
tnf_pro-c-2 yq r1	*no resequencing necessary*		
tnf_pro-c-2 yq r2	*no resequencing necessary*		
tnf_pro-c-2 zq	32.73	282.93	92%
tnf_pro-c-2 zq r1	*no resequencing necessary*		
tnf_pro-c-2 zq r2	*no resequencing necessary*		
tnf_pro-c-2 zq r3	*no resequencing necessary*		
Averages:	**39**	**353**	**94%**

Summary of Results

	1/2 rxn ethanol	1/4 rxn ethanol	1/4 rxn MultiScreen$_{384}$SEQ
Average Phred	37	37	39
Average > Phred 30 Readlength	323	321	353
% Pass Rate*	87%	83%	94%
# of Resequenced Plates**	7	8	0

*Pass rate = number of samples with Phred score > 30 and at least 100 bp of Phred 30 Read length / total number of samples

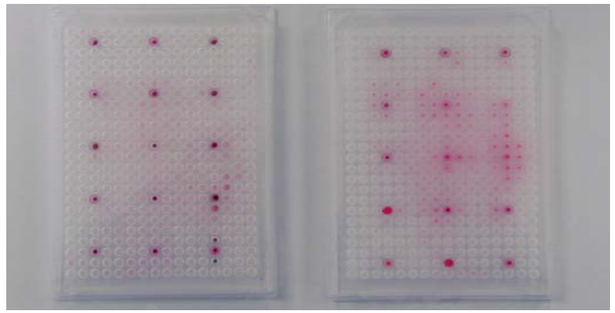
**Sample plates resequenced due to < 80% pass rate or contamination of samples.

Comparison Study Conclusions

MultiScreen$_{384}$SEQ Plates Clean-up procedure resulted in
- Higher Phred score
- Longer good quality sequence
- Higher sample pass rate
- No cross contamination of samples
- No resequences needed

Ethanol Clean-up procedure, although produces satisfactory results,
- Lacks consistency in performance, especially among different users.
- Has higher occurrences of cross-contamination → more resequences needed.

Contamination Issue with Ethanol Clean-up Procedure

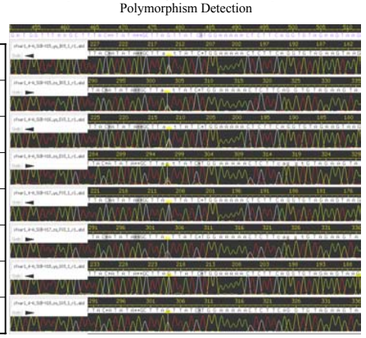


Cross contamination occurs after inversion of ethanol and extension products mixture in sealed plate during ethanol clean-up procedure.

This is demonstrated in the picture below where the "pink wells" are the results of crossing over of red food coloring + ethanol mixture from source wells into the adjacent wells during the inversion process.

Phred 40	Phred 50	Ave. Phred	Genotype
334	238	42.7	CG
328	230	41.4	CG
345	250	43.4	CG
111	46	35.0	CG
342	234	43.2	CC
292	157	40.2	CC
249	238	42.7	GG
377	301	43.4	GG

Sample of MultiScreen$_{384}$SEQ Phred 50 sequences for Polymorphism Detection



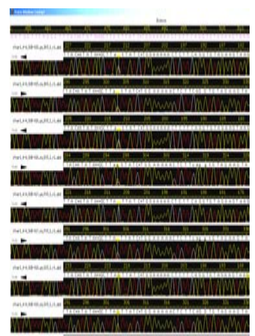
Benefits of High Quality Sequencing Data

Polymorphism Discovery
- Polymorphism Detection at phred 50 is significantly more accurate and automatable than at phred 30.
- CG polymorphism shown here at base 478 of consensus sequence (yellow tag) was called correctly by software.

Accuracy of Results
- Fewer false positives (less automated mis-calls)
- Fewer questionable calls, i.e. less human interference required.
- Data confidence is higher:
  - Phred 30 = 1 mistake in 1,000 bases
  - Phred 40 = 1 mistake in 10,000 bases
  - Phred 50 = 1 mistake in 100,000 bases

Significant cost savings - fewer data analysts are needed

Problem Detection
- Easier to spot sequencing peculiarities



Overall Conclusions

- The MultiScreen$_{384}$SEQ Clean-up procedure has allowed *DNA Sciences, Inc.* to detect polymorphisms in disease associated genes faster, cheaper, and with greater data confidence.
- The MultiScreen$_{384}$SEQ set-up on Biomek FX provides a high level of walk-away confidence.
- Excellent recovery of products from MultiScreen$_{384}$SEQ Plates can allow further miniaturization of sequencing reactions, which leads to more cost savings.
- To further increase throughput, a Biomek FX with dual pod, 384/96 head design can be implemented.