NEWBORN SCREENING **ONTARIO**
DÉPISTAGE NÉONATAL **ONTARIO**

A quality improvement initiative to investigate the utility of an open platform assay and Next Generation Sequencing for molecular genotyping in Cystic Fibrosis (CF) screening by Newborn Screening Ontario (NSO)

Ed Yeh(1), Janet Marcadier(1), Maria Karaceper (1), Hema Liyanage(2), Pranesh Chakraborty(1,3), Dennis Bulman(1,3)

[1] Newborn Screening Ontario, [2] Agena Bioscience, San Diego, CA, [3] Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON

Background

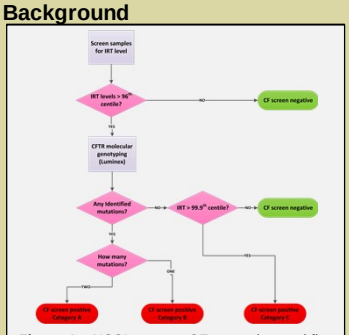


Figure 1 – NSO's current CF screening workflow

NSO screens for CF using a two tiered workflow (Figure 1) – measurement of immunoreactive trypsinogen (IRT) levels followed by molecular genotyping for those with elevated IRT levels. Currently, genotyping is done using xTAG CF 39v2, a commercially available assay (Luminex) with a fixed panel that interrogates 39 mutations commonly associated with CF (Table 1). NSO refers screen positives for diagnostic follow up within three categories - **category A** individuals carry two CFTR mutations, **category B** individuals carry one CFTR mutation and **category C** individuals do not possess any detectable mutations but have a very high IRT level (>99.9th centile). Both category B and C individuals remain at risk for CF, as the genotyping assay only ascertains common CFTR mutations (and no deletions/gene rearrangements).

A review of NSO CF screen positive referrals during the period of 04/2008-04/2013 revealed trends that are also observed by other CF screening programs – a low CF positive predictive value in category B and C screen positives, while the vast majority of referrals are category B and C (Table 2). To address these issues, NSO is investigating **1) the use of an open platform genotyping assay** and **2) usage of next generation sequencing (NGS) for category B and C screen positives.** An open platform provides the ability to create a custom screening panel catered to NSO's target population with the benefit of cost effectiveness; NGS may provide data to reduce the number of referred category B and C screen positives and/or provide basis for an alternate follow up workflow.

ΔF508	A455E	R1162X
Δ1507	1717-1G>A	3659delC
G542X	R560T	3849+10kbC>T
G85E	R553X	W1282X
R117H	G551D	N1303K
621+1G>T	1898+1G>A	5/7/9T
711+1G>T	2184delA	F508C
R334W	2789+5G>A	I507V
R347P	3120+1G>A	I506V
1078delT*	1898+5G>T	V520F
394delTT	2183AA>G	A559T
Y122X*	2307insA	S549N
R347H	Y1092X	S549R
M1101K*	3876delA	3905insT
S1255X*(Ex20, Ex19)		
ACMG recommended	variants	*yet to be detected by NSO

Table 1 – Mutations in current NSO CFTR genotyping panel

CF Classification	Category A	Category B	Category C	TOTAL
YES	110 (94.83%)	27 (1.60%)	2 (0.51%)	**139 (6.34%)**
NO	0	1 (0.06%)	361 (91.39%)	**362 (16.50%)**
VARIANT	1 (0.86%)	62 (3.68%)	2 (0.51%)	**65 (2.96%)**
INCIDENTAL	0	1554 (92.34%)	0	**1554 (70.83%)**
OTHER	0	27 (1.60%)	29 (7.34%)	**56 (2.55%)**
PENDING	5 (4.31%)	12 (0.71%)	1 (0.25%)	**18 (0.82%)**
TOTAL	**116 (5.29%)**	**1683 (76.71%)**	**395 (18.00%)**	**2194**

Table 2 – NSO screen positive referrals (2008-2013)

CF Classification meaning: YES – CF diagnosis; NO – negative CF diagnosis; VARIANT – "grey zone", atypical CF, CFTR related; INCIDENTAL - 1 disease causing allele identified; OTHER – declined follow up, lost to follow up, deceased, etc.; PENDING – results pending

Methods

Genotyping: Samples - Samples screened for CF by NSO during the period of Jan-Apr/2015 were used. Verification samples were obtained from the Coriell Biorepository (Coriell Institute), previously tested NSO samples and the Hospital for Sick Children (Toronto). A synthetic DNA sample with the S1255X allele was manufactured by IDT. **DNA Extraction from DBS -** Extractions were performed using a methanol fixation/boil method; the same sample was used for both Luminex and iPlex assays. **Genotyping Assay** - Genotyping by NSO is performed using the xTAG CF 39v2 kit as per vendor guidelines (Luminex Corp). A custom panel was designed based on iPlex/MassARRAY® (Agena Bioscience) technology against the alleles listed in Table 1 and 3. Oligos were manufactured by IDT. Reaction setup was carried out manually or using Biomek NX/4000 liquid handlers (Beckman Coulter). Assay verification was performed using known positive samples (see Samples). QC samples (positives, negatives, blanks) were included in every assay. **Data Analysis** - Analysis of Luminex data was done using TDAS v2.0 software (Luminex); MassARRAY® data using Typer 4.0 software (Agena Bioscience) and in house algorithms.

NGS: Samples – NSO archived category B and C samples with a negative CF diagnosis were used. CFTR control DNA samples NA07441, NA11275 and NA11277 were obtained from Coriell. Anonymized DNA samples were used during optimization and verification of sequencing workflows. **DNA Extraction from DBS -** DNA extractions from a single 3.2mm punch were performed using in house developed methods: a) wash/boil method or b) proteinase K/salt precipitation method. QC measurements included quantification using Picogreen and quality/integrity using Nanodrop/agarose gel electrophoresis. **Library preparation** - Two library preparation chemistries were used and compared: 1) TruSeq Custom Amplicon (TSCA, Illumina) - a PCR based enrichment approach using a custom panel against 21 genes including CFTR. A design targeting exons and 25bp of intronic padding was created in conjunction with Illumina. 2) Nimblegen SeqCap EZ Choice (Roche) - an oligo based capture approach using a custom design against CFTR only (targeting exons and 50bp of intronic padding). Workflows were performed as per vendor guidelines; samples were indexed and multiplexed for subsequent sequencing. QC measurements were performed using Picogreen, Nanodrop and a 2100 Bioanalyzer. **Sequencing –** Sequencing was done on a MiSeq (Illumina) using paired end 2X150bp reads following vendor guidelines for library input. **Analysis pipeline –** MiSeq Reporter was used (TSCA workflows: Banded Smith-Waterman for alignment, GATK for variant calling; Nimblegen workflows: BWA for alignment, GATK for variant calling). Alignments/pileups were visualized with IGV (Broad Institute); variant annotation and filtering were performed using Variant Studio v2.2 (Illumina). Primary filtering: Gene - CFTR, Quality >30, depth >30x, call freq. >25%. Secondary filtering: allele freq. < 1%, intronic variants (outside 25bp padding), presence in unrelated sequenced samples – exceptions are variants with a previously characterized disease association.

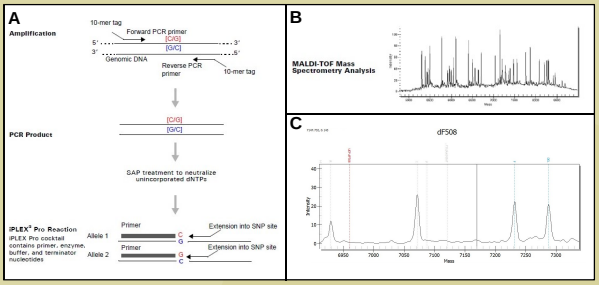
Results – Validation of iPlex/MassARRAY® genotyping assay

CFTR mutation	# alleles found
D1152H	5
Ser489X (Fr. Cdn. mut)	2
c.2657+2_2657_3insA	2
Deletion exon 2-3	2

Table 3 – Recurring CFTR alleles identified in category B and C cases diagnosed with CF – A review of B and C screen positive cases (2008-2013) with a follow up diagnosis of CF revealed specific recurring CFTR alleles. These alleles were included in the custom iPlex panel.

Results – cont'd

Figure 2 –iPlex/Mass-ARRAY® genotyping platform. A) iPlex chemistry generates molecules that differ in mass based on the presence/absence of nucleotide variation. **B)** MassARRAY® analysis resolves molecular products by mass. **C)** Typer software correlates spectrometry data with the panel design to make genotyping calls.



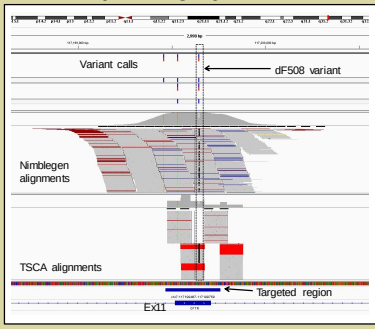
Parallel testing	Luminex	iPlex/Mass ARRAY®	Discordant results: Genotype	Discordant results: Comment
# experiments	15	15	D1152H/+	detected in 3 samples on iPlex
Total samples tested	1094	1113*	c.2173G>A /F508C	c.2173G>A mimics 2307insA allele (due to flanking sequence and design strategy) → panel design modified
# samples with alleles	76	79	5T/8T**	5T on Luminex; 5T/9T on iPlex
NSO's estimated reagent cost	$43500	$17600	7T/8T**	7T on Luminex; 7T/9T on iPlex

*additional QC samples

**8T polymorphism was verified by Sanger sequencing

Table 4 – Validation of iPlex assay – After assay verification, the iPlex assay was used in parallel with the current Luminex testing workflow. iPlex genotyping was 100% concordant with Luminex screen positive results and as predicted, increased the sensitivity of CFTR mutation detection.

CFTR sequencing by NGS



Metric	TSCA	Nimblegen
Read quality	✓	✓
Read depth	(as predicted) < Nimblegen	✓
Unique reads?	reads are duplicates	✓
Read depth uniformity	unequal across target	✓
On target	✓	flanking regions captured
large (>15bp) indel calling	?	likely (existing pipelines)
# variants called	less ✓	more
Workflow (time,complexity)	✓	new streamlined workflow yet to be tested

Figure 3 – Comparison of TSCA vs Nimblegen – Left is a screenshot from IGV showing alignments from two samples carrying a dF508 mutation. Both are correctly called. The table above lists a comparison of key metrics between the library prep chemistries. (✓ – denotes better)

Sample ID	Library prep	Filtered variants	NGS sequencing: Call freq %	Depth	dbSNP	Allele freq	Legacy allele	Sickkids db	IRT	NSO CF results: Genotype	NSO CF results: T	NSO CF results: Category	CF?
CFDBS1	TSCA	c.1585-1G>A	49.3	2127	rs76713772	0.05	1717-1G>A	CFTR2	62.4	1717-1G>A	7	B	no
CFDBS2	TSCA	c.1210-12_1210-11delTT	45.4	651		0	5T	T tract	112.3	---	5/7	C	no
CFDBS3	TSCA	c.1516_1518delATC p.I506del (p.I507del)	45.4	1785	rs121908745	0	dI507	CFTR2	58	dI507	7	B	no
CFDBS4	TSCA	c.1652G>A p.G551D	51.3	4152	rs75527207	0	G551D	CFTR2	54.1	G551D	7	B	no
CFDBS5	TSCA	c.1210-12_1210-11delTT	43.8	658		0	5T	T tract	121.5	---	5/9	C	no
		c.1210-12_1210-11dupTT	43.8	658		0	9T	T tract					
CFDBS6	TSCA	---	---	---	---	---	---	---	832.1	---	7	C	deceased
CFDBS3_2	TSCA	c.1516_1518delATC p.I506del (p.I507del)	46.3	1986	rs121908745	0	dI507	CFTR2	58	dI507	7	B	no
CFDBS_7	TSCA	c.1000C>T p.R334W	50	3728	rs121909011	0.05	R334W	CFTR2	59.7	R334W	7	B	no
CFDBS8	TSCA	c.3276C>A p.Y1092X	53.4	4703	rs121908761	0	Y1092X	CFTR2	52.7	Y1092X	7	B	no
CFDBS_9	TSCA	c.1210-13G>T	96.9	415	rs10229820	12	9T	T tract	72	R1162X	9	B	no
		c.3484C>T p.R1162X	27.2	463	rs74767530	0	R1162X	CFTR2					
CFDBS_10	TSCA	c.1210-13_1210-12insTT	29	858	rs200454589	0	9T	T tract	118.7	---	7/9	C	no
CFDBS_11	TSCA	c.1210-13G>T	41.2	818	rs10229820	12	9T	T tract	58.5	dF508	7/9	B	no
		c.1520_1522delTCT dF508	45.4	2194	rs121909001	1	dF508	CFTR2					
NA07441_250	TSCA	c.489+1G>T	53.8	145	rs78756941	0	621+1G->T	CFTR2	NA	NA	NA	NA	yes
		c.1210-13G>T	48.7	315	rs10229820	12	9T	T tract					
		c.2988+1G>A	32	169	rs75096551	0.05	3120+1G->A	CFTR2					
NA11275b	Nimblegen	c.1210-13G>T	54.4	4114	rs10229820	12	9T	T tract	NA	NA	NA	NA	yes
		c.1520_1522delTCT dF508	39.3	4981	rs121909001	1	dF508	CFTR2					
		c.3527delC	46.9	4994	rs121908747	0	3659delC	CFTR2					
NA11277	Nimblegen	c.1516_1518delATC p.I506del (p.I507del)	40.1	4977	rs121908745	0	dI507	CFTR2	NA	NA	NA	NA	no

Table 5 – Summary of NGS sequencing – The TSCA protocol was modified and optimized for starting material from DBS samples (metrics achieved were similar to those with high quality DNA extracted from whole blood - data not shown). 11 samples previously screened as either category B or C with a negative CF diagnosis were library prepped and sequenced over 2 sequencing runs (the library for CFDBS3 was sequenced twice and gave similar metrics for both runs). Optimization of Nimblegen library preparation using DBS starting material is pending; preliminary results with DNA from whole blood indicate the design is meeting performance benchmarks (metrics not shown).

Conclusion

A custom panel based on iPlex/MassARRAY® technology for CFTR genotyping was developed and validated. This provides a flexible and cost effective alternative to the current genotyping assay. While increasing the content increases the sensitivity for CFTR mutation detection, it also increases the number of category B screen positives (that have a low positive predictive value for CF). Sequencing of CFTR using NGS can provide additional molecular information that can be used to reduce the number of "true" screen positives (Figure 4). Remaining challenges include incorporating the NGS workflow within the current TAT for reporting and the ability to confidently called novel indels.

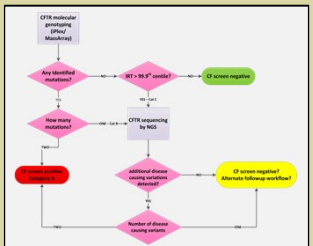


Figure 4 – Future molecular genotyping workflow?