Screening of Alu Sequences in the Human 1.7k Clone Set

Monika Sharma, Eric Ho and Neil Winegarden

Toronto General Research Institute, CGC-Microarray Center, Toronto, ON

Abstract

Alu sequences are repetitive elements often found in DNA. We have analyzed our human 1.7k clone set for any such repetitive elements that may be interfering with the hybridization studies performed in the microarray experiments.

Several hybridization experiments were carried out in order to investigate for possible presence of Alu sequences. Initial experiments (Direct labeling with universal RNA, labeling of cold Alu-oligo with tdt (terminal transferase) method and other chemical labeling kits) indicated the presence of several Alu positive clones, however a BLAST searches of these sequences with the consensus Alu sequence showed no relative homology. At this point we stopped to trouble shoot for the absence of Alu sequence in the Alu positive clones. Most of these clones were further sequenced up and down stream of the original sequence in order to reach the cloning sites. This was done to see if the Alu sequence was sitting outside our original primer targets. However, a positive hit to Alu was still not found.

Another approach for detecting Alu sequences on the arrays, was to hybridize the array with alu consensus-oligo (55mer) labeled at the 5'end with cy5. In theory this oligo should specifically hybridize to any Alu sequences. In total 125 clones had been identified as potentially containing Alu sequences within our human 1.7k clone set. 22 clones showed positive hybridization to the consensus oligo and for these sequences there was a confirmed full or partial homology to the Alu oligo as judged by an alignment. The remaining 103 clones had double amplification products band or mixed sequences. Alu sequences were found in both complete and incomplete sequences of all 125 clones. Blasting sequences with Alu oligo (55mer) using emboss (Smith-Waterman local alignment) for the other 103 clones shows alignment for as little as 5-6 bp; thus we draw a conclusion that it's possible for short sequences (5-6 bp) to be hybridized at 37°C.

Hybridization in the presence of excess cold alu (unlabelled) oligo was carried in order to see if we could effectively block hybridization of labelled cDNAs to the Alu sequences. This cold Alu oligo was added in excess to see if we could suppress the signal from the labelled 55-mer-consensus sequence. Adding 10x as much cold Alu oligo in the hybridization solution resulted in inhibition of Alu sequences without interfering with the true expression data from experimental sample. As a result we now recommend the addition of a cold Alu consensus sequence as a blocking agent during hybridization.

Introduction

Repetitive DNA sequences can be problematic for cDNA microarray analysis. Such repetitive sequences can mask actual signals and prevent the identification of true gene expression changes. There are several different family of repetitive sequence. In specific we have examined the Alu sequence in our 1.7k human microarrays. The Alu repeat is itself represented by different families each of which are present in Cot-1 DNA, which is why many microarray protocols suggest the addition of this DNA as a blocking agent during microarray hybridization. Alu repeats are interspersed repetitive DNA elements specific to primates that are present in 500,000 to 1 million copies. It is reported that alu sequences encode functional binding sites for retinoic acid receptors, which are members of the nuclear receptor family of transcription factors. Very little is known about evolutionary relationships and functional significance of repetitive elements within the genomes of individual species.

The existence of Alu sequences (a repetitive sequence) in cDNA clones within our human 1.7k clone set can be a problem. These sequences can hybridize to the labeled probe as if there was real expression and can skew the results of that particular study by having false up or down regulation. Repetitive sequences such as Alu, PTR5, PTR7, MSR1, MER13, MER22, MER32 and L1 represented a considerable portion of contamination within each of the sequence found to be contaminated.

The UHN Microarray Center wanted to investigate the impact of these repetitive sequences on our arrays. We also wanted to identify problematic clones so that they can be cleaned up for future versions of our arrays.

In this study we focused on a highly repetitive Alu sequence that is 55 bp in length. We identified the clones contaminated with this consensus Alu sequence and then worked on the solution for suppressing this sequence without interfering with the actual or true expression signal for that particular gene. To address these questions, thorough analysis of the 1.7k clone set was required. Physical hybridizations as well as bioinformatics analysis was performed for this study. 19k arrays, Mouse arrays and Yeast arrays will be evaluated next for any alu contamination within the clone sets.

Methods and Materials

Tailing With tdt Method: Terminal transferase from Roche was used for the labeling of 55-mer alu-oligo to assess alu sequences in the 1.7k clone set.

Hybridization Experiments: Various conditions were used to identify potential Alu containing clones, and to determine if Alu dependent signals could be supressed. Hybridization was carried at higher temperatures such as 42°C, 60°C and 65°C as oppose to the standard 37°C hybridization temperature to confirm the specificity of Alu positive clones. Washes were done at higher temperature and with more stringent conditions. Different labeling methods were compared as well as probe purification methods. Hybridization in the presence of excess inhibitors such as Cot-1DNA as well as excess cold alu were tested for the suppression of alu sequence.

Designing, Synthesis of Primers and Sequences: For all of the suspected alu containing clones primers were designed using the primer program of the Wisconsin GCG bioinformatics package to walk each of the sequences further to locate Alu sequence. The primers were synthesized in house on a GeneMachines Polyplex and sequencing was performed on an ABI Prism 3100.

Sequence Results: All of the primer walked sequences were blasted against the original Alu probe sequence in order to see if regions of homology could be found in these suspected clones. Two methods were used: BLASTing two sequences using the NCBI BLAST server and Emboss (Smith Waterman alignment). Also all results were compared with the NCBI searches to see if other labs have reported presence of Alu in the same clones as us.

Solution for the Problem: We tested to see if we could remove any Alu based signal by hybridization at 65°C. We also attempted to suppress the Alu signals by including cot-1 DNA in excess in the hybridization solution. Both of these produced unsatisftory results. Our final method involved using 10x as much cold Alu oligo in the hybridization buffer. This suppressed the signals obtained, but this may not be an accurate representation of all Alu variants.

Alu original probe / Alu-reverse-compliment / Alu original-Cy5-5' were obtained from Cortec and there sequences are as follows:

5'-GGC CGG GCG CGG TGG CTC ACG CCT GTA ATC CCA GCA CTT TGG GAG GCC GAG GCG G-3'

5'-CCG GCC CGC GCC ACC GAG TGC GGA CAT TAG GGT CGT GAA ACC CTC CGG CTC CGC C-3'

Cy5-5'-GGC CGG GCG CGG TGG CTC ACG CCT GTA ATC CCA GCA CTT TGG GAG GCC GAG GCG G-3'

All other labeling reagents were used as per our standard protocol that is our homemade reagents and stocks from Cortec/Invitrogen and Sigma-Aldrich.

Results

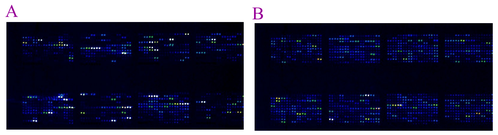

Fig 1: To investigate if same spots will light up when 1.7k array is hybridized with universal cDNA probe in comparison to Alu labeled probe. Hybridization with Alu labeled probe on Cy5 channel (A) and with cDNA from universal RNA on Cy3 channel (B) were set up for overnight at 37°C. Hybridization with Alu oligo shows strong signal intensities for different genes; produced a strong signal on genes that are weakly expressed with universal cDNA. A possibility that Alu sequence may be present.

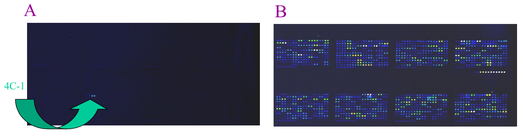

Fig 2: To investigate if same spots will light up as with Alu-oligo labeled probe: Hybridization with probe 4C-1 and with universal cDNA is performed. Probe purification for both A and B is achieved with GFX columns. A) Oligo 4C-1 is labeled using terminal transferase (tdt) enzyme and Cy3 d-CTP. Hybridization to only corresponding spots occurred. B) Human universal RNA is labeled using Direct-labeling method with Cy5 d-CTP. Good hybridization is observed.

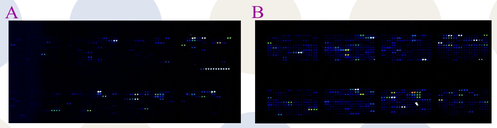

Fig 3: Alu-original probe is labeled with terminal transferase (tdt) and hybridization at 37°C for 1 hour vs. overnight. Probe clean up with GFX columns and Microcon® columns are compared. A) Probe clean up with GFX columns and Hybridization at 37°C for overnight shows more specific hybridization as oppose to probe purification with Microcon® columns. B) Hybridization at 37°C for 1hr and probe purification with Microcon® columns produced hybridization to almost every spot. Clones other than the Alu positive clones lit up.

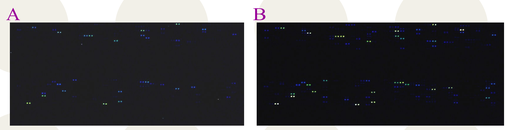

Fig 4: Alu-Cy5-5' labeled oligo is hybridized on 1.7k array at 2 different concentrations (100 pmol/ul and 10 pmol/ul) to determine any clones contaminated with Alu sequence. Only those clones contaminated with Alu sequence should light up. At 10 pmol/ul (A) 21 clones lights up as Alu positive clones. All 39 clones produced from the 100-pmol/ul (B) hybridization are investigated for Alu sequence. 22 out of 39 are with full Alu sequence. 17 other clones are found to have only partial sequence of Alu.

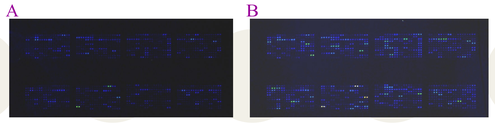

Fig 5: Effect of Cot-1 DNA in the hybridization buffer. To investigate if Cot-1 DNA can suppress hybridization to Alu sequence and to find out if universal cDNA probe will hybridize to Alu oligos (spotted at different concentrations on a chip). Labeled probe is obtained with direct labeling method followed with GFX columns for probe clean up. A) Cot-1 DNA in the hybridization buffer inhibited hybridization to Alu oligo (spotted at different concentration) as well as overall expression is relatively inhibited. B) Universal cDNA lights up Alu oligo and a very weak hybridization is expressed all around the array.

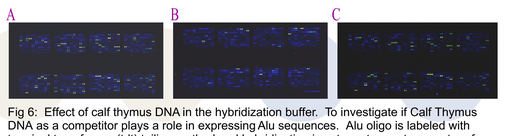

Fig 6: Effect of calf thymus DNA in the hybridization buffer. To investigate if Calf Thymus DNA as a competitor plays a role in expressing Alu sequences. Alu oligo is labeled with terminal transferase (tdt) tailing method and hybridization is set up at room temperature for overnight. A) As expected Alu oligo (spotted at different concentrations) didn't hybridized with standard conditions (in the presence of both calf thymus DNA and yeast t-RNA). B) In the presence of cot-1 DNA overall expression to the entire array is relatively low. Weak hybridization to the spotted Alu oligo is present. C) In the presence of only Yeast t-RNA as a competitor in the hybridization buffer resulted in a somewhat weaker hybridization compare to the standard conditions.

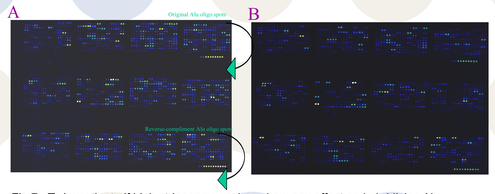

Fig 7: To investigate if high stringency washes show any effect on hybridizing Alu sequences. A wash at 50°C vs. 65°C is compared. cDNA probe is generated from human universal RNA with Direct labeling following probe purification with GFX columns. A) Universal cDNA probe hybridized to both original and reverse-compliment Alu sequence spots. B) Washing at 65°C in the presence of Calf Thymus DNA in the hybridization buffer shows some inhibition, but overall expression is comparable to 50°C washes for highly expressed genes.

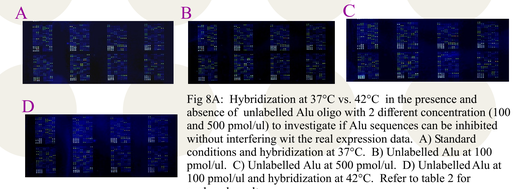

Fig 8A: Hybridization at 37°C vs. 42°C in the presence and absence of unlabelled Alu oligo with 2 different concentration (100 and 500 pmol/ul) to investigate if Alu sequences can be inhibited without interfering wit the real expression data. A) Standard conditions and hybridization at 37°C. B) Unlabelled Alu at 100 pmol/ul. C) Unlabelled Alu at 500 pmol/ul. D) Unlabelled Alu at 100 pmol/ul and hybridization at 42°C. Refer to table 2 for analyzed results.

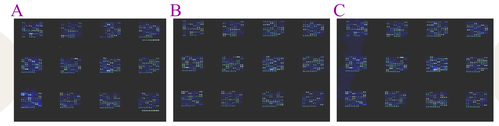

Fig 8B. Universal RNA is labeled with direct labeling method and hybridized in the presence / absence of 100pmol/ul and 500pmol/ul cold Alu oligo in the hybridization buffer. This results in 100% inhibition to Alu original and reverse-compliment spots. A) With standard conditions Alu original and reverse-compliment oligos lights up B) Hybridization is as bright as with standard conditions in the presence of unlabelled Alu (100 pmol/ul) with inhibition to the Alu sequences. C) Hybridization is as bright as with standard conditions in the presence of unlabelled Alu at 500 pmol/ul with inhibition to the Alu sequences.

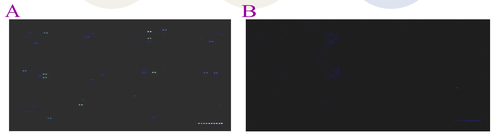

Fig 8C: Labeled Cy5-5' Alu oligo is hybridized in the presence / absence of cold Alu oligo. A) In the absence of unlabelled Alu in excess, reverse-compliment oligos as well as clones with Alu sequences appeared. B) In the presence of unlabelled Alu in excess, all clones with Alu sequences (full or partial) are suppressed. Alu oligos spotted are suppressed as well.

Clone ID	Accession Number	Intensity (bkg subtracted) 100ng/ul(37°C)	100ng/ul(42°C)	500ng/ul(37°C)	Standard(37°C)	Alu-cy5-5'(37°C)
8 F4	N57713	reddish white (35227)	reddish white(41814)	reddish white(44873)	white(46080)	reddish yellow(31986)
2 F2	H09051	green(16909)	green(16712)	yellowish(30885)	red yellow(26841)	green(11042)
10 F10	N66935	green(17538)	green(12129)	green(14376)	red(32896)	green(17481)
10 H10	N62695	yellow red(25764)	yellow red(30414)	red(36984)	white(45292)	white(43592)
3 D12	T83006	yellow(24092)	yellow(15511)	yellow(28569)	red white(36051)	red(32159)
7 F4	214904*	green(12600)	green(6981)	green(5887)	white(34533)	yellow green red(29262)
16 D1	AA149475	red(27293)	red(22841)	red(20678)	white(21331)	green(17144)
18 B1	BG149860	green(12487)	green(7000)	greenish blue(8202)	green(17804)	white(42299)
7 H11	H75494	red white(32187)	red white(31910)	red white(32300)	white(36179)	yellow(15634)
1 A5	24769*	red(36228)	yellow(22605)	red(33379)	red white(33241)	yellow(15610)
13 D7	AI826395	red white(35424)	red white(35422)	blue(42208)	red white(42151)	green(8718)
14 E11	AA203732	blue(5756)	blue(6317)	green blue(7218)	green(12706)	faint blue(576)
18 B11	R07395	spot missing	spot missing	spot missing	spot missing	blue green(231)
3 H11	T97408	blue(10923)	blue(7277)	blue(8664)	yellow green(16805)	green(8868)
7 A2	R92629	light green(14068)	light green(10546)	green(12328)	green(16246)	blue(6832)
16 E11	BF216768	blue(11669)	blue(8823)	blue(7390)	green blue(83035)	green blue(4440)
9 D7	BI772316	green(15504)	green(16464)	blue green(25117)	green(18576)	blue(7431)

Table 1: Alu-contaminating clones are suppressed in the presence of cold Alu oligo (100pmol/ul and 500pmol/ul) in the hybridization buffer. Compared hybridization at 37°C vs. 42°C. 17 Alu positive clones among all hybridization conditions are compared.

Clone I.D	Image #	NCBI Search for Alu Sequence	Complete or Incomplete Sequence	Clones with Double Bands	Identities
8 F4	248679	N	Incomplete sequence	N	50/55 emboss and 48/51 NCBI
18 B1	344295	Y	complete sequence	N	50/52 emboss and 50/52 ncbi
14 E11	449261	Y AluY,1(=peptive element)	complete sequence	Y	16/24 emboss and 36/40 ncbi
13 D7	376514	N	complete sequence	N	12/17 emboss and 45/52 NCBI
7 H11	230658	Y	Incomplete sequence	N	16/25 emboss and 41/44 NCBI
1 A5	24769	N	Incomplete sequence	N	15/20 emboss and 39/43 NCBI
7 E3	211771	N	Incomplete sequence	N	16/25 emboss and 49/51 NCBI
3 F11	113513	N(contains PTR5)	complete sequence	N	41/54 emboss and 50/55 NCBI
1 E10	20078	N	Incomplete sequence	N	50/55 emboss and 48/52 NCBI
10 F10	293032	N	Incomplete sequence	N	48/55 emboss and 48/55 NCBI
10 H10	293897	N	Incomplete sequence	N	7/7 emboss and 53/55,44/52 NCBI
7 F4	214904	N	Incomplete sequence	N	12/20 emboss and 52/55 NCBI
9 D7	264916	N	Incomplete sequence	N	17/26 emboss and 33/38 NCBI
18 B11	125581	Y	complete (Alu 2X)	Y	40/45 emboss and 51/52,34/36 NCBI
16 D1	503166	Y (also contains L1 repet. elem)	complete	N	17/24 emboss and 35/39 NCBI
7 A2	195925	Y (also contains PTR5 repet. elem)	Incomplete sequence	N	26/43 emboss and 48/55,30/34 NCBI
2 F2	45837	N	Incomplete sequence	N	21/34 emboss and 54/55 ncbi
3 D12	110608	N	Incomplete sequence	N	11/14 emboss and 39/39 ncbi
16 E11	4103438	N	Incomplete sequence	Y	15/20 emboss and 32/34 ncbi
9 D1	262826	N	complete sequence	N	52/56 emboss and 52/56 NCBI
17 F1	362584	N	Incomplete sequence	N	16/26 emboss and 43/51 NCBI
15 E5	488596	Y (MER40 repet.elem)	complete sequence	N	55/55 emboss and 55/55 NCBI

Table 2: Current list of Alu expressed clones.
* not a accession number, accession number was not available in the list

Discussion and Conclusions

We studied 1.7k clone set for possible Alu sequences with various labeling methods, probe purification methods and hybridization conditions. One of the labeling methods uses the terminal-transferase (tdt) enzyme to label the Alu consensus oligo. However, many problems were encountered with this method. Almost every spot provided a strong signal; it was impossible to distinguish clones with Alu sequence from the ones that did not contain the repeats. Terminal transferase (tdt) in theory adds a tail of C's (or whatever nucleotide is being used) at the 3' end of the oligo, hence wherever a stretch of G's are found in the array elements hybridization will occurr. (Fig1)

We designed 3 oligos from our clone set. Each of the 3 oligos (55mer were designed from 3 different complete sequences from 1.7k clone set). These oligos were ordered commercially (Cortec DNA Services). Our purpose for designing the 3 oligos in a similar fashion to the Alu consensus oligo was to verify the conditions for labeling were causing the problems. The exact same conditions were used as for Alu oligo for the labeling and hybridization. Our results indicated clones other than the corresponding spots (complementary clones) were hybridizing (Results not shown)

Our next step of investigation was labeling with chemical kits using PCR product from one of the clones. We evaluated the Label-It kit from Mirus and the Universal Linkage System (ULS) kit from Amersham. Label-It produced a weak hybridization and few spots showed any signal. (Results not shown) ULS in contrast produced a very bright hybridization with the PCR product. (Results not shown). All of the spots showed hybridization however due to the fact that universal primers are used for amplification, and these are present in every probe. Thus we avoided the PCR product approach and continued our investigation with oligo labeling method.

Previous work with microcon columns showed inconsistency between columns. GFX columns were compared to microcon columns for probe purification. GFX columns produced more specific results over microcon columns (Fig 3). Labelling one of the oligo (designed from our clone set) with probe labeling method and purifying with GFX columns produced hybridization to only the corresponding clone. (Fig 2) Other 2 oligos (results not shown) also hybridized to the corresponding spots. These findings confirmed probe clean up with GFX columns is very efficient, but calls into question the reason for the non-specific hybridization when using the Microcon columns. This will require further investigation.

Various hybridization buffers and temperatures were considered to detect Alu sequences. Hybridization at 65°C using SSC/SDS resulted in high background and relatively weak hybridization signal intensities compared to the standard 37°C with Dig Easy hybridization buffer. Another hybridization buffer, Express-Hyb was evaluated. This buffer allows hybridization to occur at higher temperature (60°C). We were able to confirm the specificity of Alu positive clones. (Results not shown) Signal intensities were comparable to Dig Easy Hyb however some background was detected. Based on these results Dig Easy is most suitable and convenient for it's temperature.

For fast /easy detection of Alu sequences, Alu Cy5-5' labeled oligo is used for hybridization. The advantage of having labeled Alu-cy5-5' is that it is ready for hybridization (no need for labeling and probe cleaning up) though bit more expensive. It's fast and consistently reliable. One can simply hybridize any array in question to identify clones contaminated with Alu sequences with this one step assay. We were able to finish this experiment in 2-3 hrs depending on the number of slides to be studied and the resulted images were quite accurate. We hybridized with two different concentrations of the labeled Alu oligo and found that at 100 pmol/ul a total of 39 clones lit up and at 10-pmol/ul 21 clones lit up. (Fig4) All of these clones were investigated for Alu sequences with NCBI blast (blasting 2 sequences) and Emboss (smith water alignment). We found that 22 clones had the full Alu sequence and few others had a partial Alu sequence. Our next question was if inhibitors in the hybridization buffer play a role in terms of Alu sequences lighting up? Hybridization with Alu-cy5-5' were carried out at 37°C in the presence and absence of inhibitors (tRNA and Calf Thymus DNA). These hybridizations showed no significant difference. Hybridization at 37°C and 42°C with Alu-Cy5-5' labeled resulted in 27 and 22 bright spots respectively. We found at higher temperature (42°C) hybridization is more specific. However, at 42°C slightly higher background was detected. (Needs further optimization for cleaner background)

Once clones with Alu sequences were identified, we wanted to see if we could eliminate the effect of these repeats. We wanted to determine if it was possible to block these sequences without affecting hybridization to the real samples. Our first attempt was to study the effect of excess Cot-1 DNA in the hybridization solution. (Fig5) Our findings indicated that in the presence of Cot-1 DNA with labeled cDNA as a probe, hybridization to Alu spots (Alu oligo spotted at different concentrations on 1.7k arrays as a control) was blocked, but signal intensities on 1.7k array overall were relatively low. We then tested the labelled Alu probe (generated from tailing) with Cot-1 DNA in excess in the hybridization solution. (Fig6) Much to our surprise the original Alu oligos spotted at the bottom of the array were hybridized in the presence of Cot-1 DNA the reverse compliment spotted Alu oligos were inhibited and overall expression to the rest of the array was relatively weak. We concluded having excess Cot-1 DNA in the hybridization buffer is not the solution to suppress Alu sequence.Omitting calf thymus DNA in the hybridization solution resulted in a somewhat weaker hybridization indicating calf thymus DNA is essential for efficient hybridization. At this point of the study we stepped back to investigate if high stringency washes has any effect to Alu sequences hybridized with universal cDNA as a probe? (Fig7) Labelled cDNA produced hybridization to both Alu strands spotted on the 1.7k arrays, some inhibition was there, but overall expression was comparable for highly expressed genes.

Our final attempt to inhibit the Alu sequences without interfering with real expression data involved the addition of 10x excess cold Alu oligo during hybridization. The cold Alu oligo apparently suppressed spurious hybridization to Alu sequences and resulted hybridization was as bright as expected to the rest of the array. (Fig8A, 8B, 8C) We compared hybridization at two different temperatures and 2 different concentrations of cold Alu oligo. Table 1 compared intensities and colors of the spots for all hybridizations. Unlabelled Alu oligo at 100 pmol/ul showed a similar pattern at both hybridization temperatures. Cold Alu at 500pmol/ul affected some of the spots on the array while leaving others unchanged. It is suspected that this high an excess of the cold Alu is causing non-specific hybridization and is inappropriately blocking non-Alu containing signals. For this reason we recommend 100 pmol/ul for suppressing hybridization to Alu sequences.

Acknowledgments: We thank Elizabeth Sakac for synthesizing primers. Shani Mintzberg and Quyen Tran for their work on sequences. We couldn't have done it without them.

The UHN Microarray Centre
University Health Network
200 Elizabeth Street
MBRC, 5R-419
1-877-294-4410 (Toll Free)
416-340-4259 (Phone)
416-340-4004 (Fax)
www.microarrays.ca

UHN Microarray Centre