

**LARGE-SCALE GENERATION OF SYNTHETIC DNA LIBRARIES: SEQUENCE-
SPECIFIC PRIMING OF REVERSE TRANSCRIPTION**

by

Yusuf Esmail Murgha

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biomedical Engineering)
in The University of Michigan
2012

Doctoral Committee:

Professor Erdogan Gulari, Chair
Professor Shuichi Takayama
Associate Professor Michael Mayer
Assistant Professor Peter Woolf

© Yusuf Murgha 2012
All Rights Reserved

To my Mom and Dad

ACKNOWLEDGMENTS

I would like to thank Prof. Erdogan Gulari for giving me the opportunity to pursue my career goals and for his advice and support during the doctoral work. I like to thank Dr. Jean-Marie Rouillard for being a great mentor and motivator. His calm, confident and professional demeanor under pressure is a quality I aspire to achieve.

I would like to thank Prof. Shuichi Takayama, Prof. Michael Mayer and Prof. Peter Woolf for their guidance and feedback towards the completion of my work. I thank past and current members of Gulari research group for their help and support. I like to thank all my friends and peers for their continuous motivation and making this an enjoyable journey. I like to thank the administrative staff in biomedical and chemical engineering department.

I dedicate this work to my family, my mom and dad, brother and sister-in-law, and all my relatives. I thank them for their continual support and encouragement. They are my rock.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGMENTS.....	iii
LIST OF FIGURES.....	viii
ABSTRACT.....	xi
CHAPTER 1 INTRODUCTION.....	1
1.1 Applications of oligonucleotide libraries.....	2
1.1.1 Synthetic gene assembly	2
1.1.2 Cloned libraries.....	4
1.1.3 Baits.	7
1.1.4 Oligo-Selective Sequencing (OS-Seq)	9
1.1.5 Fluorescent In Situ Hybridization (FISH)	11
1.2 Source of oligonucleotide libraries	13
1.2.1 Conventional state-of-the art oligonucleotide synthesis.....	13
1.2.2 Microarray Technology	14
1.3 Current limitations	18
1.4 Motivation	20
1.5 Overview of the dissertation	21
CHAPTER 2 FROM DNA MICROARRAYS TO SEQUENCE-SPECIFIC SINGLE- STRANDED NUCLEIC ACIDS.....	23
2.1 Introduction	23
2.1.1 Process outline.....	24

2.1.2	Emulsions background	27
2.1.3	W/O emulsions in biology	33
2.1.4	Methods to prepare single-strand DNA from double-stranded DNA	37
2.2	Experimental methods and protocols	43
2.2.1	Process outline	43
2.2.2	Yeast Knock-out (YKO) deletion collection.....	46
2.2.3	Enzyme based methods to prepare single-strand DNA from duplex DNA..	48
2.2.4	Preparation of single-strand DNA with 5'-modified end.....	49
2.2.5	Optimize digestion time of Lambda exonuclease	49
2.2.6	In vitro transcription and reverse transcription	50
2.2.7	SYBR gold gel staining.....	50
2.2.8	Polyacrylamide gel (PAGE) purification	51
2.2.9	Phenol/ Chloroform/ Isoamyl Alcohol (PCIA) purification.....	51
2.3	Experimental results and discussion.....	52
2.3.1	Yeast knock-out collection (YKO).....	52
2.3.2	Interpretation of gel images	52
2.3.3	Dependence of emulsion PCR on template size and concentration	52
2.3.4	Dependence of emulsion PCR on type of oil.....	55
2.3.5	Enzymatic hydrolysis methods to prepare single-stranded DNA from duplex DNA.....	57
2.3.6	Preparation of single-strand DNA with 5'-modified end.....	61
2.3.7	Optimal digestion time of Lambda exonuclease	64
2.3.8	Amplification of microarray-derived DNA library.....	65
2.3.9	In vitro transcription and reverse transcription (IVT-RT).....	76
2.4	Conclusion	81

CHAPTER 3 PRIMING REVERSE TRANSCRIPTION WITH SEQUENCE-SPECIFIC PRIMER LIBRARIES.....	84
3.1 Introduction	84
3.1.1 Sequence-specific primer libraries for reverse transcription.....	86
3.1.2 Background	88
3.1.3 Insights from reverse transcription-polymerase chain reaction (RT-PCR)	90
3.2 Experimental methods and protocols.....	93
3.2.1 One round aRNA amplification with 5'-end dangling gene-specific primer.....	93
3.2.2 Removal of excess sequence-specific primers.....	93
3.2.3 Optimize RT conditions when using sequence-specific primers.....	95
3.2.4 Detect differential expression of transcripts with sequence-specific primers.....	96
3.2.5 Activity of reverse transcriptase in Actinomycin D.....	97
3.2.6 Two step process to remove excess long primers	99
3.2.7 Reverse transcription with sequence-specific long primer library.....	100
3.2.8 <i>Saccharomyces cerevisiae</i> growth conditions and total RNA extraction.....	101
3.2.9 Polymerase chain reaction (PCR)	101
3.2.10 Real-time PCR and Melt Curve	102
3.2.11 Microarray analysis.....	102
3.3 Experimental results and discussion.....	103
3.3.1 Interpretation of real-time PCR plots.....	103
3.3.2 One round aRNA amplification with long gene-specific primers	103
3.3.3 Removal of excess sequence-specific primers	106

3.3.4	Optimize reverse transcription conditions.....	114
3.3.5	Detect differential expression of transcripts with sequence-specific primers.....	117
3.3.6	Activity of reverse transcriptase in Actinomycin D.....	122
3.3.7	Differential expression of yeast genes induced in glucose and galactose media	124
3.3.8	Two step process to remove excess long primers	131
3.3.9	Reverse transcription with sequence-specific long primer library	134
3.4	Conclusion	139
CHAPTER 4 CONCLUSION AND RECOMMENDATIONS.....		141
4.1	Conclusion	141
4.2	Recommendations for future work.....	143
4.2.1	Under generation of single-strand nucleic acid libraries	143
4.2.2	Under sequence-specific primer libraries.....	146
APPENDIX		148
BIBLIOGRAPHY		162

LIST OF FIGURES

Figure 1-1 Synthetic gene assembly.....	3
Figure 1-2 Cloned libraries – RNAi and peptide libraries.....	5
Figure 1-3 Solution hybrid selection – Baits.....	8
Figure 1-4 Padlock probes – MIPS and SGC.....	8
Figure 1-5 Oligonucleotide selective sequencing.....	10
Figure 1-6 Fluorescent In Situ Hybridization (FISH).....	12
Figure 1-7 Oligonucleotide synthesis	13
Figure 1-8 Microarray synthesized oligonucleotide library	18
Figure 2-1 Process overview.....	25
Figure 2-2 Emulsion types	29
Figure 2-3 Emulsion breakdown processes.....	31
Figure 2-4 Emulsion PCR versus Conventional PCR.....	35
Figure 2-5 Exonucleolytic hydrolysis of duplex DNA to generate ssDNA	39
Figure 2-6 Formation of single-stranded DNA.....	41
Figure 2-7 Emulsion PCR.....	44
Figure 2-8 Restriction digest of yeast genomic DNA	53
Figure 2-9 Effect of template size on emulsion PCR	54
Figure 2-10 Effect of oil type on emulsion PCR.....	56
Figure 2-11 T7 gene 6 exonuclease and phosphorothioate backbone	59
Figure 2-12 Lambda exonuclease hydrolysis of 5' phosphoryl end	60

Figure 2-13 Lambda exonuclease activity on modified 5'-ends.....	62
Figure 2-14 Optimal digestion time (lambda exonuclease).....	64
Figure 2-15 Measurement of droplet size.....	66
Figure 2-16 Real-time PCR – Standard curve.....	69
Figure 2-17 Emulsion PCR versus Conventional PCR.....	71
Figure 2-18 Lambda exonuclease - Oligonucleotide library to ssDNA.....	74
Figure 2-19 Oligonucleotide library – IVT-RT method.....	78
Figure 3-1 cDNA synthesis with sequence-specific primers 1.....	104
Figure 3-2 cDNA synthesis with sequence-specific primers 2.....	105
Figure 3-3 Spin column removal of double-stranded primer.....	107
Figure 3-4 Effect of SSB protein on amount of ssDNA digested.....	109
Figure 3-5 Real-time PCR - Effect of SSB protein on amount of ssDNA digested.....	110
Figure 3-6 Size dependent fractionation - PEG concentration (1.25 M NaCl).....	112
Figure 3-7 Size dependent fractionation PEG and NaCl concentration.....	113
Figure 3-8 Effect of temperature on reverse transcription.....	115
Figure 3-9 Effect of primer concentration on reverse transcription.....	116
Figure 3-10 Density Plot – cDNA synthesis with mouse sequence-specific primer.....	118
Figure 3-11 Differential gene expression – mouse sequence-specific primers.....	118
Figure 3-12 Exonuclease I digestion and reverse transcription of ssDNA.....	120
Figure 3-13 'no RNA' reverse transcription with sequence-specific primers.....	121
Figure 3-14 Activity of reverse transcriptase in presence of Actinomycin D.....	123
Figure 3-15 Melt curve of GAL1 gene - optimal sample processing method.....	125
Figure 3-16 Number of gene detected across sample processing methods.....	126
Figure 3-17 List of yeast genes expressed in galactose - glucose.....	128
Figure 3-18 Yeast gene expression – Effect of actinomycin D.....	130

Figure 3-19 Exonuclease I digestion in presence of actinomycin D.....	132
Figure 3-20 AMPure bead - Effect of DNA concentration	133
Figure 3-21 AMPure bead - Effect of PEG8000 concentration.....	133
Figure 3-22 Gene expression – exonuclease I and 11 % PEG 8000	135
Figure 3-23 Gene expression – exonuclease I and 10 % PEG 8000	136

ABSTRACT

Custom-designed DNA and RNA oligonucleotide collections are used as building blocks in synthetic biology, complementary probes for targeted sequencing, and nucleic acid aptamers. They encode information for technologies like RNA interference, protein engineering and DNA-encoded chemical libraries. These applications require an economical source of diverse libraries.

High-throughput microarray technology produces hundreds of thousands of diverse sequences on a single planar substrate at low cost (<\$200 for 20,000 oligonucleotides). However, the quantity and quality of individual oligonucleotides obtained from this technology is low (~ 5 fmol). In order to produce feasible amounts of high quality oligonucleotides (10 – 100 pmol), it is necessary and more economical to go through a molecular amplification procedure. Existing methods of amplification lead to the formation of chimeric products. We developed a 2-step amplification process which gives an overall ~450 fold increase in the amount of single-stranded oligonucleotides. The first-step, emulsion polymerase chain reaction (PCR), provides the initial amplification while preserving the library complexity (~25 μ M product). However, the overall product yield from a single reaction after removal of primer binding sequences and single-strand DNA formation is limited (~15 pmol) as a result of restrictions imposed on the input reactant quantity. Therefore, this step alone is labor intensive to produce large amounts of products due to lack of scalability. This limitation is addressed via transcription-reverse

transcription of the emulsion PCR amplicons. This two stage process allows scalability of microarray synthesized oligonucleotides for preparation of large quantities of ssDNA libraries (~6500 pmol).

As an application of the amplification technology we developed an integrated method for cDNA library construction with sequence-specific primers incorporating a unique tag and universal primer sequence. The method suffers from the formation of three types of false positives that need to be sufficiently removed to reduce contribution of false-positives signals. A 3 step process is implemented to reduce the false positives contributors and still detect differential expression of yeast genes in galactose and glucose conditions. The sequence-specific primers libraries can be used for applications not limited to detection of low abundance and rare RNA and identification of aberrant splicing variants and gene-fusions.

CHAPTER 1

INTRODUCTION

The last decade has seen the emergence of a broad range of applications for DNA and RNA oligonucleotide libraries. DNA oligonucleotides are the synthetic biology building blocks for the assembly of single genes [1-3] to whole genomes [4, 5]. Targeted next-generation sequencing relies heavily on oligonucleotide libraries as a source of baits to capture, either in the form of DNA padlock probes for the circularization of targeted sequences [2, 6] or in the form of RNA baits for the direct capture of sequencing genomic DNA library fragments [7]. Recently Millykangas et al. pushed the application of oligonucleotide libraries for targeted sequencing even further by integrating the target capture into the sequencing device, using a DNA oligonucleotide library to customize the primer lawn on a sequencing flowcell [8]. Oligonucleotides libraries are widely used to encode active RNA such as shRNA [9], or peptides [10, 11] after cloning in appropriate vectors. Labeled libraries can be also used for molecular detection probes in fluorescent in situ hybridization (FISH) techniques such as OligoPaint [12].

1.1 Applications of oligonucleotide libraries

1.1.1 Synthetic gene assembly

Gene-synthesis is the process of creating a gene in vitro without a precursor template [Figure 1-1]. The process of gene-synthesis begins with hybridization of melting temperature-normalized oligonucleotides to each other to make the gene backbone. The oligonucleotides overlap at their ends to simultaneously act as templates and primers for the synthesis of other strand. These are subsequently joined by various PCR techniques (ligation PCR, assembly PCR and fusion PCR) to form a new gene. Here, the primary requirement is for long oligonucleotides so that fewer numbers are required to make longer gene segments or genes. Also, a large amount of oligonucleotides are required to make many genes and its variants in a single reaction.

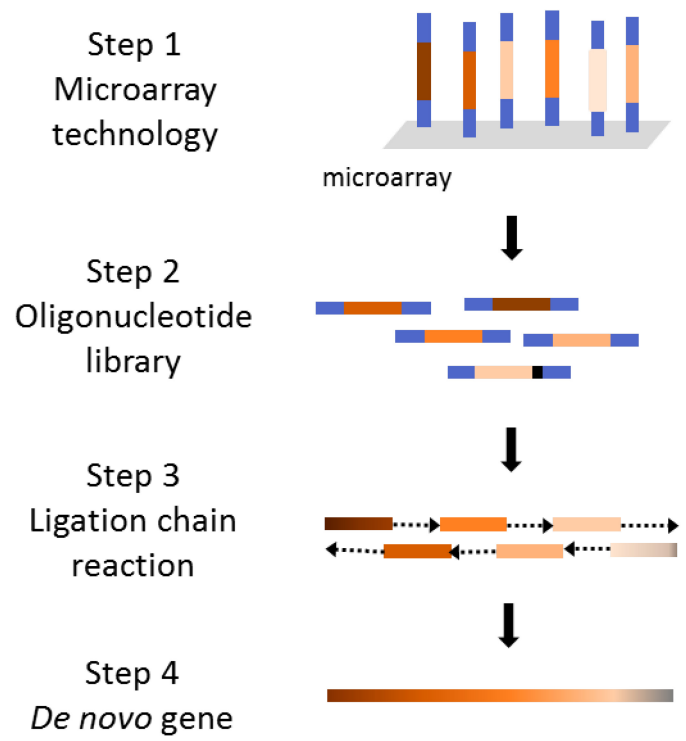


Figure 1-1 Synthetic gene assembly

In the field of synthetic biology, oligonucleotide libraries have accelerated the optimization of gene-expression and protein activity [13], enabled the assembly of difficult to make antibodies and antigens [14, 15], engineering of metabolic pathways [16] and construction of mitochondrial, viral and bacterial genomes [4, 17, 18]. Compared to conventional methods of gene-generation, which pick elements from existing elements by PCR and DNA cloning, they provide greater flexibility to gene-engineering to ultimately create organisms with new properties (bioremediation, biosafety, biofuels, drugs and vaccines). Another variant of assembly of synthetic oligonucleotides is in the field of DNA computing and storage [19, 20].

1.1.2 Cloned libraries

Here, the oligonucleotide libraries are used to encode information for downstream applications. This can be either for a short hairpin RNA (shRNA) or a gene coding for a peptide [Figure 1-2]. Small interfering RNAs (siRNAs) are used for specific silencing of many genes. The process has been used to study cell behavior in many mammalian cells [21]. For this application the oligonucleotide libraries has self-complementary ends separated by common loop, which then folds to form short-hairpin DNAs (shDNA) constructs. The individual shDNA are cloned into vectors to produce shRNAs (or siRNA) that get incorporated into RNA-induced silencing complex (RISC) to guide mRNA cleavage and degradation [22, 23]. Unlike other methods to make siRNA, namely i) chemical in situ synthesis of RNA, which is cost prohibitive, ii) biological sources that give variable amounts of siRNA depending on mRNA expression levels and abundance in the tissue and iii) enzymatic digestion (RNase III, Dicer) of long dsRNA molecules into randomly chopped molecules, complex oligonucleotide libraries give more defined and

uniform representation of siRNA without any of the non-specific effects such as binding to genes with homologous sequences or to splice isoforms [Figure 1-2] [24, 25].

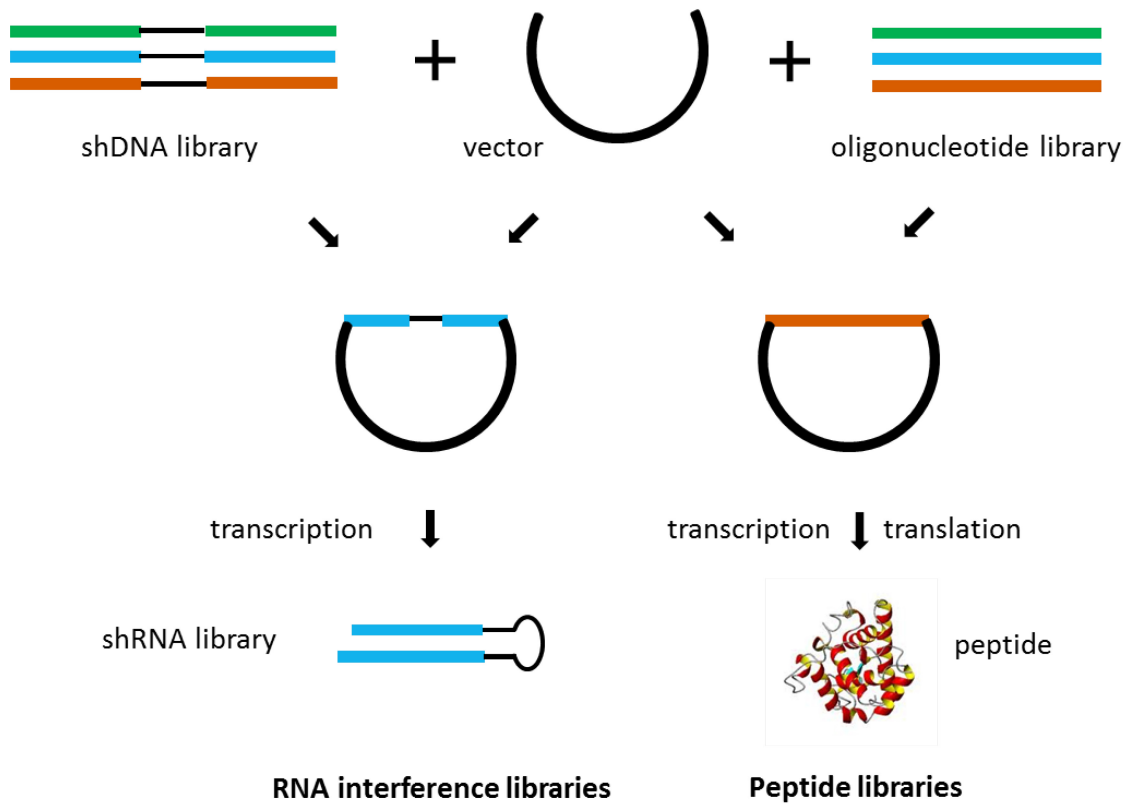


Figure 1-2 Cloned libraries – RNAi and peptide libraries

Phage display is a high-throughput technique to screen, select and optimize for protein interactions with other molecules such as ligands, peptides, other proteins and DNA [26]. Here the phenotype (individual peptide) is linked to its genotype (coding sequence) which enables the isolation of binding clones from large populations. Historically, random nucleic acid libraries used to encode for peptides or proteins expressed on the bacteriophage cell surface is a great method to identify good candidates. However, they are not able to cover the entire diversity of peptide library (e.g. for a 12 mer peptide, there are possible $12^{20} = 4 \times 10^{15}$ individual sequences of which less than 0.001% are screened in a typical experiment) [27]. Also, there is a decrease in the diversity of the library during the panning and selection rounds, further reducing the number of binding clones identified during the screening process [28]. Thus once good candidate sequences are found, they can be mutated in a controlled manner by using oligonucleotide libraries to access the best candidates with the highest binding affinities.

Another application of oligonucleotide libraries is to produce peptide or short protein libraries [Figure 1-2]. Here systematic combinations of DNA fragments are cloned into vectors to express peptides and short proteins to identify biologically active peptides (e.g. anti-microbial), discover ligand-binding proteins, study and optimize enzyme functionalities, epitope mapping and generate vaccines [29]. Although, solid-phase peptide synthesis competes with the flexibility provided by DNA libraries, it is limited by the peptide length (<18 amino acids), and high costs of synthesizing large libraries [30]. For both peptide libraries and phage display, oligonucleotide libraries provide flexibility in choosing codons to optimize for protein expression.

1.1.3 Baits

The past few years has witnessed rapid development in 'next-generation' or massively parallel sequencing technology leading to sequencing whole genomes [31-33]. Whole genome sequencing of humans is resource-intensive as multiple runs are required to get sufficient depth and coverage. In addition, most disease-causing mutations (85%) are estimated to be located within the exome (protein coding regions), which constitute only 1 % - 1.5 % of the human genome (3 Gb) [34, 35]. Hence, methods are required to selectively enrich and amplify gene-subsets for targeted sequencing. The underlying principle of the methods namely, molecular inversion probes (MIP) [36], selective genomic circularization (SGC) [37] and solution hybrid selection (SHS) [38] is solution-based hybridization of targets with complementary sequences. For these applications a large number and quantities of diverse oligonucleotides are required to increase genome coverage and sensitivity and reproducibility of the assays. In case of MIPs, the use of longer oligonucleotides enables the incorporation of barcodes of sample multiplexing [39].

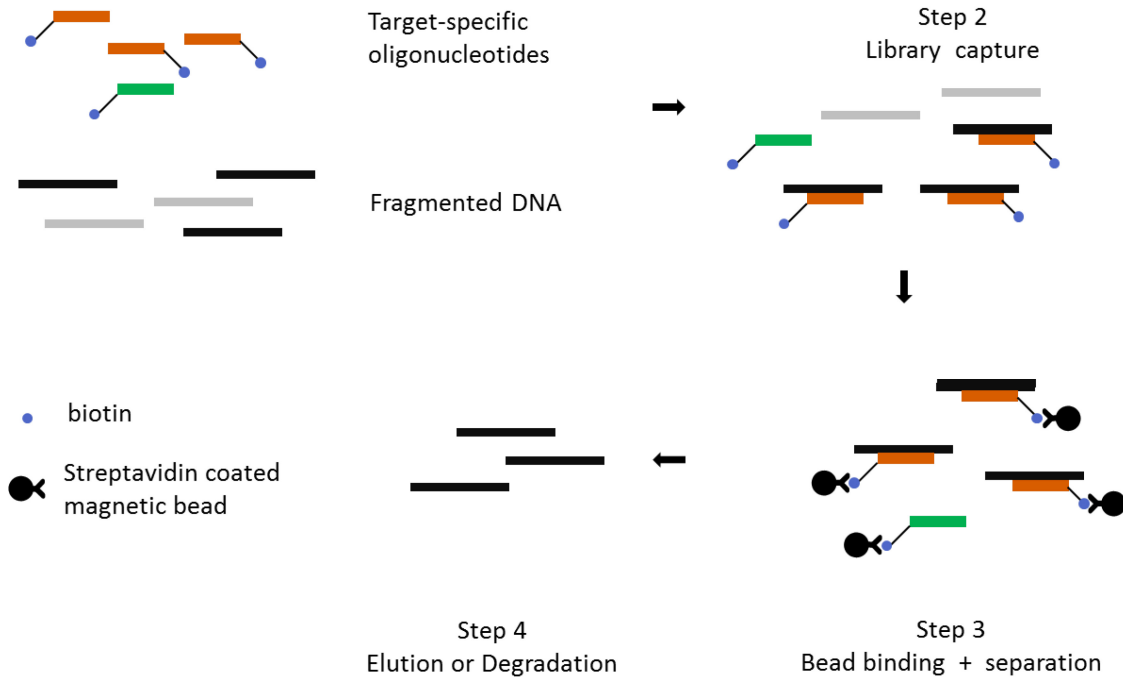


Figure 1-3 Solution hybrid selection - Baits

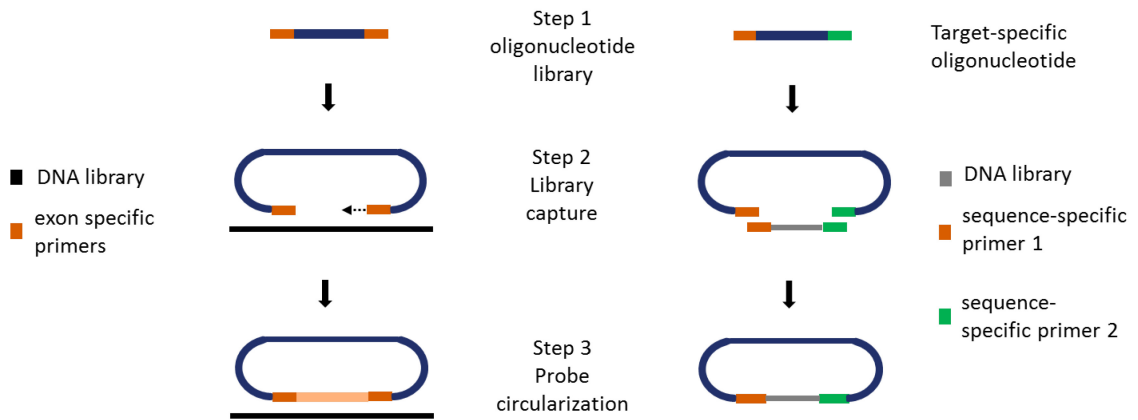


Figure 1-4 Padlock probes - MIPS and SGC

In SHS method, biotinylated target-specific RNA probes (or baits) bind to fragmented genomic DNA libraries and separated by capture on magnetic beads. As single-stranded RNA display a different spectrum of sensitivity to chemical agents and enzymes they can be easily removed from DNA prior to sequencing [Figure 1-3]. On the hand, DNA baits (MIPs or padlock probes - SGC) are used for copying target sequences between two primer binding sites [Figure 1-4]. The DNA probes have common linker flanked by two target-specific sequences. In MIPs method the baits hybridize to both end of the intended target, followed by primer extension (DNA polymerase) and ligation (DNA ligase) to form circular probes. Linear species molecules are removed by exonucleases to enrich for the circularized probes. In SGC method, the biotinylated probes anneal to either end of the target DNA fragments and are ligated to form circles. Unlike, MIPs method where the circularized probe is a complement of the DNA, in SGC method the DNA is directly incorporated into the circular probe [40].

1.1.4 Oligo-Selective Sequencing (OS-Seq)

Recently, a new approach to targeted sequencing of single nucleotide variants (SNV) in cancer has been published that uses oligonucleotides libraries as a tool to re-engineer the Illumina GaxII flow cell [8].

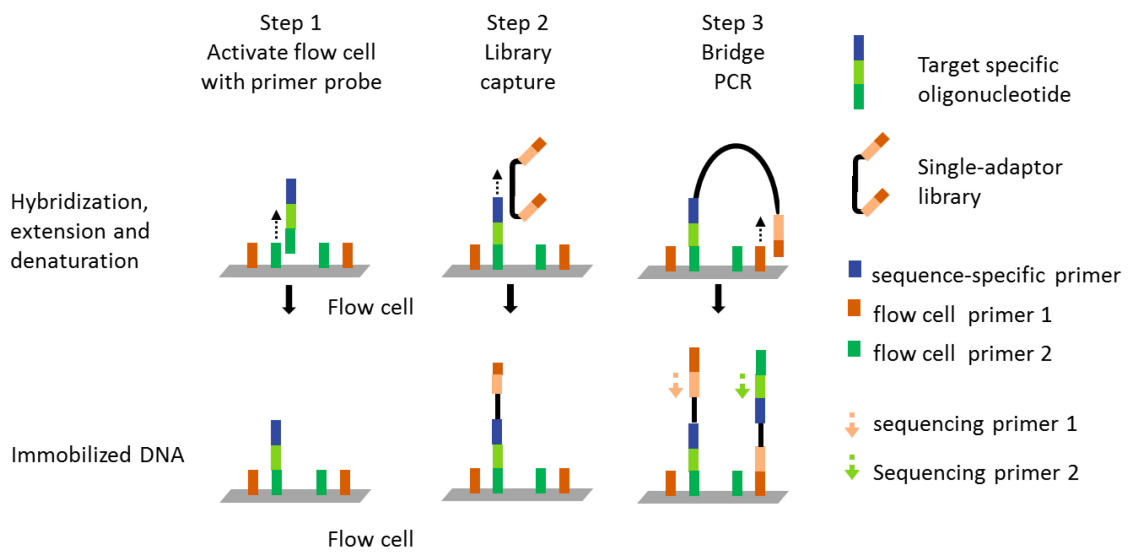


Figure 1-5 Oligonucleotide selective sequencing

The surface functionality of flow cell is changed using oligonucleotides that serve as probes to capture specific DNA library fragment and as sequencing primers [Figure 1-5]. The probe binds to one of the two immobilized short PCR primers on the flow cell and is extended by DNA polymerase to build a highly customized flow cell (Step 1). In step 2, the probes hybridize to their target fragments. The sheared DNA library is prepped such that either end is flanked by primer sequences to flow cell primer. The library is amplified on the flow cell by bridge PCR. This results in both strands being immobilized on the flow cell for bio-directional sequencing (Step 3).

1.1.5 Fluorescent In Situ Hybridization (FISH)

Fluorescent in situ hybridization (FISH) is a method to visualize and map an individual's genetic material. The method uses fluorescent probes to tag chromosomal regions within cells to detect chromosomal rearrangements, translocations, deletions and copy number variations. It is used for clinical diagnosis of many diseases such as Down syndrome (trisomy 21) and Angelman syndrome (deletion of genes on maternal chromosome 15). This techniques has been used to detect localized mRNA in bacteria [41] and yeast [42]. Currently, dye-labeled oligonucleotide libraries are being investigated to increase the resolution and dynamic range of existing FISH technologies [12]. This application of FISH probes as Oligopaints in particular requires microgram amounts of diverse oligonucleotides for significant coverage of chromosomal regions [Figure 1-6].

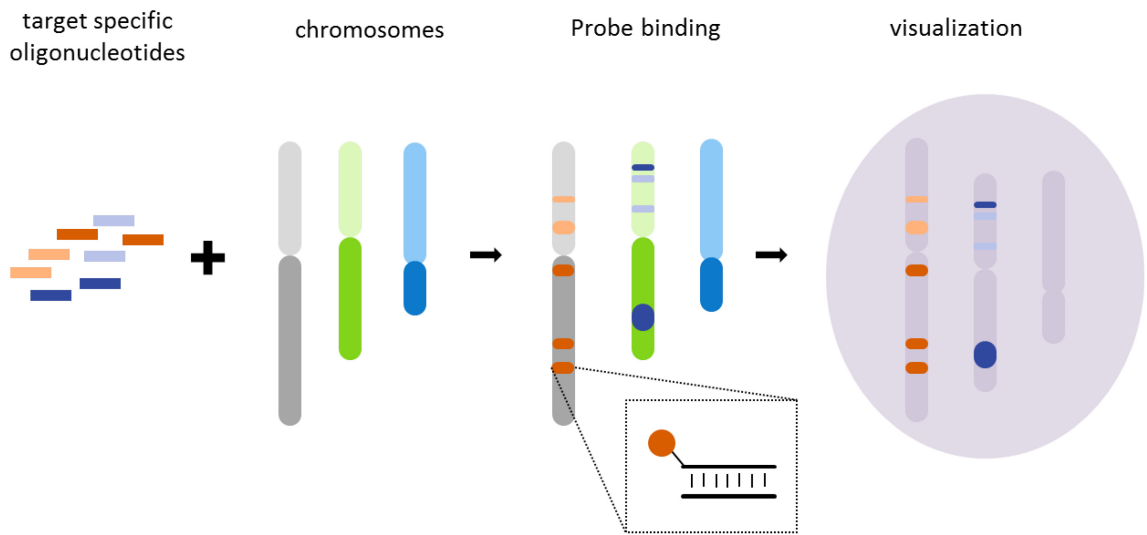


Figure 1-6 Fluorescent In Situ Hybridization (FISH)

1.2 Source of oligonucleotide libraries

1.2.1 Conventional state-of-the-art oligonucleotide synthesis

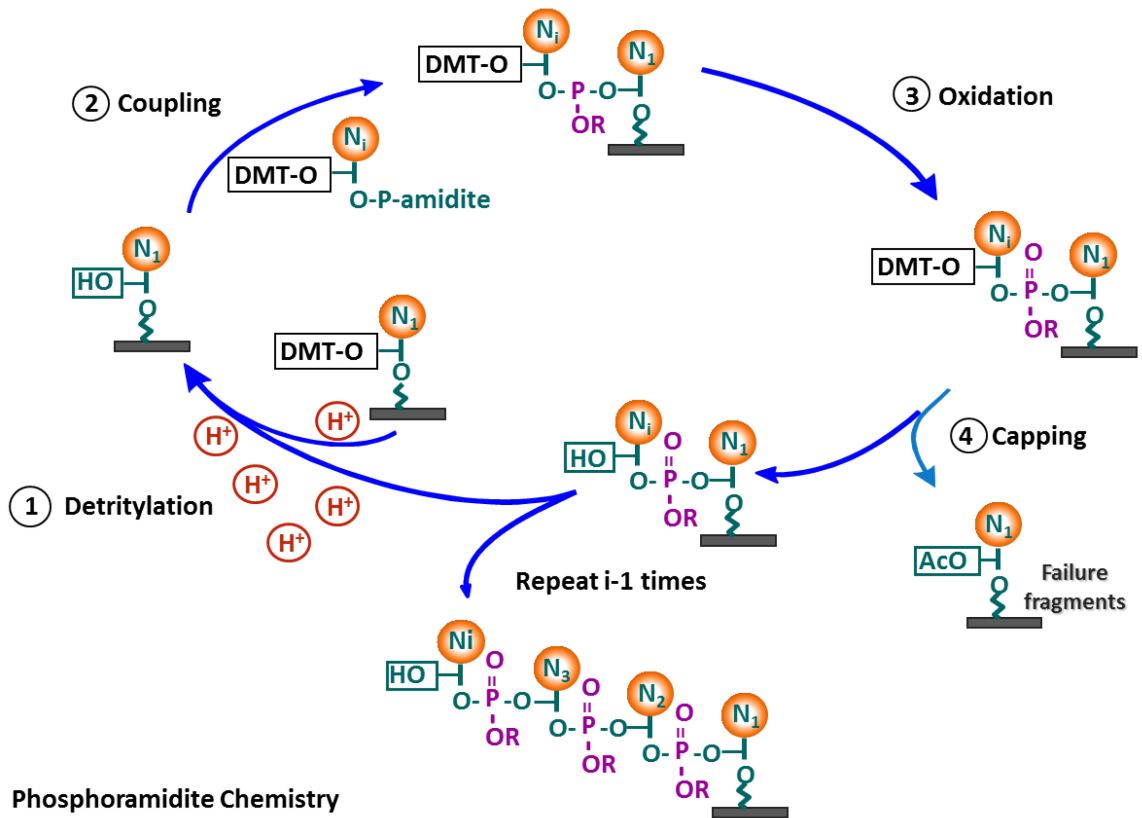


Figure 1-7 Oligonucleotide synthesis

Oligonucleotides are synthesized using standard phosphoramidite chemistry starting from 4 tiny bottles of DNA monomers namely adenine, guanine, cytosine and thymidine [Figure 1-7]. Acid-cleavable phosphoramidite monomers are added in cyclical process to grow the nucleotide polymer chain from 3'-end to the 5'-end. In the first synthesis cycle, the desired monomer is linked to a solid support (e.g. controlled pore glass) pre-functionalized with a 3' end tethered monomer. This is followed by step-wise addition of deoxynucleoside monomer in 4 step-process, which are i) deprotection (detritylation) – weak acid such as trichloroacetic acid (TCA) is used to remove the dimethoxytrityl (DMT) group from the 5' end of monomer, ii) coupling – 5'-hydroxyl end of monomer is covalently linked to the desired monomer in the presence of a activator agent (e.g. 1-H tetrazole), iii) capping – unreacted 5'-hydroxyl monomers are capped from further reactions by acetyl group to minimize deletion sequences and vi) oxidation – the two monomers on the oligonucleotide chain linked by unstable phosphite triester is converted to phosphate (e.g. iodine, pyridine solution). This process is repeated several times to grow DNA polymer, which is then cleaved from the surface using a base-catalyzed reaction (ammonium hydroxide), which also removes the remaining protecting groups [43].

This process is the principle method used by most commercial DNA synthesis companies such as Integrated DNA technologies. Individual oligonucleotides (up to 200 bases) are synthesized in large reactors (column-based synthesis) resulting in yields from 10 - 100 nmol (10^{-9}) at low costs (\$0.1 per base) [44].

1.2.2 Microarray Technology

DNA microarrays are an abundant source of inexpensive complex DNA libraries. Currently, commercially available, *in situ* synthesized oligonucleotide

microarray come under two categories i) photo-labile 5'-O protecting group (Affymetrix, Nimblegen) and ii) acid-labile 5'-O protecting group (Agilent, CustomArray, LC Sciences, Mycroarray). The protection group serves the same function as DMT group used for column synthesis, i.e. prevents monomers to form homopolymers.

Photo-labile protecting group

Affymetrix is the dominant player in the field of microarrays. They leverage semiconductor-based photolithographic fabrication techniques and use photo-labile groups, which enables them to synthesize a large variety of defined probes (1 million) on a large surface with feature size ranging from 1 – 5 μm . Their arrays are made using photolithographic masks, which require up to 100 masks to make a 25-mer oligonucleotide. The high cost of making pre-fabricated masks significantly adds up when used to make custom oligonucleotide libraries required for the applications above [45, 46].

Nimblegen like Affymetix uses photo-labile protection groups but differs in two aspects, i) they use mask-less projection technology for deprotection – Digital Micromirror Device (DMD) and ii) they use a different protecting group. The DMD consists of electromechanically controlled mirrors that give them the flexibility to target any area on the chip. Also, their protecting group (5'-ortho-nitrophenyl-type [NPPOC]) enables them to achieve higher step-wise efficiency (~96%) than Affymetrix (ortho-nitro-benzyl-type [MeNPOC]; <94%) and they have been able to make approximately million probes of 80 bases length [47, 48]. As both methods are dependent on photolytic cleavage of protecting groups, and as extent of cleavage is dependent on the exposure time (order of 1-2 min to get 99% deprotection), they

are unable to make long oligonucleotide libraries (100 – 150 bases). For 100-mer oligonucleotide, 96% step-wise yield gives $0.96^{100} = \sim 1.69\%$ full length sequences).

Acid-labile protecting group

Agilent microarrays are based on inkjet printer technology that enables the synthesis of high density arrays (244k spots) with up to 150-mer long oligonucleotides [49]. An inkjet printer head controlled by motion controllers delivers the synthesis reagents to desired spots on the array. Like, column-based synthesis, they use standard phosphoramidite chemistry which allows them to achieve close to 99.5 % step-wise efficiency [50].

CustomArray uses electrochemical reactions for DNA synthesis on silicon substrate. The acid produced then catalyzes the removal of the protecting group to enable coupling with incoming monomer. Their arrays are fabricated using CMOS technology to produce a matrix of individually addressable microelectrodes. The application of current, leads to oxidation of electrolyte at the anode to generate acid that diffuses to the spot for oligonucleotide synthesis and in a defined region, which is then neutralized at adjacent cathodes. They can make up to 12k spots on the array with 50 base oligonucleotides [51].

LC sciences and Mycoarray have adapted the standard phosphoramidite chemistry to digital photolithography for synthesis of DNA microarrays. This is achieved by replacing the acid with a precursor molecule which generates acid upon UV illumination for synthesis at desired spots. Both technologies achieve > 99% step-wise yield and can synthesize oligonucleotides up to 120 – 150-mer [52, 53]. However, because the acid generated is in the liquid phase, microfluidic devices with individual microchambers are fabricated from silicon to prevent diffusion. This limits the density of features to $\sim 3,698$ for PicoArray chip marketed by LC Sciences.

On the other hand Mycoarray have further optimized the technology to perform synthesis on open planar substrates. This has enabled them to achieve 20 k spot density.

1.3 Current limitations

The common theme to all the above applications is the requirement of complex oligonucleotide libraries. Commercial DNA synthesizers make individual oligonucleotides in large reactors (column-based synthesis) and have been scaled up to produce 96 or 384 oligonucleotides in a single batch. They give really high yields (nanomoles) of individual oligonucleotides at relatively low cost (\$0.1 per base). Nevertheless, they are still too expensive to make libraries for applications such routine assembly of denovo genetic circuits or organisms. For instance, Craig Venter Institute used this option to synthesize bacterial genomes of *M. genitalium* (580 kb) and *M. mycoides* (1.08 Mb) at costs of several hundred thousand dollars ($1.08\text{Mb} \times \$0.1/\text{base} \times 2 \text{ strands} = \$216,000$) [4, 5].

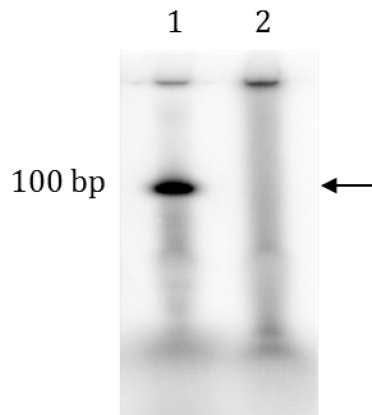


Figure 1-8 Microarray synthesized oligonucleotide library

1. Control oligonucleotide : 100 bp is run as marker (1 pmol)

2. library: 1 pmol of 100 bp array library

Diverse oligonucleotide synthesis prices can be reduced by using high throughput microarray technology. They produce hundreds of thousands of oligonucleotide however their Achilles heel is the limited spot size where the synthesis occurs, which is usually well below 100 microns in diameters [2, 48, 51]. This results in very small synthesis scale. The overall yield of individual oligonucleotides per chip is in the range of 10 - 15 fmol, which is 10^6 times less than column-synthesis. This amount of oligonucleotides is further reduced as chemical reactions are not 100 % efficient; there are errors most often deletions. This leads to the formation of truncated DNA molecules as depicted in Figure 1-8. Thus, a 100 base oligonucleotide synthesized using state-of-art column synthesis (>99.5% step-wise efficiency) gives about $0.995^{100} = 60.5$ percent full length oligonucleotides. For array based synthesis the yield is about 36.6 percent (99% step-wise efficiency) resulting in about ~3.6 - 5.5 fmol of individual oligonucleotide. This dictates a molecular amplification step; polymerase chain reaction (PCR) being the optimal choice to high amplification (> 10^6 -fold) amount. Conventional PCR has two major limitations related to amplification of multiple templates:

- Preferential amplification of some templates
- Formation of chimeric products.
- Truncated products within the oligonucleotide library can act as primers to make non-specific products.

Also, PCR leads to formation of double-stranded DNA (dsDNA) and incorporates primer binding sequences (PBS). The addition of PBS increases the length of oligonucleotide synthesized, raising the percentage of undesired

sequences. After PCR amplification, the PBS is removed by type II restriction enzymes. For applications like MIPs or gene-assembly, the enzyme used must cut outside their recognition site [36]. The former i.e. formation of dsDNA requires additional processing steps to make them single-stranded DNA (ssDNA). Most methods to prepare single-stranded DNA (ssDNA) (e.g. exonuclease digestion) protect the desired strand from exonuclease digestion. For these methods:

- The amount of single-strand DNA produced depends on the starting amount of PCR products

Finally, the single-strand DNA reaction products consists of 5 different DNA species; desired ssDNA, dsDNA that is not digested by the restriction enzyme and that is not converted to single-stranded form, as well as any leftover PCR primers and cut primer binding DNA fragments. High quality ssDNA is obtained by purification on polyacrylamide gels.

1.4 Motivation

Many current (targeted sequencing, synthetic biology) and upcoming applications (FISH, RNA interference and bioactive peptides) use oligonucleotide libraries pools as a starting point. These applications require anywhere from 50 ng (1.6 pmol) – 1 µg (30 pmol) of single-strand nucleic acid (size 100bp) per reaction [39]. Technologies like targeted-sequencing, RNAi and fluorescent in situ hybridization (FISH) are being developed for clinical applications. Just during the development stages they require 10 - 100 reactions (even more for pilot studies) for robust data analysis. The current method to amplify microarray derived libraries i.e. conventional PCR falls short as it amplifies non-specific products. Also,

exonucleolytic hydrolysis of unprotected strand combined with polyacrylamide purification to remove PCR artifacts and primer binding sequences leads to > 50 % loss of single-strand DNA. Typical yields obtained per reaction ranges from 5 – 15 pmol.

This dissertation aims to develop a methodology to make large quantities of diverse oligonucleotides by integrating solid-phase oligonucleotide synthesis technology with molecular biology methods. In the second part, it develops the method to use oligonucleotide libraries as primers for first strand cDNA synthesis. Finally, the oligonucleotide libraries will help accelerate progress in the fields of synthetic biology, protein engineering, biotechnology and clinical diagnostics and therapeutics.

1.5 Overview of the dissertation

Chapter 2 presents an overall process outline starting from getting diverse oligonucleotide pools to generation of single-stranded DNA (ssDNA) and describes in detail two key components, i) molecular amplification method and ii) methods to make ssDNA. First emulsion PCR amplification technique is adapted to increase the yield of chip-derived oligonucleotides. Here, we investigated parameters to get amplification with the objective of preserving sequence fidelity and library. Second, many molecular biology methods are tested to obtain large amount of single-stranded DNA. Two of these methods are compared on the basis on yield and depending on the method different strategies are used to remove the PCR primer sequences.

Chapter 3 describes in a step-wise manner the problems encountered when using a library of sequence-specific primers to make complementary DNA (cDNA)

from RNA templates. In this case many enzymatic and non-enzymatic methods are tested in varying combinations to effectively reduce any false positives. The success of the methods is determined by measuring the level of genes induced in the yeast galactose versus glucose growth media. Gene-expression is studied initially using microarrays and later by polymerase chain reaction.

CHAPTER 2

FROM DNA MICROARRAYS TO SEQUENCE-SPECIFIC SINGLE-STRANDED NUCLEIC ACIDS

2.1 Introduction

DNA microarrays combine the benefits of solid phase synthesis and combinatorial chemistry [54]. They have ten to hundreds of thousands of polynucleotides fabricated with high resolution at precise locations. This makes them making them an inexpensive source of diverse DNA libraries [55]. Traditionally, microarrays have been used for hybridization based studies to i) measure changes in gene expression levels [56], ii) identify genomic gains or loss (microarray comparative genomic hybridization - CGH) [57], and iii) identify DNA mutations (e.g. single nucleotide polymorphisms – SNP) [58, 59]. Applications of microarrays have been described for almost every field of biological science.

Recently, there have been reports that use microarray synthesized DNA (oligonucleotide) libraries to assemble short DNA fragments into synthetic genes [2, 60-62]. Complex oligonucleotide libraries have been used in a variety of ‘front-end’ capture methods to selectively enrich and amplify regions of the genome for targeted sequencing [36, 37, 63, 64] as well as generation of mutant DNA libraries for peptide evolution [10]. For all these applications the bound oligonucleotides has to be released from the surface. This is done by either breaking the covalent bond between the first nucleoside and array surface or an on-chip second strand DNA

synthesis followed by DNA melting to release single-strand DNA. The success of latter method (i.e. on-chip PCR) depends on many factors such as steric hindrance, surface passivity and evaporation and is reviewed here [65].

Using the former method, the oligonucleotides are released from the surface in one big pool. Two major drawbacks limit their direct use namely, i) low yield of synthesized oligonucleotides released per chip and ii) less than 100% step-wise monomer coupling efficiency. A step-wise efficiency of 99%, will give only $0.99^{25} = 77.7\%$ full length DNA compared to 36.6 % for a 100-mer. Also at each synthesis reaction cycle there may be a shift in the acid diffusion boundary may by \pm a few nanometers. This leads to formation of undesired oligonucleotide products whose size and sequence are highly variable.

The methods to select and purify correct size oligonucleotides such as polyacrylamide gel (PAGE) purification and high performance liquid chromatography (HPLC) gives less than 50% recovery yields further reducing the amount of oligonucleotides. To ensure enough probes are available for downstream applications, the template library is amplified by PCR, which necessitates extra steps to get single-stranded DNA (ssDNA). Here, both the amplification and ssDNA formation method is important to get high quality nucleic acid probes.

This chapter details the issues encountered and process to make large quantities of oligonucleotide libraries starting from microarray-derived DNA pool

2.1.1 Process outline

An overall outline of the process to get functional ssDNA from oligonucleotide probes is shown in Figure 2-1. Two methods integral to the overall process, i) Emulsion PCR and ii) conversion of double-strand DNA (dsDNA) to single-strand DNA (ssDNA) are described in more detail in separate sections.

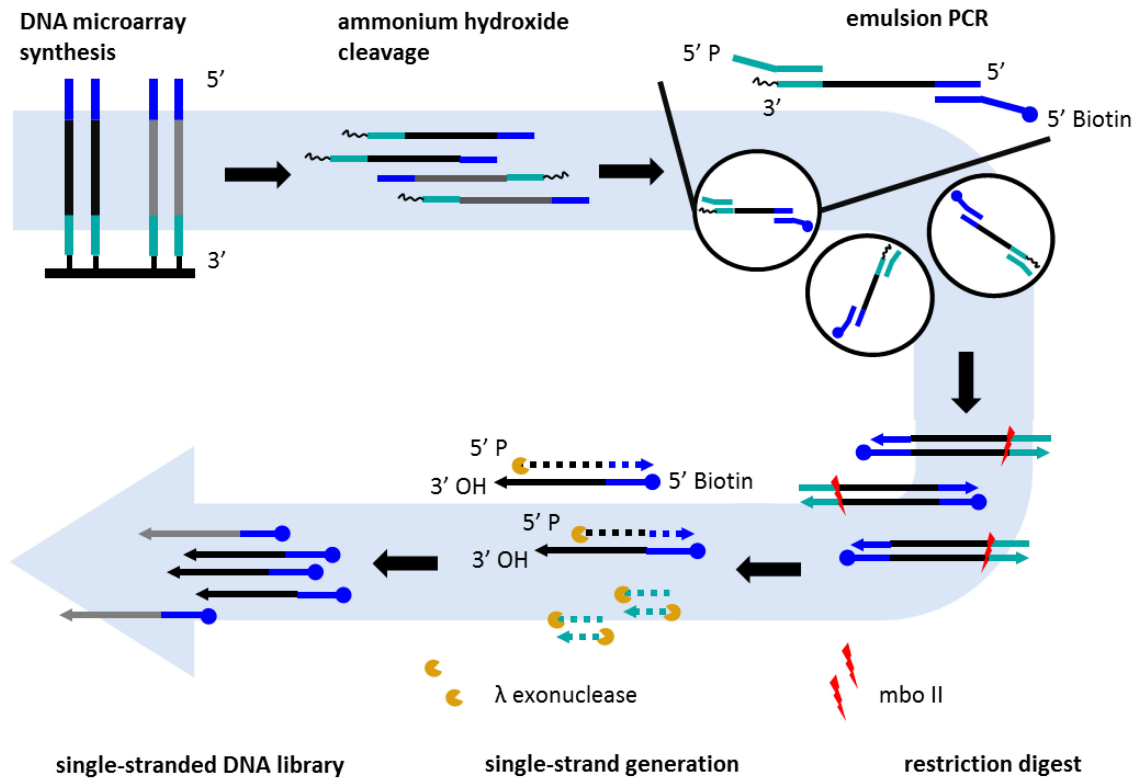


Figure 2-1 Process overview

- (1) DNA microarray synthesis of custom oligonucleotide library.
- (2) Overnight ammonium hydroxide release of surface bound oligonucleotides.
- (3) Emulsion PCR – amplification step
- (4) Removal of primer binding sequence
- (5) Exonucleolytic hydrolysis of 5'-phosphoryl strand

Purification

The cleaved DNA library is purified from short DNA fragments (<10 nt), linkers and other contaminants that inhibit the yield of polymerase chain reaction. In addition they give inaccurate spectrophotometric quantification of full length oligonucleotides. Some methods to purify oligonucleotides are i) size exclusion chromatography, ii) silica-membrane spin-column purification, iii) magnetic bead selection, and iv) gel-size selection. Size exclusion chromatography is based on gel matrix capture of low molecular weight contaminants (salts, NTP's and other LMW compounds) while large molecules are exchanged with the buffer of choice and eluted. Purification with size exclusion columns does not completely remove the unwanted molecules. In contrast, there is loss of full length oligonucleotides when purified with commercially available silica-membrane based spin columns. This is because of their high cutoff threshold of 200 bp for ssDNA (cut off threshold for single-strand DNA = 2 * cut off threshold for double strand DNA). Magnetic selection for isolation of single-strand DNA can be done either by binding biotinylated probe to streptavidin coated beads [66] or using beads with probes complementary to a common sequence [67]. The former method requires an additional on-chip coupling of biotin phosphoramidite to 5' prime terminus of oligonucleotide which labels probes irrespective of their size and consequently truncated and full length probes are purified upon magnetic bead selection. The latter method requires custom manufacturing of magnetic beads with common sequence that anneals to 3' prime end of DNA probe to achieve good purification yields [68]. As probes are synthesized from 3'-5' prime end, majority of truncated probes have the common binding sequence making this method infeasible. Thus, the only method that to some degree removes unwanted molecules prior to amplification is size exclusion spin-columns.

Amplification

The oligonucleotides have common primer binding sequence (PBS) at either ends for one-tube PCR amplification of all templates. To avoid loss of specific oligonucleotides, or introduce bias in template to product ratio as well as prevent any recombination events, the templates are amplified individually by compartmentalization in tiny droplet reactors formed upon emulsification of PCR mix.

Restriction digest

The primer binding sequences (PBS) flanking the user-sequence, especially the one at the 3' end of the user-sequence is removed with restriction enzymes that cut outside their recognition. This make the gene-specific sequence available for polymerase mediated extension and creates 5' phosphoryl terminus on the opposing strand.

Single-strand generation

After the primer removal step, one strand of dsDNA has 5'-phosphoryl terminus, while the complementary strand is not phosphorylated. Only the strand with 5'-phosphoryl group is hydrolyzed with lambda exonuclease followed by purification of products on native acrylamide gel to get high quality single-stranded DNA.

2.1.2 Emulsions background

General considerations

Emulsion is a mixture of two or more completely or partially immiscible liquids (e.g. oil and water) where one liquid (dispersed phase) is distributed in the

form of droplets (e.g. oil droplets) in the other (continuous phase; e.g. water). As oil and water do not mix spontaneously, the formation of dispersion phase requires the input of external energy, such as mechanical energy (e.g. stirring) or sound energy (i.e. ultrasound). Here, continuous energy input is required to stabilize the large surface energy (interfacial tension) associated with an increase in surface area/volume ratio. Upon cessation of external energy, the oil droplets coalesce, forming a single blob of droplet that floats to the surface (generally, oil is less dense than water). Thus emulsions are thermodynamically unstable [69, 70].

Emulsifying agents such as surfactants (surface active agents) are added to stabilize emulsions. Surfactants are amphiphilic molecules i.e. their molecular structure has both a hydrophilic and a hydrophobic part that form a protective coating (steric or electrical barrier) between the phases. Surfactants reduce surface tension, provide repulsive interaction potential between droplet interfaces and inhibit coalescence of emulsion system [71]. Surfactants are characterized by an empirical value known as Hydrophile - Lipophile Balance (HLB) on the basis of their solubility in water or oil (higher the HLB value - more water-soluble). Also, known as Bancroft's rule, it qualitatively determines the type and stability of emulsions [72]. On a scale of 1 to 20, emulsifiers with HLB value < 10 , are good to stabilize w/o emulsion i.e. water droplets dispersed in oil phase and those with HLB value > 10 stabilize o/w emulsion in which oil is dispersed in the water phase. The principle of w/o and o/w emulsion is shown in Figure 2-2 [73, 74].

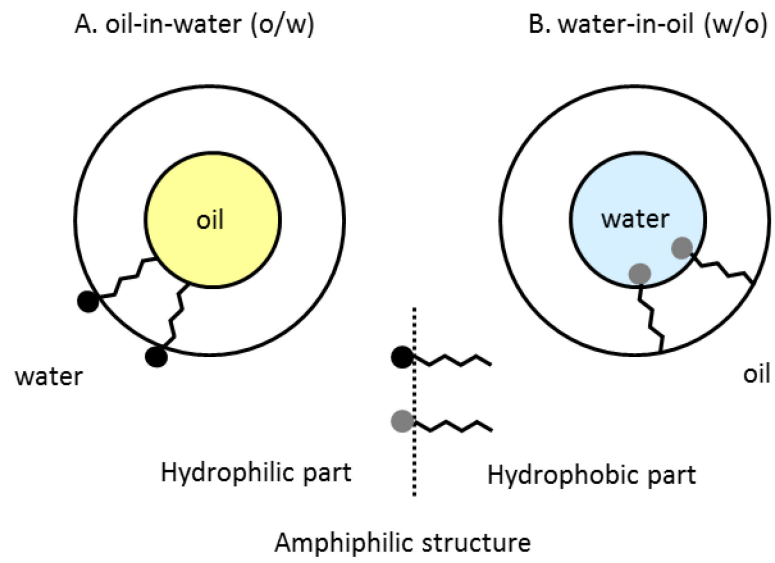


Figure 2-2 Emulsion types

Stability and breakdown mechanisms

The stability of an emulsion is measured by the rate of droplet coalescence. Factors that contribute to emulsion stability are i) Surfactant characteristics – they must diffuse rapidly to the new created interface, be soluble in both the internal and external phases and form relatively elastic films at the interface. Blends of two more surfactants (high and low HLB) make more robust emulsions (Gibbs-Marangoni effect) and improve stability against temperature changes. ii) Nature of electrical or steric interfacial barrier – interplay of repulsive effects (electrostatic, elastic, entropic, or mixing energy) with van der Waals force of attraction. iii) Viscosity of continuous phase – increase in viscosity reduces the diffusion coefficient (Einstein-Stokes Equation), thereby reducing the frequency of collisions and lowers coalescence. iv) Droplet size and distribution – monodisperse droplets form more stable dispersion against aggregation [75], v) Phase volume ratio – increase in volume of dispersed phase, decrease stability of emulsion. At high temperatures, o/w emulsions invert to form w/o emulsions. vi) Electrolyte concentration – w/o emulsions with >74% dispersed phase volume become more stable on addition of electrolytes. Here, addition of electrolytes helps to lower the refractive index difference between two phases [76]. vii) Density difference between phases and viii) Temperature – temperature effects factors such as strength of interfacial film, interfacial tension, solubility of surfactant, kinetic energy of droplets, and viscosity of liquid.

Emulsions are prone to various breakdown processes that may occur simultaneously as shown in Figure 2-3 [77, 78].

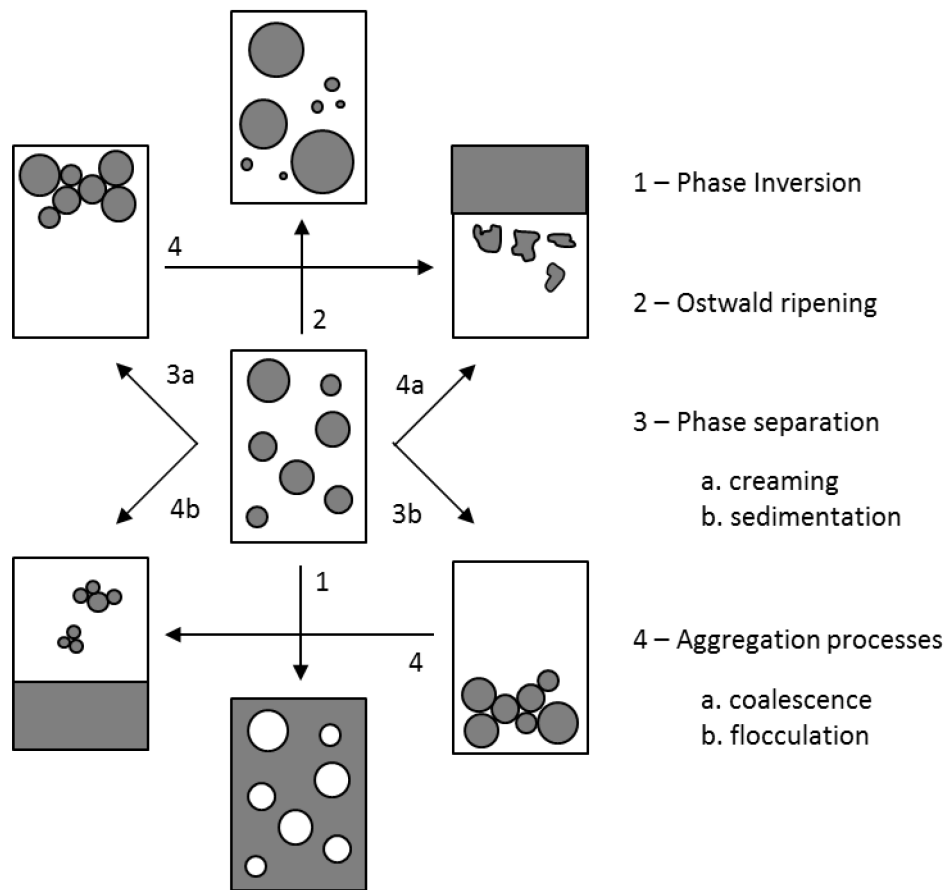


Figure 2-3 Emulsion breakdown processes

Phase Inversion – There is a sudden reversal of phases whereby an o/w emulsion becomes w/o type or vice versa. It is due to either increase in phase volume ratio above a critical value or change in solubility of surfactant towards the two phases (Bankcroft's rule). Factors that affect phase inversion are temperature, polarity of the oil phase, kind of electrolyte and its concentration, other additives (organic, water-soluble solvents increasing oil solubility in water), and oil volume fraction.

Ostwald Ripening – Is a direct consequence of the polydispersity in the droplet size distribution. It is characterized by time dependent increase in droplet size. The difference in solubility (chemical potential) between droplets with different diameters (solubility increases with decreasing droplet size – effect of surface area/volume ratio), leads to the diffusion of molecules from small droplets to large ones. For w/o emulsions, the addition of electrolytes to dispersed phase decreases ripening rate [79]. Other approaches to reduce Ostwald ripening are i) increase surfactant concentration (below critical micelles concentration, it hinders rate of new molecule incorporation), ii) add components that reduce rate of diffusion of molecules within dispersed phase. Finally, Ostwald ripening does not take place in monodisperse solutions.

Phase separation – Gravitational forces may cause large droplet (diameter > 1 μm) to settle preferentially to the top (creaming – o/w emulsion) or bottom (sedimentation – w/o emulsion) of the container. This is observed due to density difference between the phases. Shaking the container reverses the process back to its initial state. Increase in viscosity of continuous phase or lowering density difference between phases reduces probability of phase separation.

Aggregation processes – During flocculation individual dispersed droplets cluster together without losing their identity, while coalescence is the process

whereby two or more droplets fuse together to form a larger droplet. Fusion occurs as a consequence of the rupturing of the thin liquid that separates droplets at flocculation stage under the action of attractive forces or due to thermodynamic instabilities. Flocculation is the initial step towards phase separation. It can be prevented by using polymeric surfactants which provide good steric stabilization. One method to enhance emulsion stability (i.e. prevent flocculation and coalescence) is addition of small amount of ionic surfactant to high concentrations of nonionic surfactant.

2.1.3 W/O emulsions in biology

Emulsion Polymerase Chain Reaction

During replication, cells produce multiple copies of genes (DNA), transcripts (RNA) and encoded peptides and proteins. The content of an individual cell is separated from others by a phospholipid barrier (cell wall), which helps to link genotype and phenotype. Two *in vitro* methods that mimic this replicating function of cells have formed the basis of a major number of significant advances in the field of biotechnology. The first technique employs bacterial cells (yeast, mammalian cells) to clone and express multiple genes whereby populations arising from single molecules are spatially separated in colonies for easy identification and differentiation[80]. The second technique uses a cell free method known as polymerase chain reaction (PCR) to multiply molecules [81, 82].

During PCR, some templates are preferentially amplified over others attributed to factors such as primer binding energy, amplicon G+C content, amplicon length, secondary structure of gene and its accessibility within genome, gene copy number and amplification efficiency[83-87]. The co-amplification of genes or organisms (viruses, bacteria) with high sequence similarity or repetitive sequence

stretches leads to formation of chimeric DNA molecules [88, 89]. Also, DNA artifacts are generated during PCR by cross-hybridization of heterologous sequences [90, 91]. Techniques to lessen PCR bias include optimizing template concentration with number of PCR cycles [66, 84], mixing multiple replicate reactions [84, 86], optimize annealing temperature and elongation times [85, 92], and increasing temperature ramp rates during PCR cycling[93].

Other techniques involve modifying the basic PCR protocol. For example, touchdown PCR (gradual decrease of annealing temperature during cycles) and hot-start PCR (addition of polymerase at denaturation temperature) increase the specificity of DNA amplification enabling detection of low template numbers. The former prevents mispriming during gene-amplification, while the latter prevents pre-PCR mispriming and primer-dimer formation. A two-step PCR (Nested PCR) where amplification first proceeds with external primer set, followed by an internal primer set that binds within the target has been shown to improve the specificity (closely related sequences) and sensitivity (very small amount of starting material; rare DNA) over single-step conventional PCR. Finally, commonly used TAQ DNA (T. aquaticus) polymerase introduces base incorporation errors during replication process. Although, these random inserted incorrect bases are not detected during total amplicon DNA sequencing, they may prove problematic for subsequent functional analysis (for example, large scale production of proteins, development of nucleic or peptide assays) and therefore require additional selection steps. The discovery of thermostable polymerases with 3' – 5' exonuclease activity has made the fidelity of in vitro DNA replication comparable to cell-based cloning methods [94].

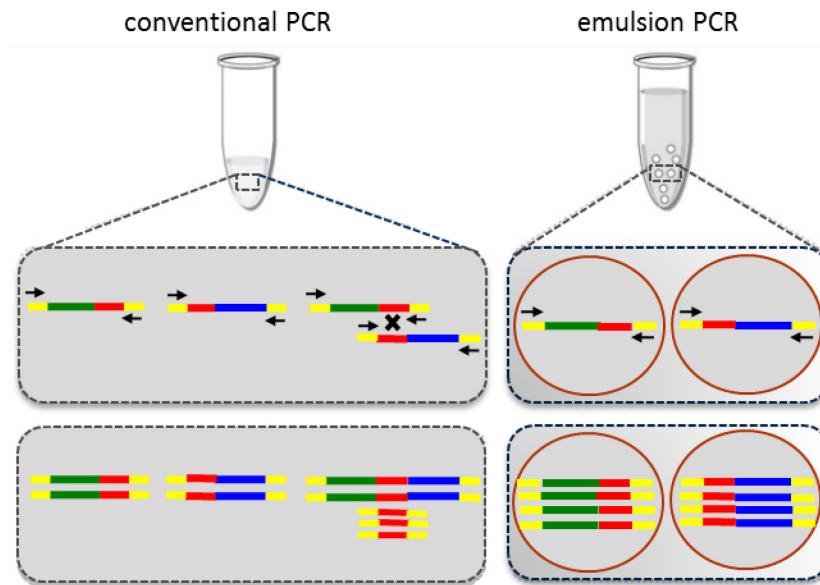


Figure 2-4 Emulsion PCR versus Conventional PCR

conventional PCR: Recombination of homologous regions leads to formation of chimeric products.

emulsion PCR: Isolated amplification in an emulsion results in equal representation of each fragment.

The use of w/o emulsions to carry out polymerase chain reaction (Emulsion PCR), whereby individual templates are segregated and amplified in millions of aqueous droplet micro-sized reactors eliminates biases associated with multi-template PCR [Figure 2-4] [95]. The knowledge of average droplet size and aqueous phase volume enable scientists to tune the template concentration such that there is only 1 template per reactor droplet. This ensures that the abundance and diversity of different templates is unaltered. The detection of single molecules requires optimization of reaction components, cycling conditions and rigorous primer design. Additionally, isolation of individual templates increases probability of primer binding to template and prevents dominance of non-specific product amplification within the reaction. Emulsion reactions have been described for detecting low copy number DNA and RNA molecules [96, 97].

Other applications

Directed evolution - Is the process of evolving proteins within the laboratory to generate protein with novel functionalities. The complexity of protein structure makes it impossible to predict the effect of change in structure on function. Many mutants of wild-type nucleotide sequence have to be analyzed to generate a protein molecule with desired traits. *In vitro* compartmentalization of genes and their products in emulsion droplets have been described for the evolution of many enzymes [98-103]. High throughput screens are possible as $>10^{10}$ reactors are formed within 1 ml of emulsion with reaction volumes in the order of femtoliters (10^{-15} L) [104]. The resultant increase in effective concentration of reactants and reduction in diffusional distance between them may cause increase in rates of reaction [105].

BEAMing (beads, emulsions, amplification and magnetics) - Is a process of solid-phase immobilization of amplicons during emulsion PCR [106]. Besides preserving clonal information upon emulsion breaking, it lends to ease in downstream manipulation of desired products, for example fluorescence activated cell-sorting [107]. Methods to immobilize amplicons to solid support either bind the amplicon directly to the bead surface via streptavidin-biotin interaction or the amplicons anneal to a covalently surface bound complementary DNA fragment. Currently, BEAMing methods are used to prepare samples for next generation sequencing with 454 life science [108] and ABI SOLiD platforms [109]. The number of templates within droplets is modeled as a Poisson distribution translating to the possibility that a large number of emulsion droplets have beads but no templates [96]. To enrich for amplicons carrying beads and increase the number of sequencing reads, modifications have been introduced during sample preparation steps [110].

2.1.4 Methods to prepare single-strand DNA from double-stranded DNA

The preparation of single stranded DNA is an important component of many procedures in molecular biology such as strand-specific microarray hybridization [111], next-generation sequencing [112], solution-based target enrichment [113, 114] and Systematic Evolution of Ligands by EXponential Enrichment (SELEX) [115].

Existing methods to convert library of double-strand DNA (dsDNA), specifically PCR amplicons to get single-strand DNA (ssDNA) can be divided into two categories, i) enzyme based or ii) affinity capture based. Enzymatic methods either selectively degrade or preferentially amplify one strand over the other. Exonuclease III and T7 gene 6 exonuclease digest both strands of dsDNA at equal rates to form

two shorter DNA fragments of half the dsDNA template length [116, 117]. Both methods require modification of the dsDNA to prepare full length ssDNA.

Legend

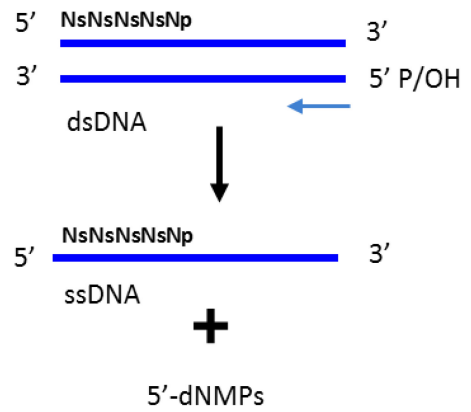
N – deoxyribonucleotide

p – phosphate bond

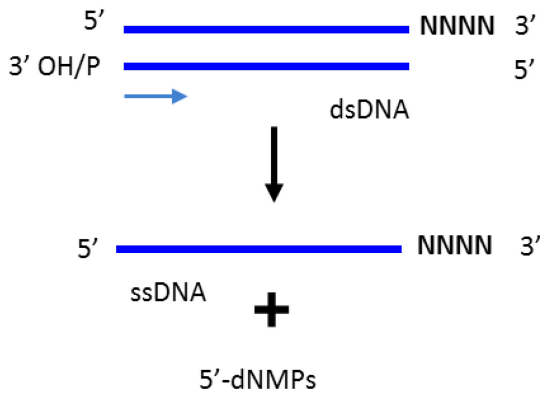
s – phosphorothioate bond

dNMPs – deoxymononucleotides

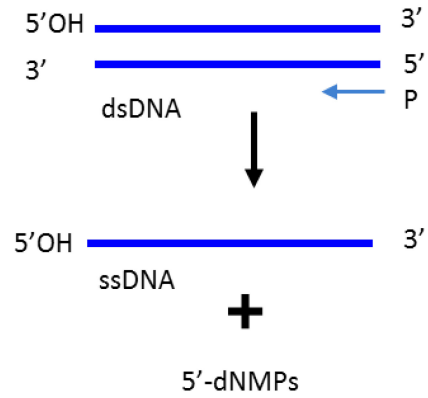
← enzyme mode of action



(a) T7 gene 6 exonuclease



(b) Exonuclease III



(c) Lambda exonuclease

Figure 2-5 Exonucleolytic hydrolysis of duplex DNA to generate ssDNA

T7 exonuclease attacks double-strand DNA substrates from 5' end. The use of phosphorothioate bonds in the DNA backbone of the first few bases at 5' end of DNA strand makes it resistant to exonucleolytic hydrolysis. PCR primers with the above modification have been used to prepare single-strand DNA for PCR based Elisa-type single nucleotide polymorphisms (SNP) detection assays [118], and Sanger DNA sequencing [Figure 2-5.a].

Exonuclease III digests duplex DNA from 3' end. The incorporation of 2-phosphorothioate bonds at 3' end of DNA strand makes them resistant to exonucleolytic hydrolysis. As Exonuclease III is inactive on ssDNA, duplex DNA digested with restriction enzymes that generate four-base or longer 3'-protrusions to generate single-strand DNA have been successfully used for sequencing [Figure 2-5.b] [119].

Lambda exonuclease is a highly processive enzyme that preferentially digests duplex DNA strands with 5' terminal phosphate [Figure 2-5.c] [120]. The enzyme is well characterized and has been shown to have greatly reduced activity for 5' terminal hydroxyl strand of native and denatured DNA Article [121, 122]. Lambda exonuclease has been used to prepare single-stranded DNA for wide variety of applications. For example, probes for *in situ* hybridization [123], capture probes for next-generation DNA sequencing [124], dye-labeled targets that improve microarray hybridization signal-to-noise ratios [125], and DNA-based sensors [126].

Asymmetric PCR (AsymPCR; thermostable DNA polymerase) is used for preferential amplification of one strand over the other by using unequal primer concentrations to get single-strand DNA targets [Figure 2-6.a]. The technique requires rigorous primer design dependent on thermodynamic properties of both the product and primers and is limited by the size of PCR product [127]. Although, difficult to optimize and requiring 50+ cycles to produce high single-strand to

double-strand DNA ratio, variations of asymmetric PCR have been used in diverse applications such as profiling of yeast knockout strain collection [128], in multiplexed detection of drug resistant genes [129], and single cell genetic diagnostic [130].

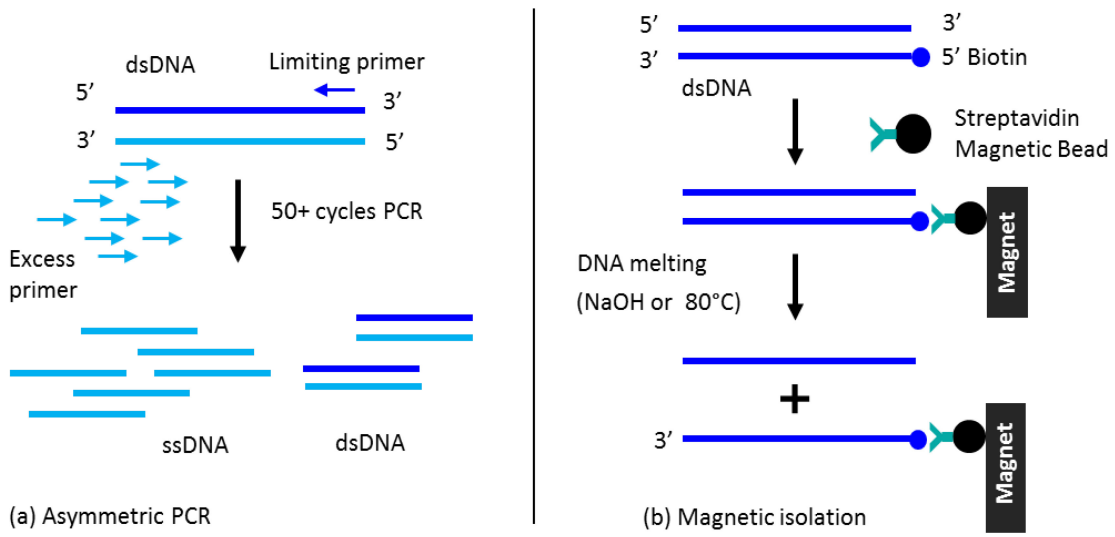


Figure 2-6 Formation of single-stranded DNA

Magnetic isolation - An alternative non-enzyme method to get single-stranded targets from double-strand DNA is strand separation using biotinylated primers and streptavidin-coated magnetic beads [Figure 2-6.b]. Depending on downstream application, they can be captured either directly onto the bead surface [131, 132] indirectly captured via hybridization on pre-immobilized probes [133]. Next-generation-sequencing technologies use both capture techniques during library preparation and sequencing steps, for example enrichment of starting sample (poly (A) tail RNA, known DNA or RNA sub-populations), capture of biotinylated DNA for emulsion PCR, and selection of beads that have emulsion PCR products [134].

***In vitro* transcription and reverse transcription** - Reverse transcriptase in the presence of a suitable primer synthesizes a complementary DNA copy of an RNA template [135, 136]. This coupled with an upstream RNA amplification has the potential to generate large quantities of ssDNA probes [137].

2.2 Experimental methods and protocols

2.2.1 Process outline

Processing chip-derived oligonucleotide libraries

The synthesized oligonucleotide library is released from microarray in large volume of alkaline solution (30% ammonium hydroxide). The volume is reduced to 30 – 40 μ l by either evaporating the excess solution under vacuum or ethanol precipitation. The DNA library is purified on size-exclusion spin columns using recommended protocol (CS-20, Princeton Separations) and quantified by Nanodrop spectrophotometer (Thermo Scientific).

Standard curve – Real time PCR

115-mer gel-purified oligonucleotide is obtained from Integrated DNA Technologies. 10 x serial dilutions of the oligonucleotide ranging from 100 femtomole (~60 billion molecules) to 100 yoctomole (~6 molecules) are amplified by real time PCR (Chromo4-PTC 200 thermocycler, MJ Research). The PCR master (devoid template) consists of 0.5 μ M forward and reverse primers, 2.5 mM MgCl₂, 1 X GoTaq-flexi buffer, 0.2mM dNTP, 2 μ M SYTO13 dye and 2.5 U GoTaq-flexi polymerase (Promega Inc). The final reaction volume at each template concentration is 50 μ l. The amplification program is comprised of 35 cycles with the following steps: initial denaturation at 94°C for 2 min, 35 cycles of denaturation at 94°C for 15 s; primer annealing at T_m -5°C for 20 s; primer extension at 72°C for 20; and final elongation at 72°C for 5 min. The signal intensity of SYTO13 dyes is measured immediately after the annealing and extension temperature [138]. A melt curve is generated by slowly increasing the temperature in steps of 0.5°C from 60°C

to 95°C. The tubes are held for 5 s at every temperature increment, followed by measurement of dye intensity. The data is analyzed using Opticon software available with the thermocycler.

Emulsion PCR

80 μ l of ABIL EM90 (4%) and 1 μ l of Triton X-100 (0.5%) surfactants are added to 2 ml of light mineral oil and rigorously mixed for 1 min and out on ice. The PCR (aqueous) mix contains 1 – 5 μ l of template, 0.5 μ M each of forward and reverse primer, 1 X GC buffer (proprietary solution), 1 mM MgCl₂, 0.5 μ g/ μ l BSA and 2 U/50 μ l hot-start high-fidelity phusion DNA polymerase.

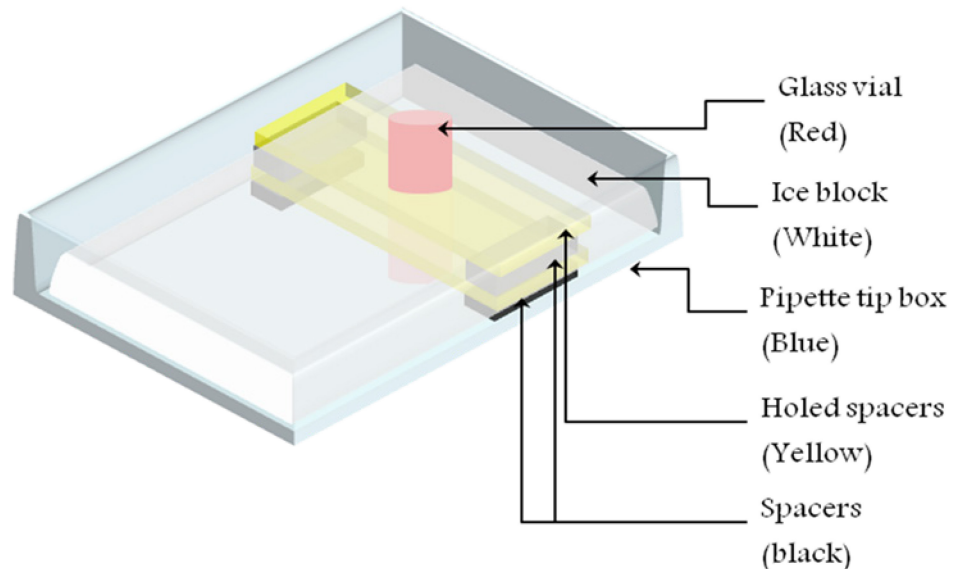


Figure 2-7 Emulsion PCR

The oil phase (4% ABIL EM90; 0.05% Triton X-100) is transferred to 2 ml glass vial containing magnetic stirrer [Figure 2-7]. The emulsion is prepared by drop-wise addition of PCR solution to oil phase (stirred at 1000 rpm), followed by additional stirring for 15 minutes. The ratio of aqueous: oil phase is maintained at either 1:4 or 1:6 and the final volume of emulsion is kept between 200 – 600 μ l. The emulsified product is transferred to 0.2 ml PCR tubes or PCR plate in aliquots of 50 μ l. PCR is performed in an Eppendorf thermocycler. The amplification program is comprised of 35 cycles with the following steps: initial denaturation at 98°C for 1 min, 30 cycles of denaturation at 98°C for 15 s; primer annealing at T_m -5°C for 20 s; primer extension at 72°C for 20 s; and final elongation at 72°C for 5 min. After PCR amplification, the products equivalent to 100 μ l of aqueous phase volume (i.e. 400 – 600 μ l emulsion volume) are pooled together in 1.6 ml centrifuge tube containing 5 μ l of loading buffer (indicator – orange G/xylene cyanol, to help identify water droplet in organic solvents), followed by addition of mineral oil to 1.4 ml final volume. The contents of tubes are mixed by vortex for 1 min. The emulsion is separated into oil and aqueous-surfactant phase by centrifugation at 13,000 g for 10 min. The upper (oil) phase is removed and the aqueous-surfactant phase is broken by successive washes with 1 ml water saturated solutions of diethyl ether and ethyl acetate. In each case, the contents of the tubes are mixed by vortex (10 s) and collected by centrifugation (5 s), followed by removal of upper solvent phase. The bottom phase contains the PCR products. After the final diethyl ether wash step the tubes are placed on 37°C heat block to evaporate trace amounts of leftover diethyl ether.

Restriction digest

1 µg of PCR amplicons are digested with 20 units of Mbo II restriction enzyme (New England Biolabs Inc.) in 50 µl reaction for minimum 2 h at 37°C. The buffer formulation is 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM Dithiothreitol pH 7.9 @ 25°C (1 X NEB buffer 2). For DNA size >100 bp, the enzyme is heat inactivated at 65°C for 20 min. To avoid possible strand separation and incorrect strand re-annealing, the enzyme is not heat activated for duplex DNA < 100 bp. dsDNA is purified with MinElute PCR purification kit (Qiagen Inc.) following manufacturer's protocol with following modifications, i) nucleic binding step is performed at 6,000 g for 1 minute, ii) a second elution step is performed with 10 µl of EB buffer and centrifuged for 3 minutes. The concentration of DNA (2 µl) is measured with Nanodrop spectrophotometer and saved for gel analysis.

Preparation of single-strand DNA

The 5'-phosphoryl strand of restriction digest product is hydrolyzed in 1 X lambda exonuclease buffer (67 mM Glycine-KOH, 2.5 mM MgCl₂, 50 µg/ml BSA, pH 9.4 @ 25°C) at 37°C for 30 minutes (unless stated). 1 unit of enzyme is used for every 5 pm of double strand DNA. The concentration of dsDNA in final reaction volume is kept between 0.5 pm/µ to 2 pm/µl. Finally, the enzyme is inactivated by heating at 75°C for 15 min and the products are run on acrylamide gel for purification.

2.2.2 Yeast Knock-out (YKO) deletion collection

Frozen (-80°C) samples of pooled mat-alpha mating YKO deletion strains is thawed on ice and diluted in 2% glucose rich media. The cells are spread on rich media supplemented with 200 µg/ml geneticin (G-418; Gibco) and grown at 28°C (2

days). They are recovered in ringer's solution (0.4% NaCl) and cells are counted using hemacytometer.

Yeast genomic DNA (20kb – 40kb) is extracted from 10 OD₆₀₀ of cells (1 OD₆₀₀ = 2.7E+07 cells/ml) using 'smash and grab' method [139]. Acid washed glass beads (400 – 500 µm) and phenol chloroform solution (pH 8.0) are used to break open the yeast cell wall under rigorous mixing followed by multiple steps of phenol/chloroform extraction. The protocol is modified to include a protein digestion (Proteinase K) and additional phenol/chloroform and chloroform extraction steps after RNase A treatment to yield high quality genomic DNA (OD₂₆₀/OD₂₈₀ > 1.8).

Touchdown PCR

PCR amplicons corresponding to the either UPTAG or DNTAG modules are amplified using universal primer sequences. The PCR reaction mix (50 µl) contains 200 ng yeast genomic DNA, 0.5 µM forward and reverse of UPTAG or DNTAG primers, 2.5 mM MgCl₂, 1 X GoTAQ flexi buffer, 0.2mM dNTP and 2.5 U GoTAQ polymerase (Promega Inc). The cycling parameters are as follows: initial denaturation at 94°C for 2 min; 4 cycles of dsDNA denaturation at 94°C for 20 s, primer annealing at initially T_m+6°C for 30 s with 2°C decrease per cycle; primer extension at 72°C for 30 s; 30 cycles of denaturation, primer annealing and extension at constant annealing temperature T_m-4°C; and final elongation at 72°C for 5 min.

Restriction digest of yeast genomic DNA

1 µg of genomic DNA is separately digested with 20 U DraI (recognition site: 5'-TTT/AAA-3') and 20 U NdeI (recognition site: 5'-CA/TATG-3') for 37°C for 60

minutes. The buffer components are 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate and 1 mM dithiothreitol (DTT) buffer (1 X NEB 4 buffer). In a third tube, 20 µg of genomic DNA is then digested overnight with 60 units each of DraI and NdeI in 1 X NEB buffer at 37°C. In all cases, the enzymes are inactivated after digestion by incubation at 65°C for 20 min.

2.2.3 Enzyme based methods to prepare single-strand DNA from duplex DNA

PCR amplification and restriction digest

300 base-pair (bp) amplicons with UPTAG module is amplified with modified U1 universal primer and 5'-hydroxyl KanB by touchdown PCR. In case of T7 gene 6, the first 5 bases at 5' end of U1 primer are linked by phosphorothioate bonds. In case of lambda exonuclease, the U1 primer has a 5'-phosphoryl group. The amplicons are purified on QiaQuick PCR purification columns and eluted with 50 µl of Tris-HCl pH 8.0. 1 µg of purified amplicons are digested with 20 units of Bgl II restriction enzyme (20 µl, 1x NEB buffer 4) for 30 min at 37°C, followed by heat inactivation of the enzyme for 20 min at 65°C.

T7 gene 6 exonuclease and 5'-phosphorothioate backbone strand

0.5 µg of amplicon (phosphorothioate backbone) is digested with 50 U of T7 gene 6 exonuclease (Affymetrix Inc.) for 45 min at 37°C in 1x T7 gene 6 exonuclease reaction buffer (20 µl). The enzyme is inactivated by heating at 75°C for 10 min.

Lambda exonuclease and 5'-phosphoryl terminus strand

0.5 µg of PCR amplicons with 5' phosphorylated ends are digested with 10U of lambda exonuclease (New England Biolabs) for 60 min at 37°C in 1x lambda exonuclease buffer (20 µl). The enzyme is inactivated by heating at 75°C for 10 min.

Exonuclease I hydrolysis of ssDNA

An aliquot of ssDNA product is hydrolyzed with 1 U/ μ l exonuclease I (20 μ l; New England Biolabs) for 60 min at 37°C to detect for presence of ssDNA. Equal concentrations of unpurified PCR, BglII digest, single-strand formation reaction and exonuclease digestion products are electrophoresed on native polyacrylamide gels.

2.2.4 Preparation of single-strand DNA with 5'-modified end

300 bp UPTAG module region of YKO collection is amplified in six separate PCR reactions with forward U1 primer modified by 3 different 5'-terminus groups (Cy3, Cy5 and hydroxyl) and reverse KanB primer with 5'-hydroxyl or 5'-phosphoryl group. 5 pmol of PCR product is digested with lambda exonuclease (1 U enzyme / 5 pmol DNA) at 37°C for 45 min (10 μ l). The enzyme is heat inactivated at 75°C for 15 minutes. 2.5 pmol of PCR and lambda exonuclease products are digested with exonuclease I for 30 min (5 μ l). Equal concentrations of the 4 products are run on 6% native acrylamide gel.

2.2.5 Optimize digestion time of Lambda exonuclease

146 bp region of the YKO genomic DNA is amplified using 5'-phosphoryl forward primer and 5'-hydroxyl reverse primer by PCR. 400 ng (~ 4 pmol) of PCR amplicons are digested with lambda exonuclease (1 U enzyme / 5 pmol dsDNA) in 5 μ l reaction at 37°C at varying time points (1, 2, 5, 10, 15, 30 and 60 min) in separate 0.2 ml PCR tubes. Lambda exonuclease samples at 15, 30 and 60 minutes as well as PCR sample are hydrolyzed with exonuclease I enzyme at 37°C for 30 min (10 μ l). The products are run on 8% native acrylamide gel.

2.2.6 In vitro transcription and reverse transcription

4 pmol of emulsion PCR products is transcribed in 40 µl reaction with AmpliScribe™ T7-Flash™ Transcription Kit (Epicentre Biotechnologies Inc.) at 42°C for 4 h. 1 µl of TURBO DNase is added to the reaction and incubated at 37°C for 15 min, followed by inactivation of DNase with 2 µl of inactivation reagent following TURBO DNA-free kit (Ambion Inc.). 36 µl of the DNase inactivated solution is suspended in 144 µl nuclease free water and 20 µl 3 M sodium acetate (pH 5.2) and purified on two RNeasy columns (Qiagen Inc.). The RNA is eluted with 50 µl of nuclease free (total 150 µl). The concentration is measured with Nandrop spectrophotometer. RNA is reverse transcribed with superscript II (Invitrogen Inc.) following manufacture's protocol with a few modifications; i) 100 pmol of RNA is used per 20 µl reaction, ii) use of 1.5 x sequence-specific primer (150 pmol), iii) the hybridization reaction is heated at 65°C for 5 min and slowly cooled (-0.1°C/s) to 37°C, iv) the reaction after addition of enzyme is incubated at 37°C for 3h and v) the reaction is scaled to process 1200 pmol (240 µl) of RNA. RNA is hydrolyzed with 0.43 volume (103.2 µl) of 1 M NaOH at 65°C for 15 min, followed by inactivation with equal volume (103.2 µl) of 1 M HCl. The cDNA is ethanol precipitated with 0.3 M sodium acetate (pH 5.2, 50 µl). Post precipitation the DNA pellet is suspended in 75 µl nuclease free water and purified with three centri-spin20 column (25 µl each).

2.2.7 SYBR gold gel staining

Polyacrylamide gels are stained using SYBR gold dye [80]. The top plate is removed with the gel on the bottom plate. Excess gel not containing any nucleic acid products is cut out, followed by washing the gel with deionized water. 5 µl of 10,000x SYBR gold (Invitrogen Inc.) is suspended in 10 ml of tris-HCl pH 8.0 (10x) and poured over the gel. The nucleic acids are stained in the dark for at least 20 min.

The gel is washed to remove excess dye and transferred on to a UV transilluminator for analysis and image capture with camera (fitted with orange filter).

2.2.8 Polyacrylamide gel (PAGE) purification

The desired DNA band is excised from the gel and transferred to 1.6 ml tube. The DNA is extracted from the gel using 'crush and soak' method [140]. Post incubation (>20 h), the buffer solution is run through 22 μ M spin column filter to separate out the gel fragments and the DNA is ethanol precipitated. The re-suspended DNA is purified on silica-based filter columns or size-exclusion columns to get cleaner product.

2.2.9 Phenol/ Chloroform/ Isoamyl Alcohol (PCIA) purification

For enzyme inactivation or purification of reactions where DNA or both DNA and RNA are needed for downstream applications, we use PCIA pH 8.0, following standard protocol [141] with minor modifications; i) if minimum reaction volume (100 μ l) is not met, it is increased by addition of nuclease free water, ii) 0.5 M NaCl is added to aqueous phase iii) the entire upper phase is recovered and processed with chloroform pH.0 to remove any trace amounts of phenol, and iv) the contents of the tube are vortex for 1 min followed by centrifugation for 5 min at >10,000 g to separate the two phases. Post chloroform extraction, the upper phase is recovered and precipitated using ethanol.

2.3 Experimental results and discussion

2.3.1 Yeast knock-out collection (YKO)

The Saccharomyces Genome Deletion project has made as complete a set as possible of yeast knock-out (YKO) strains with the overall goal of assigning function to the ORFs through phenotypic analysis of the mutants [142-144]. PCR-based gene deletion strategy is used to generate a start- to stop- codon deletion of each of the ORFs in the yeast genome. As part of the deletion process, each gene disruption is replaced with a KanMX module and uniquely tagged with one or two 20mer sequence(s), assigned as UPTAG and DNTAG sequence. The tags are detected via hybridization to a high-density oligonucleotide array, enabling growth phenotypes of individual strains to be analyzed in parallel. YKO genomic DNA is used to validate and optimize protocols for emulsion PCR and enzyme based ssDNA generation methods.

2.3.2 Interpretation of gel images

Each gel image has a ladder that is used as a size marker to indicate the position of the DNA or RNA product. The size of relevant DNA or RNA band is shown to the right of the figure for interpretation.

2.3.3 Dependence of emulsion PCR on template size and concentration

The goal of this experiment is to study the effect, if any of the template size on the success of emulsion PCR. Whole and digested fractions of genomic DNA are analyzed on 0.8 % TBE agarose gels [Figure 2-8]. 10 kb ladder are run as size markers (lane 1, 4, 7). We see that uncut genomic DNA (lane 3, 7) runs above 10 kb size marker. In restriction digest with one enzyme, there is significant amount of

high molecular weight DNA (lane 2, 5; smear from 4 kb to > 10 kb). Double enzyme restriction digest eliminates the high molecular weight DNA and gives DNA size distribution between 4 kb and <0.5 kb (lane 8).

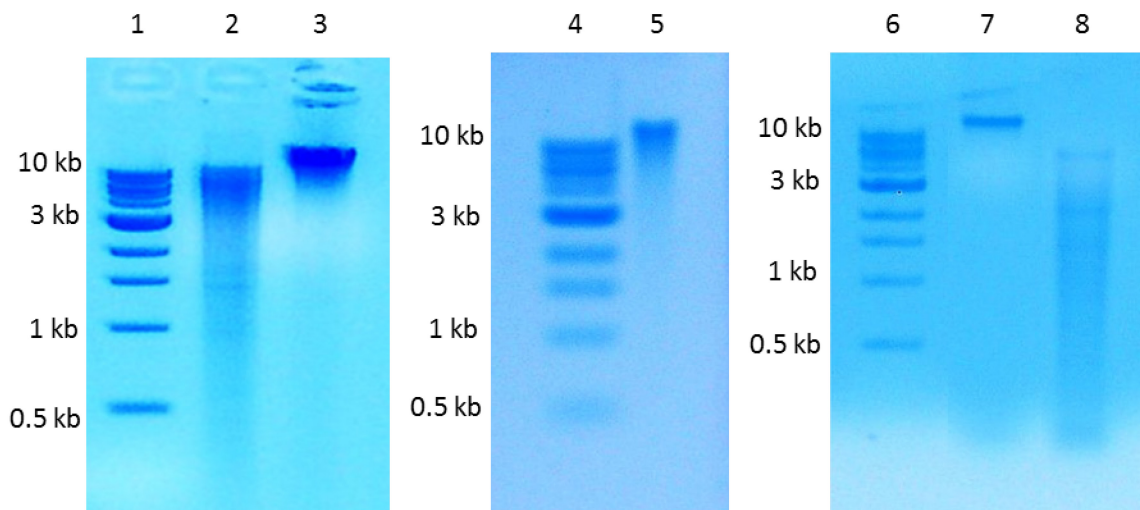


Figure 2-8 Restriction digest of yeast genomic DNA

200 ng of genomic DNA is amplified by both PCR methods (conventional PCR and emulsion PCR) to detect two different amplicons (200 bp, 300 bp) and double digest genomic DNA (DraI / NdeI digestion) is amplified by both methods to detect only the 300 bp amplicons. The basic emulsion PCR protocol described in methods sections has a few modifications, i) 5U GoTaq polymerase (Promega Inc.) and ii) aqueous: oil phase ratio is 1:8. Here, no-template' control is used to test for PCR contamination. 20% of non-emulsified and 100% of emulsified reactions products are run on 1 % TBE agarose gel [Figure 2-9].

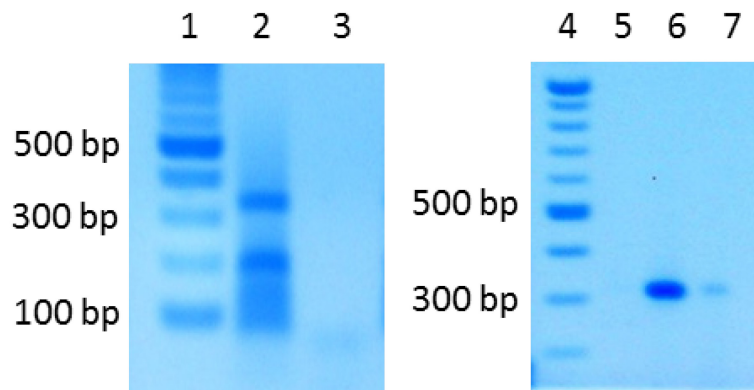


Figure 2-9 Effect of template size on emulsion PCR

In case of whole genomic DNA, there is no amplification of products in emulsion PCR (lane 3), compared to conventional PCR (lane 2). For PCR amplification on digested genomic DNA, there is desired DNA band for both conventional PCR (lane 6) and emulsion PCR (lane 7). This experiment shows that the success of emulsion PCR amplification is dependent on the size of template and droplet diameter i.e. the DNA template molecule in an expanded form should fit into the emulsion droplets. The maximum DNA length when droplet diameter is 5 μm should be less than $34 \text{ nm} * x = 5 \mu\text{m}$, $x \sim 14.7 \text{ kb}$. Now, there is a significant difference between the yields of PCR product between the two PCR methods. The low amount of product in emulsion PCR may be due to low concentration of digested DNA fragment with the target sequence such that there is no amplification in many of the emulsion droplets.

2.3.4 Dependence of emulsion PCR on type of oil

The stability of w/o emulsions is influenced by the viscosity of the oil (continuous phase), viscosity of aqueous phase (dispersed phase) and the viscosity difference between the phases. Also higher viscosity of continuous phase can reduce droplet size and prevent coalescence [145]. To test for effect of oil type on emulsion PCR outcome, two oils namely mineral oil (viscosity – 15 mPa.s at 40°C, non-polar) and TEGOSOFT® DEC (viscosity – 5 mPa.s at 25°C; medium polarity) are used to prepare the oil phase.

The surfactant concentration is kept constant at 4% ABIL EM 90, 0.05 % Triton X-100, in two different oil formulations; i) 50 % mix of mineral oil and TEGOSOFT® DEC or ii) 100 % light mineral oil to amplify the 300 bp sequence. For this experiment we used a 2 kb PCR amplified fragment of YKO genomic DNA

(gDNA). This has two advantages namely, the template size is small enough to fit in the emulsion droplet (no need of digestion) and the templates have uniform length (can have 1 template per droplet). Here, we used two different template concentrations equivalent to approximately 2 (5 fmol) and 20 (50 fmol) templates per droplet. The expected amplicon size is 300 base pair. 20 % of conventional and emulsion PCR products are run on 1% TBE agarose gel [Figure 2-10].

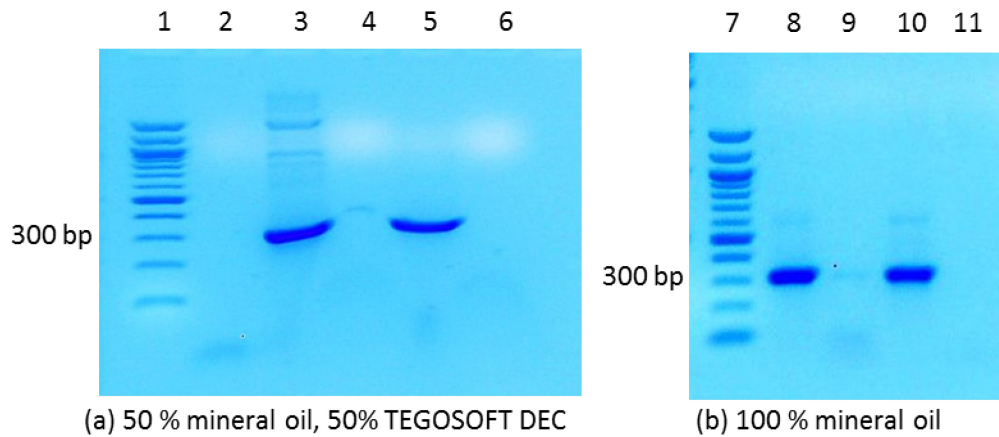


Figure 2-10 Effect of oil type on emulsion PCR

From the gel image, there is no PCR product in emulsions formed with blend of two oils (lane 4, 6), whereas with 100% mineral oil based emulsion, there is amplification of the desired 300 bp target sequence for both template concentrations (lane 8 – 5pmol and lane 10 - 50 pmol). The failure of amplification in the oil blend is attributed to the breakdown in emulsion as concluded from the settling of emulsion at the bottom of tube. Now, the oil blend of TEGOSOFT® DEC and mineral oil when stabilized with another surfactant ABIL WE09 (cetyl dimethicone copolyol-co-polyglycerol-4-ido-stearate-co-hexyl laurate) does amplify PCR products [106]. This validates that the selection of oil and surfactant are critical factors for success of emulsion PCR [146].

In this experiment the template concentration is increased to get roughly 1 template per droplet. This significantly increases the yield of emulsion PCR (Figure 2-10, lanes 8 and Figure 2-9, lane 7). Also, in conventional PCR method the starting template concentration plays a role in the formation of non-specific products (lane 3, 5). There are present multiple DNA bands above the 300 bp sequence for 50 fmol template concentration (lane 3), which are not detected at lower template concentration (lane 5 – 5 pmol). Here, we that the ‘no template’ PCR controls (lanes 2, 9 and 11) show no contamination.

2.3.5 Enzymatic hydrolysis methods to prepare single-stranded DNA from duplex DNA

We focused on two enzyme hydrolysis methods i.e. T7 gene 6 exonuclease and Lambda exonuclease to prepare single-strand DNA from duplex PCR amplicons. Other methods such as LATE-PCR and exonuclease III are not tested due to restrictions imposed by upstream processing of surface-cleaved oligonucleotides and downstream applications of the single-strand DNA library. The 50+ cycles

required to get high single-strand DNA to duplex DNA ratios, rigorous primer and amplicon design criteria for reproducible amplification and stability of emulsion (< 35 cycles) make LATE-PCR unsuitable to prepare for single-strand DNA. In addition, truncated sequences in the microarray derived library would be amplified by this method. Recently, exonuclease III has been shown to digest single strand DNA [147]. The authors also report that only phosphorothioate linkages with R stereochemistry are resistant to exonucleolytic cleavage, not those with S stereochemistry. This conflicts established results where the S stereoisomer and not R stereoisomer imparts resistance to exonuclease degradation. Experiments to study the hydrolysis of unmodified strand of dsDNA by both T7 gene 6 exonuclease and Lambda exonuclease to generate ssDNA is performed on UPTAG sequence incorporated in Yeast Knock-out (YKO) deletion collection.

T7 gene 6 exonuclease and 5'-phosphorothioate backbone strand

Oligonucleotide strands with phosphorothioate backbone are resistant to exonucleolytic hydrolysis by enzyme T7 exonuclease [118, 148]. To test of the above, template DNA is amplified by touchdown PCR to incorporate phosphorothioate bonds in the last 4 bases at 5'-terminus of one strand, while the complementary strand has phosphate backbone.

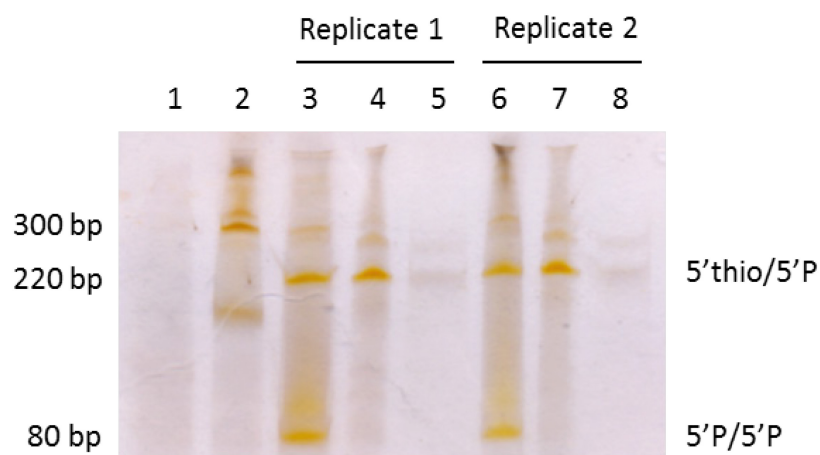


Figure 2-11 T7 gene 6 exonuclease and phosphorothioate backbone

From Figure 2-11, the 300 bp PCR product (lane 2; orange band) is digested (Bgl II) to generate two fragments of 220 bp and 80 bp (lane 3, 7; orange bands). Here, the 220 bp dsDNA fragment has one strand with phosphorothioate backbone (5'thio) while the other has phosphate backbone (5'P) (lane 3, 6). On the other hand the 80 bp dsDNA fragment has both strands with phosphate backbone (lane 3, 6). The 80 bp (5'P/5'P) dsDNA get digested with T7 exonuclease (lane 4 and 7, no band). When the 220 bp dsDNA band is hydrolyzed with T7 exonuclease, we see DNA bands (lane 4, 7; 200 bp). The formation of 220 bp ssDNA (lane 4, 7) is confirmed by digestion with 3'-5' acting single-strand DNA specific exonuclease I (lane 5, 8).

Now, the restriction digest reaction is incomplete as indicated by presence of 300 bp band in all lanes except lane 1, 5 and 8. Lane 1 is negative PCR control (no-template) and shows no products. The presence of background bands and smear are due to silver deposition on proteins and molecules other than desired DNA. The

reactions are performed and run in duplicate to minimize errors. Thus, 4 phosphorothioate bonds in the DNA backbone protect it from digestion with T7 exonuclease and this method can be used to make single-strand DNA.

Lambda exonuclease and 5'-phorylated strand

A similar experiment as described above is done to determine the efficiency of lambda exonuclease to preferentially digest the 5'-phosphorylated strand of double-strand DNA.

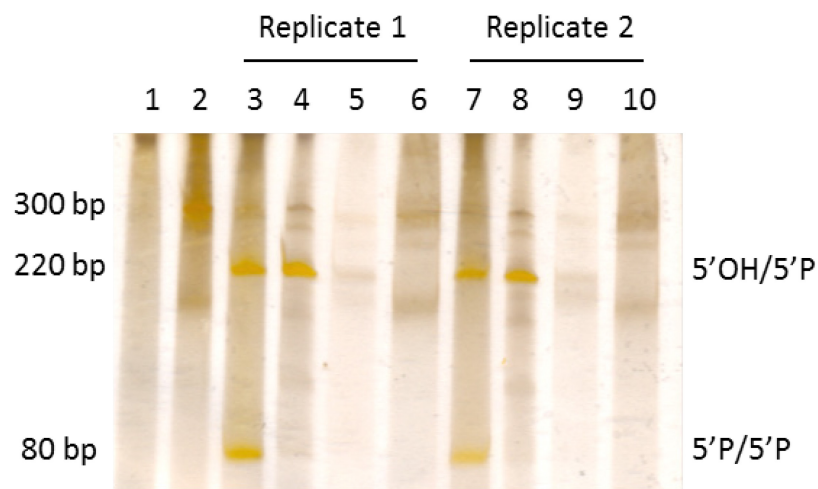
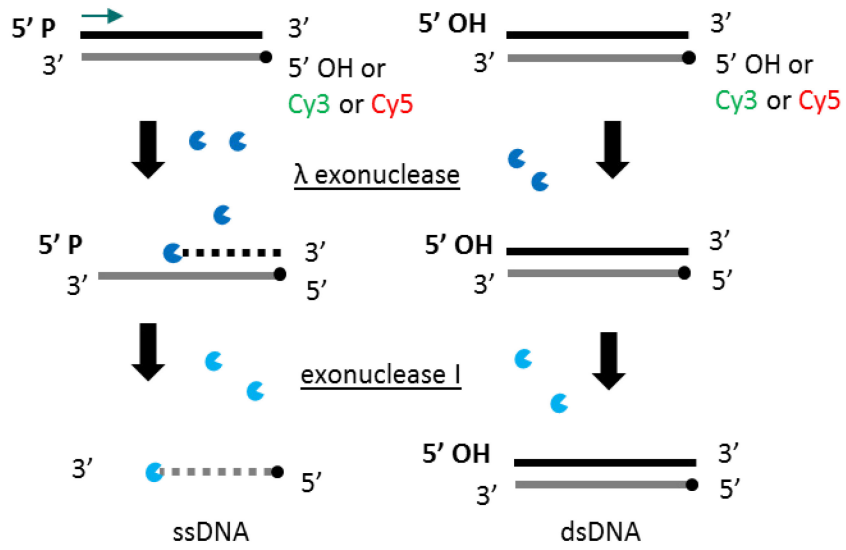


Figure 2-12 Lambda exonuclease hydrolysis of 5' phosphoryl end

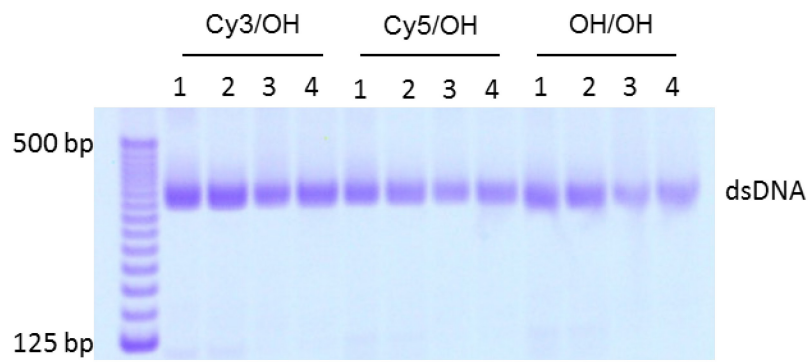
From Figure 2-12, 300 bp PCR amplicons (lane 2) with one 5'-phosphoryl (5'P) strand and other with 5'-hydroxyl (5'OH) complementary strand are digested (Bgl II) to generate two smaller DNA bands of size 220 bp and 80 bp (lane 3, 7). The 220 bp dsDNA has 5'P/5'OH strand configuration, while the 80 bp dsDNA has 5'P/5'P strand configuration. For the 80 bp dsDNA both strands with 5'-phosphoryl termini get hydrolyzed completely by lambda exonuclease (lane 4, 8). In case of the 220 bp dsDNA only the strand with 5'-phosphoryl termini get hydrolyzed leaving behind the 5'OH nucleic acid band (lane 4, 8), which is hydrolyzed upon addition of single-strand exonuclease (lane 5, 9). Also, exonuclease I is able to digest some amount of dsDNA (lane 5, 10). This is attributed to sufficient 'breathing' of duplex DNA structure (formation of single-stranded regions) during enzyme inactivation at 80°C for 20 min. Although, there is degradation of duplex DNA by exonuclease I, the disappearance of 80 bp band during hydrolysis reaction confirms formation of ssDNA by lambda exonuclease. The unidentified background bands are due to indiscriminant silver staining of enzymes, proteins, lipids and molecules other than nucleic acids. Both T7 gene 6 exonuclease and lambda exonuclease methods result in conversion of dsDNA to ssDNA with equal success.

2.3.6 Preparation of single-strand DNA with 5'-modified end

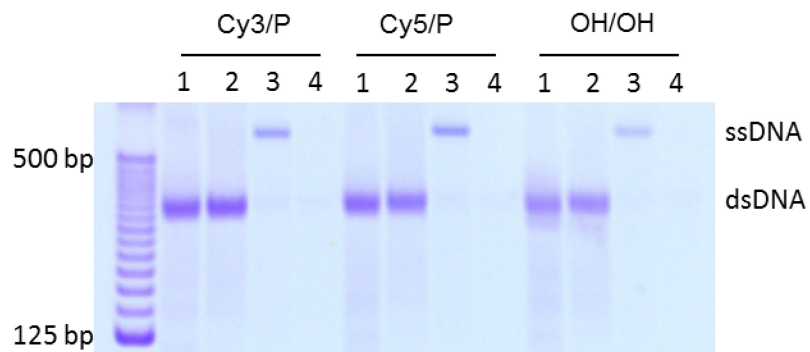
For downstream applications, especially to improve signal intensity on microarrays [149], we wanted to have the flexibility to make ssDNA with 5'-fluorescent dyes or biotin moiety. As formation of ssDNA by T7 exonuclease depends on protection of DNA backbone from exonucleolytic attack and not the identity of 5'-terminus group, this method is not investigated further.



(a) Experimental outline



(b) 5' mod / 5'OH



(c) 5' mod / 5'P

Figure 2-13 Lambda exonuclease activity on modified 5'-ends

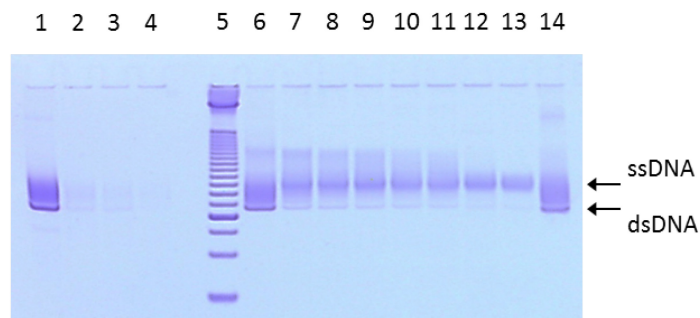
Here, 3 separate forward primers with 5' - fluorescent (5'Cy3, 5'Cy5) or hydroxyl (5'OH) termini are used with either 5'-phosphoryl (5'P) or 5'-hydroxyl reverse primers are used to independently amplify 6 different PCR products. The PCR products with the experimental process and expected results are shown in [Figure 2-13.a]. Both dsDNA and ssDNA are digested with exonuclease I. The exonuclease I inactivation protocol is modified to prevent digestion of dsDNA due to 'breathing' of dsDNA at the ends. This is done by adding 20 mM EDTA to inhibit some of the exonuclease I activity immediately followed by heat inactivation at 80°C for 20 min.

Equal concentration of template is used for each step of the reaction and all the products are loaded on the gel [Figure 2-13.b. and Figure 2-13.c]. Each of the 6 independent PCR products has 4 lanes on the gel. The 1st and 3rd lanes contain PCR and lambda exonuclease products; while the 2nd and 4th lanes contain exonuclease I digested PCR and lambda exonuclease products respectively. Across the 6 reactions, exonuclease I is unable to digest dsDNA (lane 1, 2), indicating the success of the inactivation method. Henceforth, this method is used to infer formation of ssDNA.

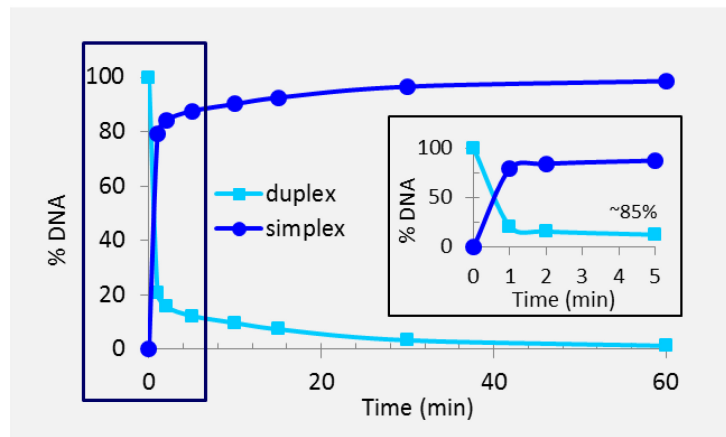
When both strands are non-phosphorylated (5'-mod/5'OH), there is no exonuclease digestion [Figure 2-13.b]. In cases of (5'-mod/5'P) dsDNA [Figure 2-13.c], the 5'-phosphoryl strand is hydrolyzed to yield ssDNA (lane 3), which is subsequently degraded by exonuclease I (lane 4). This shows that lambda exonuclease can be used for generation of 5'-fluorescent ssDNA. Similar results are obtained when biotin group is linked to 5'end of one strand while the complementary strand is phosphorylated.

2.3.7 Optimal digestion time of Lambda exonuclease

Studies to determine lambda exonuclease processivity show that it digests single-strand DNA and non-phosphorylated DNA at greatly reduced rates independent of the length of substrate. The amount of 5'hydroxyl dsDNA digested plateaus at 20% after first 10 minutes of incubation time [150]. A time series experiment is done to determine the conversion of dsDNA to ssDNA [Figure 2-14.a].



(a) Gel image



(b) Percent digestion of dsDNA

Figure 2-14 Optimal digestion time (lambda exonuclease)

Double-strand DNA with one strand 5'phosphorylated and other non-phosphorylated is digested with lambda exonuclease at multiple digestion time (1, 2, 5, 10, 15, 30 and 60 min, lane 6 -14). Formation of ssDNA after 15, 30 and 60 min is verified by digestion with exonuclease I (lane 2 – 4). In all three cases there is still some dsDNA detected.

Following, gel staining and image capture, the dsDNA band signal intensity in each lanes is analyzed with Image J software. Percentage of single-strand DNA formed at each time point is determined as follows, i) subtract background from each dsDNA band intensity, ii) calculate average band intensity of 'no-lambda exonuclease' controls and set as 100% (lane 1, 14), iii) normalize intensity of dsDNA band at lambda exonuclease time-points with respect to 'no-lambda exonuclease' controls to get % dsDNA, iv) calculate % ssDNA = 100 - % dsDNA. The presence of high molecular weight smear in the PCR and lambda exonuclease products as well as low separation resolution of dsDNA and ssDNA makes quantification of ssDNA inaccurate. Multiple repeats of the experiments gave the same result. Although this method, is not useful to quantify digestion of 5'-hydroxylated ssDNA, majority (>85 %) of duplex DNA is hydrolyzed within the first 5 min of the reaction [Figure 2-14.Inset].

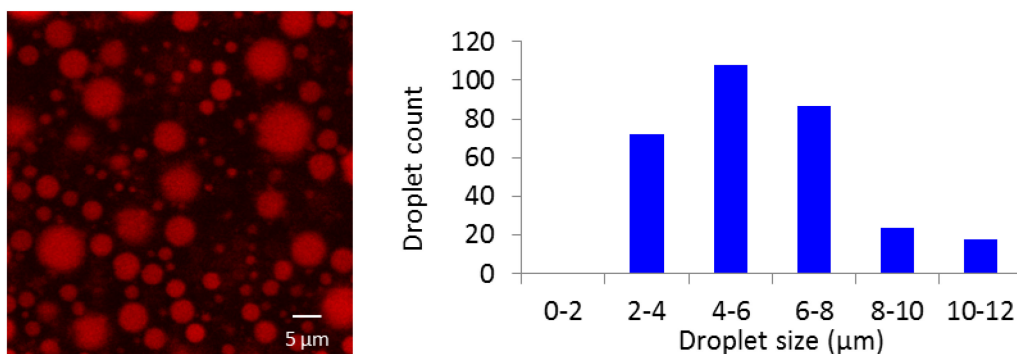
2.3.8 Amplification of microarray-derived DNA library

W/O emulsion droplet size

For emulsion PCR, thermostable non-ionic surfactant ABIL EM 90 (chemical name: cetyl dimethicone copolyol; HLB value 5; oil-soluble) is used to prepare the oil formulation. A trace amount of a water-soluble surfactant (Triton X-100; HLB value 13.4) is added to the oil to enhance the adsorbability of ABIL EM 90 on the surface of aqueous droplet. This reduces the interfacial tension between oil and

water droplets , increasing droplet stability and giving a narrower size distribution of droplets [151].

We measure the size of the emulsion droplets by confocal microscopy after PCR amplification. W/O emulsions with 4 parts oil to 1 part water are prepared with addition of 0.1 mM fluorescein to the aqueous phase. Confocal images are acquired along Z-axis at different XY planes. The emulsion droplets have size ranging between 2 – 12 μm , with >85% droplets distributed between 2 – 8 μm [Figure 2-15].



(a) Confocal image

(b) Emulsion droplet size distribution

Figure 2-15 Measurement of droplet size

The droplet size across multiple experiments gave similar spread in distribution, though number of droplets in each bin kept fluctuating. Comparing with droplet distributions obtained using same oil-surfactant mix a droplet size of 5 μ m is used to estimate 1 template per droplet concentration [95]. From the droplet diameter and volume of aqueous phase, one can calculate the starting template concentration to achieve theoretically one template per droplet [Equation 2. 1]. For example, 100 μ l PCR (aqueous) gives approximately 1.5 billion droplets with 5 μ m diameter.

$$V * (10)^9 = n * \frac{4}{3} * \pi * \left(\frac{d}{2}\right)^3$$

Equation 2. 1

Where,

V = volume of PCR aqueous phase (μ l)

n = number of droplets

d = droplet diameter (μ m)

Now, the emulsion droplet size and stability depend on several factors; i) input energy during emulsion formation, ii) duration of energy, ii) volume of oil and aqueous phases, iii) type of oil, iv) viscosity of both phases [152], v) interfacial tension, and vi) pH of aqueous phase [153]. Scientists have achieved tighter distributions of nanometer size droplets (100 - 300 nm \pm 10 nm) using ultrasonic energy to successful perform emulsion PCR, but they conclude that the large amount of PCR enzyme required for every droplet make the reaction cost prohibitive [146].

Oligonucleotide library concentration measurement

Microarray synthesized oligonucleotides released from the array surface have a large fraction of oligonucleotides of random size. These undesired oligonucleotides bias any absorbance based quantification of desired oligonucleotide and cannot be removed using conventional spin-column purification methods.

For emulsion PCR to be effective (i.e. no chimeric DNA products) the concentration of templates per droplet has to be less than or equal to one. To estimate the concentration of correct oligonucleotides, a real-time PCR dilution series calibration or standard curve is made using a commercial oligonucleotide with a) length (± 5 nt) similar to the DNA library and b) same universal primer binding sequences. The amplification plot and standard curve are shown in [Figure 2-16].

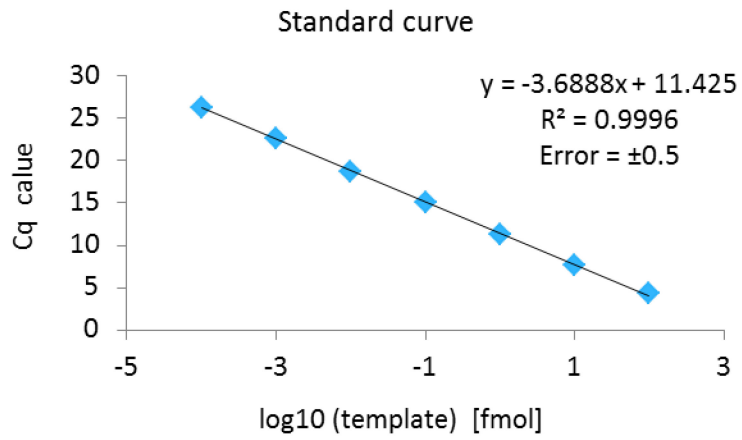
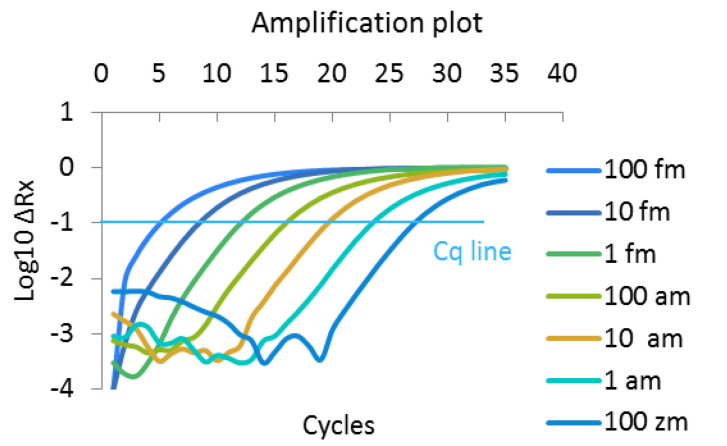


Figure 2-16 Real-time PCR – Standard curve

For reproducibility, two independent template dilution series are made starting from 10 pmol to 1 zmol. Each dilution series is amplified twice by real-time PCR with a separate master mix to give quadruplicate standard curves. There are non-specific products in high (>100 fmol) and low (<100 zmol) template concentration tubes and this were discarded for analysis. The 4 C_q values (i.e. the cutoff threshold at the exponential PCR phase) at each template concentration are averaged and plotted against the template concentration to give calibration curve [Figure 2-16]. The error in template concentration measurement ranges from ±0.1 to ±0.5 PCR cycles, corresponding to 1.1- to 1.4- fold error in predicting template concentration. The standard curve is used to derive an equation to calculate the DNA concentration from C_q value [Equation 2. 2].

$$DNA = (Cq - 11.425)/-3.689$$

Equation 2. 2

Where,

DNA = DNA concentration (fmol)

C_q = real-time PCR threshold value (exponential phase of PCR)

The efficiency of the PCR reaction is 86.67% ($E=10^{-1/\text{slope}}$) and the curve is valid for estimating DNA concentration between 100 fmol (10^{-15} ; 60 billion molecules) to 100 zmol (10^{-21} , 6000 molecules). The oligonucleotide library released from the chip surface is amplified in duplicate at 3 different dilutions (10 x dilutions; 1 ng – 10 and the corresponding C_q values are used to calculate the concentration of correct length DNA. The amount of PCR amplifiable product estimated using standard curve is 30- to 100- fold less than using spectrophotometer. 100 µl PCR mix forms ~1.53 billion emulsion droplets,

requiring 2.6 fmol of template to achieve 1 template per droplet. With 100 – fold less template i.e. 0.026 fmol, this would give only 0.01 template per droplet (i.e. 1 template / 100 droplet) corresponding to only 1% droplets with template. For the 30- fold less template, only 3% droplets have template.

Comparison of PCR methods and effect of buffer composition

The goal of this experiment is to show that emulsion PCR preserves library complexity. Chip-derived oligonucleotide library (117 bp) is amplified by both conventional PCR (convPCR) and emulsion PCR (emPCR) using two different buffers compositions. A common 150 μ l PCR master mix is used for both PCR, where 50 μ l is used convPCR and the rest for emPCR. 10 % of products are run on 2 % TBE agarose gel with 1 μ g of 25 bp ladder and ‘no template’ conventional PCR controls [Figure 2-17].

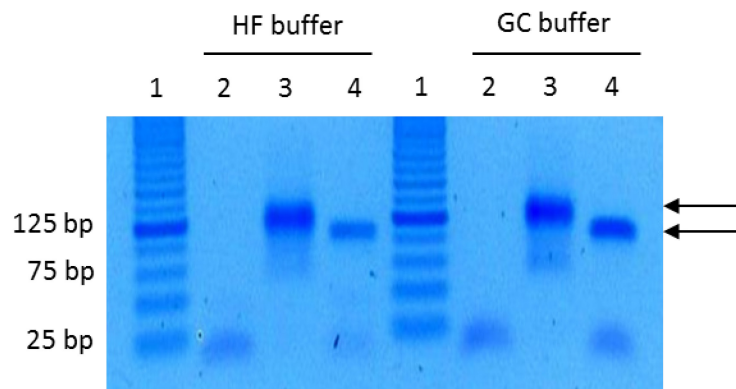


Figure 2-17 Emulsion PCR versus Conventional PCR

From gel image, comparing only the emPCR (lane 4, 8) and convPCR (lane 3, 7) products we see that emPCR results in sharp DNA band between 100 bp and 125 bp (expected size 113 bp, bottom arrow) than convPCR, where there is a smear of products between 75 – 125 bp. Also the amplicons size is different between the two PCR methods. In convPCR the amplicons are bigger than the expected size (top arrow, >125 bp), indicating formation of PCR artifacts. Comparing the two buffers, there is more amount of full length product when using buffer for GC rich templates (lane 8) than improving the fidelity of the enzyme (lane 4). The use of appropriate buffer will depend on the complexity of DNA library and specific downstream application. Finally, not all PCR primers (25 bp) are consumed in the emPCR (lane 2, 4). This indicates that there are still some droplets that are empty, despite adjusting template concentration to get 1 template per droplet. This is due to the fact that the dye used to quantify concentration of oligonucleotide library in method above binds to any non-specific products, giving a false higher concentration.

Finally, the amount and distribution of non-specific smear band during conventional PCR depends on the composition of the DNA library. DNA libraries with highly similar sequences show more amount of non-specific product both above and below the desired DNA band. Similar to all PCR reactions, emulsion PCR reaction has to be optimized with regards to annealing temperature and magnesium chloride concentration to only get desired product size.

Importance of 1 template per droplet

In our lab, we have shown that increasing the template concentration has undesirable effects on the DNA library composition. Here, a DNA library containing only mutant sequences of wild -type (WT) protein (no wild-type sequence), when amplified using 40 ng of template (roughly 3 - 4 templates / droplet) by emulsion

PCR, followed by cloning in E.coli and antibiotic selection, upon sequencing gave roughly 57% WT clones. When the same library is amplified at lower template concentration (10 ng, 1 template / droplet), the number of WT clones sequenced were reduced to only 2.5%. [10]. This demonstrates the requirement of < 1 template per droplet and power of emulsion PCR to reduce amplification artifacts.

From oligonucleotide array probes to ssDNA library

PCR amplification requires the presence of primer binding sites at each end of the oligonucleotides and will produce a double stranded amplicon. Both of these are undesirable for most other downstream applications [154-156]. The process to make ssDNA is shown in Figure 2-18.a. The microarray-derived DNA library obtained from the chip is processed through multiple steps to obtain high quality ssDNA as follows; i) quantification by real-time PCR standard curve, ii) emulsion PCR, iii) restriction digestion to remove primer binding sequence, and iv) lambda exonuclease digestion. Both PCR and restriction digest products are purified on commercially available silica-based spin columns.

Equal concentrations of the products are run on native acrylamide gel [Figure 2-18.b]. Here, 25 bp DNA ladder is run as size marker (lane 1, 6). Spin-column purified PCR amplicons (lane 2; 120 bp) are digested with restriction enzyme (MboII) to give two smaller DNA bands (lane 3; 90 bp and 25bp). The 100 bp dsDNA is the desired sequence-specific primer and 25 bp dsDNA is primer 1 [Figure 2-18.a]. Lane 4 shows hydrolysis of the digested products with lambda exonuclease to form ssDNA (95 bp), whose formation is confirmed by exonuclease I digestion (lane 5, blank). The ssDNA (lane 3, near 150 bp dsDNA band) runs slower than 120 bp dsDNA (lane 2, PCR product). This poses a challenge to purify the ssDNA from the gel. The need for purification is explained below.

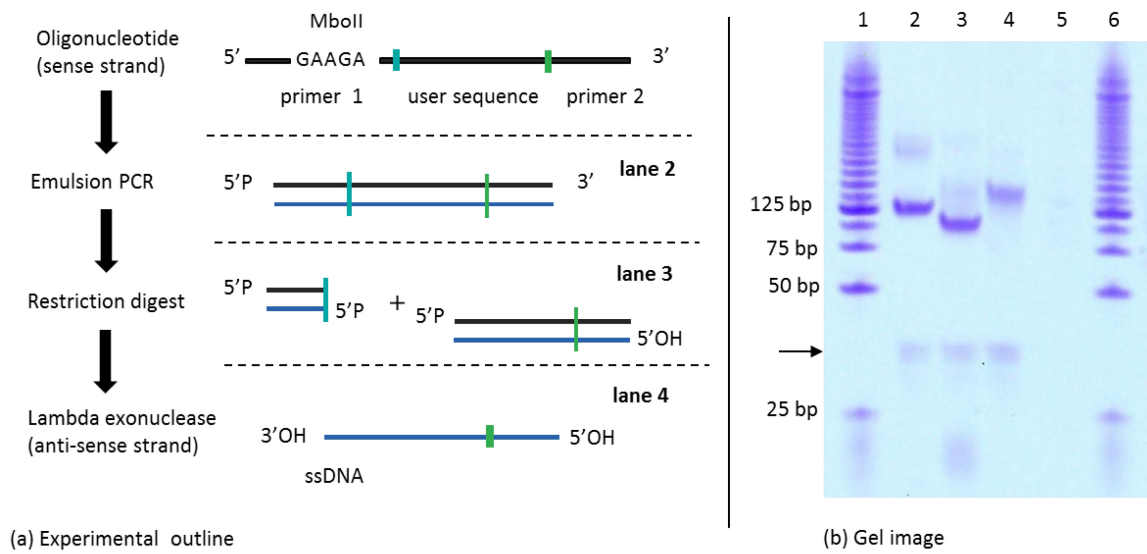


Figure 2-18 Lambda exonuclease - Oligonucleotide library to ssDNA

There is some primer-dimer formation during PCR amplification (lane 2, 3; arrow) which become single-stranded (lane 4; arrow) and is subsequently removed by exonuclease I (lane 5; arrow). These short ssDNA fragments may take part in downstream applications and have to be removed. Attempts to remove them by optimizing PCR conditions or using spin-column purification steps after both PCR and restriction digest give variable results. To get a sharp cut off, such that there is no contamination from non-restriction enzyme digested PCR product and primer-dimers, the ssDNA library is purified on polyacrylamide gel. This results in > 50% loss of desired ssDNA.

During the course of multiple experiments, we observed that there is some non-restriction digested dsDNA and also the hydrolysis of 5'OH/5'P dsDNA gave highly variable amounts of ssDNA for digestion time >30 min, and for <30 min results in unconverted dsDNA. The experimental design is modified to differentiate between dsDNA and ssDNA products during gel purification and get more quantity of single-strand DNA. The library is amplified by PCR such that the undesired strand has 5'phosphoryl group and desired strand has 5'-modification (5'-biotin or phosphorothioate backbone). This has two advantages; i) lambda exonuclease hydrolyze 5-phosphoryl strand of any non-restriction digested dsDNA to form ssDNA, and ii) the lambda exonuclease reaction can be performed for longer time period to completely hydrolyze the 5'phosphorylated strand of dsDNA without reducing ssDNA concentration.

Quantification of ssDNA produced by lambda exonuclease method

In a typical reaction, 200 µl of aqueous phase is mixed with 4 parts of oil-surfactant formulation to form emulsion for PCR amplification. This gives approximately 3.05×10^9 w/o droplets (average droplet diameter 5 µm) which

equates to approximately 5 fmol (0.005 pmol) of template to get 1 template per droplet. The average yield of emulsion PCR (30 cycles) is 50 pmol products. As the amount of ssDNA formed by lambda exonuclease method is directly one to one correlated to starting dsDNA concentration, we expect maximum 50 pmol of ssDNA. The final working concentration of ssDNA library after gel purification step is 14 pmol i.e. a recovery of 28%. The amount of single-strand DNA obtained is not enough to prime first strand cDNA synthesis discussed in Chapter 3.

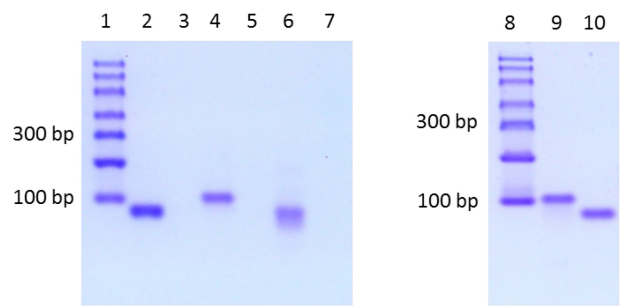
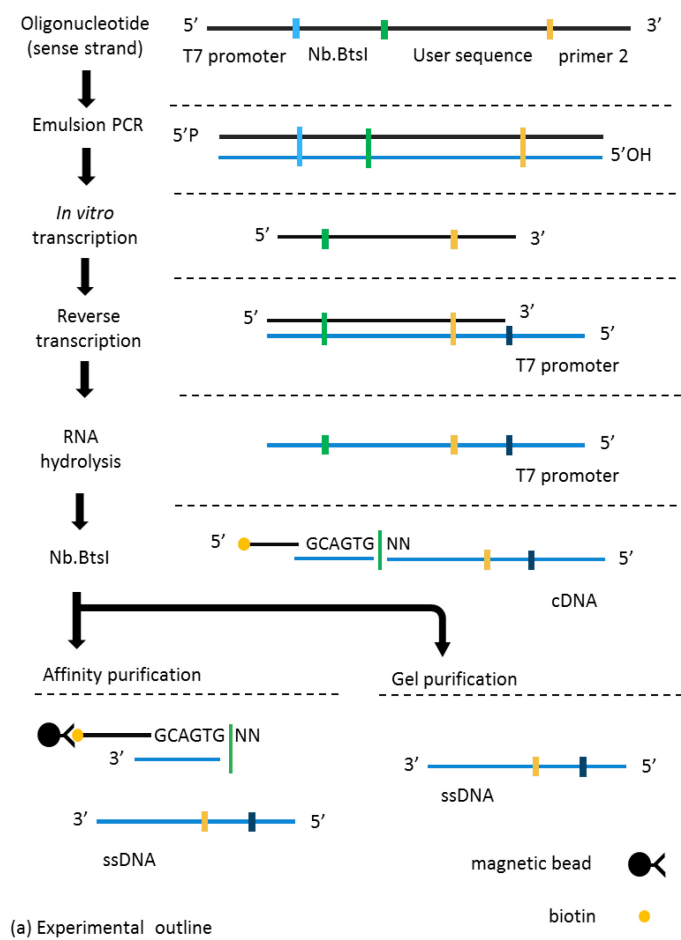
Also, the amplification factor for emulsion PCR is 10,000 x which is 100- to 500- fold less than conventional PCR (typical amplification factor > million fold). A simple method to increase yields would be to increase the reagents concentrations and perform more number of PCR cycles; this is not feasible as beyond 35 cycles the emulsion begins to breakdown and there is formation of non-specific products. Also, during the course of our experiments we do not see much yield difference between 30 and 35 cycles of PCR and in general keep the reaction to 30 cycles to prevent formation of non-specific products. Another drawback of emulsion PCR is that the primary limiting reagent is the TAQ polymerase [146]. Increasing concentration of polymerase leads to formation of non-specific PCR products and we already use double the recommended enzyme concentration to increase yield. Thus, emulsion PCR is great technique to maintain the complexity of chip-derived oligonucleotide libraries during the PCR amplification step, but it has yield limitations.

2.3.9 In vitro transcription and reverse transcription (IVT-RT)

Another method to increase the amount of ssDNA is to transcribe the emulsion PCR products using T7 RNA polymerase. This has several advantages; i) T7 polymerase will copy each DNA template >1000 times, ii) it produces single-stranded RNA molecules, which display a different spectrum of sensitivity to

enzymatic or chemical reagents, iii) RNA molecules can be copied back into DNA molecules by reverse transcription [157].

The experimental design is presented in Figure 2-19.a. PCR amplification of oligonucleotide library is done with T7 promoter forward primer and universal reverse primer (UP2) pair. T7 phage polymerase is used to transcribe the amplicons to form RNA library which is primed with sequence-specific primer containing a T7 promoter. The resultant cDNA library is hybridized with complementary Nb.BtsI (5'-GGGAGAAGTCACGCAGTG|NN- 3'; recognition sequence underlined) primer to remove 3'end PBS. Nb.BtsI only cuts the bottom strand (i.e. cDNA). The digestion products are purified on polyacrylamide gel to give ssDNA library. The design below shows an alternative method to purify the desired ssDNA from the cut primer binding sequence using biotin-streptavidin magnetic beads. The process remains the same except that the complementary Nb.BtsI (5'-biotin-GGGAGAAGTCACGCAGTG|NN- 3'; recognition sequence underlined) primer has 5'-biotin moiety. The design also shows the scheme to use T7 promoter at both ends; first to transcribe the dsDNA library and then to prepare ssDNA library with 5' T7 promoter for downstream reactions.



(b) Affinity purification (b) Gel purification

Figure 2-19 Oligonucleotide library – IVT-RT method

Figure 2-19.b shows the products of IVT-RT method (1% agarose gel). Lane 1, 8 are 100 bp ssRNA ladders. T7 RNA polymerase transcription of emulsion PCR product gives a sharp 95 nucleotide (nt) RNA band (lane 2, 9). The RNA is reverse transcribed to make 113 nt ssDNA copies (lane 4). Now, post reverse transcription there are excess RT primers (data not shown). Two methods are tried to remove these excess primers; i) exonuclease I hydrolysis of RT primers before RNA hydrolysis step, and ii) spin column purification of cDNA (~200 nt ssDNA cutoff threshold). We observe that exonuclease I hydrolyzes some of the cDNA in the RNA:DNA hybrids (data not shown), while the purification on spin-column leads to loss of cDNA and gives inconsistent removal of RT primers. As both methods do not work well, the cDNA is purified on size-exclusion columns for buffer exchange. This step does not completely remove the RT primer.

For future investigation, two methods are compared to purify ssDNA after the removal of primer binding sequence with Nb.BtsI nicking endonuclease. Lane 6 shows formation to desired ssDNA (95 nt) after 4 rounds of purification on magnetic beads. Here, we see an unexpected band below the desired fragment. This band is not seen when another enzyme (Nb.BspQI) is used to cut the primer binding sequence (data not shown). Lane 10, shows the same ssDNA after PAGE purification. Purification of ssDNA with streptavidin magnetic beads helps recover more amount of ssDNA (60 – 65 % ssDNA recovered). This is a two-fold improvement over PAGE purification method which gives only ~30% ssDNA. Finally, lane 3 is alkaline hydrolysis of RNA product and lane 5 (113 nt ssDNA) and lane 7 (95 nt ssDNA) are exonucleolytic hydrolysis of cDNA to show formation of correct products.

Quantification of ssDNA produced by IVT-RT method

In a typical reaction, *in vitro* transcription of 4 pmol of dsDNA produces approximately 3400 pmol ssRNA after purification (amplification factor ~850). The RT products are purified by size exclusion chromatography to give anywhere from 1700 – 2040 pmol of cDNA (50 - 60 % amount produced). After removal of 3' end primer binding sequence and PAGE gel purification, the final concentration of ssDNA is anywhere between 475 pmol – 570 pmol. This is 30- to 38- fold more than the amount of ssDNA generated using lambda exonuclease method.

2.4 Conclusion

In this chapter, we tested multiple methods to generate large quantities of single-stranded probes starting from synthesized oligonucleotides released from microarray.

Exponential amplification in femtoliter volume w/o emulsion reactors helps to preserve library complexity and overall abundance. Some of the parameters important to the success of emulsion PCR are i) size of template, and ii) template concentration. The knowledge of droplet size helps to determine the size of template that can be amplified and the number of emulsion droplets formed. The template size should be less than the droplet size for the DNA molecule to elongate during PCR amplification to get any products. Besides amplifying ssDNA libraries, we also demonstrated the amplification of fragmented genomic DNA by emulsion PCR. This will prevent any PCR bias during sample preparation of fragmented genomic DNA for 'next-generation' sequencing.

From the droplet size and aqueous phase volume, one can calculate the number of droplets to get 1 template per droplet. There is formation of chimeric DNA molecules concentrations when there is greater than 1 template per droplet. On the other hand, too less a concentration significantly reduces the yield and waste reagents. A 10- fold decrease in template concentration (i.e. 0.1 templates per droplet) leads to 90% of droplets being empty. This is particularly true for microarray libraries, where truncated sequences without any of the primer binding sequences results in 30- to 100-fold higher absorbance readings. Thus, real-time PCR standard curve is used as a tool to determine the concentration of amplifiable DNA molecules in the oligonucleotide library.

Two methods are used to convert dsDNA to get ssDNA molecules. The first method uses lambda exonuclease to preferentially hydrolyze 5' phosphorylated DNA strand in one-step process. Lambda exonuclease is able to digest >85% dsDNA to ssDNA within the first 10 min of the reaction. The digestion of 5'-hydroxylated strand is minimized by adding protective groups such as 5'-biotin or a phosphorothioate backbone. This method is good to prepare 5'-fluorescent probes for hybridization-based assays in short period of time (< 2h). However, when used to convert emulsion PCR products, it does not produce enough of ssDNA for priming reactions. In the second method, the emulsion PCR amplicons are transcribed to produce >850-fold RNA, followed by reverse transcription to get ssDNA. Here, the ssDNA produced is not limited by emulsion PCR product concentration. One emulsion PCR gives enough material to perform 12 IVT reactions (4 pmol per reaction), which has the potential to produce anywhere from 5700 – 6800 pmol of single-strand DNA. This is ~ 400 times greater than emulsion PCR – lambda exonuclease method to produce ssDNA. Now, 5700 – 6800 pmol quantity of ssDNA is obtained after PAGE purification. The use of streptavidin-coated magnetic beads to remove primer binding will result in 10400 – 13600 pmol of ssDNA (i.e. $10400 \text{ pmol} \times 330 \text{ g/mol} \times 95 \text{ nt} = 326 \text{ }\mu\text{g}$).

Two approaches are used to remove the 3' end PBS blocking the user (gene-specific sequence). Both approaches use type II restriction enzymes that cut outside their recognition sequence, but differ on their preference of template. i.e. dsDNA or ssDNA. For double-stranded templates, the 3' end PBS is first removed followed by hydrolysis of undesired strand, whereas for single-stranded templates, the 3' end PBS is hybridized with a complementary sequence, followed by removal with nicking endonuclease.

In conclusion, we have developed and validated a robust process to make large-quantities of high-quality single-stranded DNA starting from chip-derived oligonucleotide library.

CHAPTER 3

PRIMING REVERSE TRANSCRIPTION WITH SEQUENCE-SPECIFIC PRIMER LIBRARIES

3.1 Introduction

Reverse transcription mediated by reverse transcriptase enzyme is the cornerstone of molecular biology investigations as it catalyzes the synthesis of complementary DNA (cDNA) from messenger RNA (mRNA). [158, 159]. Three different strategies are used to prime cDNA synthesis; i) oligo(dT), ii) random primers or iii) gene-specific primers. Each of the methods differs significantly with respect to target specificity, capture sensitivity, transcript coverage, and cDNA yield. Many factors are taken into account before selecting particular primer. They are size (1 -100 cells) and quality (degraded RNA) of the sample [160-162], representative amount of genes in the sample (i.e. low abundance or rare mRNA), and the amount of background RNA i.e. total RNA [163].

Oligo(dT) primers are DNA fragments containing multiple thymidine residues only. For eukaryotic organisms, they are used to reverse transcribe polyadenylated mRNA molecules within the total RNA population. Their specificity is increased by addition of one or two extra nucleotides at 3' end, as they anneal to the 5' end of the poly(A) at the junction with mRNA limiting the portion of poly(A) tail reverse transcribed [164]. However, as they always initiate reverse

transcription at the 3' end of the transcript, any RNA secondary structure may lead to incomplete cDNA synthesis and poor representation of 5' end (3' bias) [165]. Another drawback of oligo(dT) primers is that they prime at internal stretches of A residues in the mRNA resulting in further synthesis of 3'-end truncated cDNA products [166], though the substitution of some of the thymidine residues of oligo(dT) primer with base analog 3-nitropyrrole has been shown to decrease the frequency of false priming [167]. Now, there is a significant portion of transcripts that have no poly(A) tail (e.g. histones or viral RNA) or have shortened poly(A) tails [168], while other transcripts are found in both poly(A)+ and poly(A)- forms within the population [169]. Use of oligo(dT) primers alone for detection or expression analysis of these RNA will give irreproducible error results. Oligo(dT) primer cannot be used when dealing with fragmented RNA samples, such as those derived from formalin-fixed paraffin-embedded (FFPE) tissue specimens [165].

Random primers For RNA lacking poly-A tail, the reverse transcription is usually primed with short random oligonucleotides (usually 6 to 9-mer) or with oligo(dT)_n after adding a poly-A tail in vitro to the RNA with a poly-A polymerase. Unlike oligo(dT) priming method, random primers initiate cDNA synthesis at multiple points in the RNA template to produce more than one cDNA transcript per original target. Random primers are used when high yields of cDNA are required and for low quality RNA samples. Also they are not affected by RNA secondary structure. The major inconvenience of using random primers is that they also prime reverse transcription on other types of RNA like ribosomal and tRNA which represent more than 90% of the total RNA, and may not detect rare mRNA transcripts. Second, random hexamers (6 bases) show a 19- fold increase in the amount of RNA calculated compared to sequence-specific primer [170] and over-represent the 5' end of transcripts. The latter effect has been reduced to some

degree by using pentadecamer random primers [171]. Finally, the ratio of random primers to mRNA concentration has to be optimized to prevent formation short cDNA molecules [157].

Gene-specific primers anneal to complementary regions of the target mRNA with high specificity and give accurate mRNA quantification [170]. They allow better discrimination between closely related gene sequences by having the 3' end of the oligonucleotide targeting sequences non-conserved between genes since only primers perfectly hybridized at their 3' end can be extended during RT. They can be used for applications such as; i) target specific transcripts in complex mixtures, like viral or bacterial RNA contaminated by host cell RNA [16], ii) identify and map exon-exon junctions and splice variants [172, 173], iii) detect rare or low abundance transcripts [174], and iv) for quantitative RT-PCR analysis. One drawback of gene-specific primer requires a new cDNA synthesis reaction for each gene studied, making them less than optimal for processing limited samples.

3.1.1 Sequence-specific primer libraries for reverse transcription

The emergence of next-generation sequencing, especially massively parallel complementary DNA (cDNA) sequencing (RNA-seq) is poised to make valuable impact across biosciences. Some applications of RNA-seq include *de novo* sequencing, targeted sequencing, gene expression analysis, and simultaneous analysis of transcripts from multiple cells, tissues or organisms. Unlike DNA microarrays, RNA-seq technology has the potential to give absolute quantification of transcript levels. It does not suffer from hybridization related array issues such poor detection sensitivity of low abundance transcripts, inability to differentiate between closely related sequences, and variable target-probe hybridization efficiency [175].

As the sequence of transcript is read during RNA-seq, it can accurately detect transcript isoforms, gene fusions, and map exon junctions [175, 176].

Current methods to prepare RNA samples for 'next-generation' sequencing (NGS) have lots of intermediary steps before the sample is ready to go on the sequencing machine [177]. These introduce bias in the transcript level leading to misinterpretation of data. Biases can be introduced at RNA enrichment and fragmentation, adaptor ligation, PCR amplification and size-selection gel purification steps [177-180]. The priming of first reverse transcription reaction is usually done with oligo(dT) and random primers, and is shown to give 3' or 5' transcript coverage bias [177, 181].

The motivation of using a library of sequence-specific primers for reverse transcription is to give accurate RNA quantification and improve the detection sensitivity. The use of sequence-specific primers tagged with forward sequencing primer does not require pre-reverse transcription sample preparation steps (mRNA enrichment and fragmentation, polyadenylation or adaptor ligation). Post reverse transcription, random nonamer (N₉) with upstream reverse sequencing primer can be used to synthesize second strand cDNA followed by direct sequencing with 'next-generation' sequencing platforms. The advantage of using sequence-specific primers is that it can be used to target subsets of RNA (e.g. metabolic pathway, rare mRNA, mRNA splice variants) across large number of samples [182]. This is achieved by incorporating identification tags (barcodes) for each sample (or target). It has been proposed to design the tags using hamming codes to identify and correct sequencing errors [183].

3.1.2 Background

The presence of reverse transcriptase was reported in 1970, whereby certain viruses replicated their genomes via a DNA intermediate [135, 136]. Retroviral RT enzyme has three basic enzymatic activities; i) reverse transcriptase – synthesis of DNA from an RNA template, ii) DNA polymerase – synthesis of DNA from a DNA template and iii) RNase H activity – digestion of the RNA strand from an RNA:DNA hybrid to leave ssDNA [184]. In case of retroviral genome replication, a transfer RNA (tRNA) primer associated with viral genome is extended to initiate minus-strand DNA synthesis [185, 186], while plus-strand DNA synthesis is initiated from an RNase H-resistant polypurine tract of genomic RNA that remains tightly bound to the newly synthesized minus-strand DNA [187, 188]. Retroviruses have played a big part in the shaping of extant genomes of complex organisms [189, 190].

Catalytic properties of reverse transcriptase

In molecular biology, reverse transcriptases are used to create cDNA libraries from mRNA, and along with other enzymes, allow cloning, sequencing, gene expression analysis and characterization of RNA [191]. Commercially available reverse transcriptases are derived from avian myeloblastosis virus (AMV), moloney murine leukemia virus (MMLV), and human immunodeficiency virus reverse transcriptase (HIV) either by purification from the virus or expressing them in E.Coli.

DNA polymerase activity - Unlike other DNA polymerases, the RT enzyme can use either RNA or DNA to prime DNA synthesis from an RNA or DNA template [192, 193]. The RNA-dependent DNA polymerase activity (RdDp) is required for synthesis cDNA while the DNA-dependent DNA polymerase activity (DdDp) enables synthesis of dsDNA [194-197]. There exists a high degree of overlap between the

two catalytic domains i.e. the two activities are inseparable and may share a common active site [157, 198, 199] .

RNase H activity - is required for three distinct steps during reverse transcription; i) generation of the plus strand RNA primer, ii) removal of the plus and minus strand primers from the nascent DNA strands, and iii) degradation of the RNA genome following minus strand DNA synthesis to make the DNA template accessible for plus strand synthesis [200, 201]. RNase H activity is independent of the polymerase activity. During the initiation step of cDNA synthesis, binding of DNA primer to mRNA leads to competition between primer extension (polymerase) and RNA degradation (RNase H) functions of RT [157]. This limits the maximum number of priming events and cDNA yield [202]. Further, the presence of degradative activity is responsible for decreased efficiency in synthesis of long or full length cDNA [188, 203]. Commercially available, RT enzymes (e.g. MMLV) are mutated to be devoid of RNase H- activity (MMLV H-), while maintain their polymerase activity to improve the efficiency of cDNA synthesis [204].

Template-switching or retroviral recombination activity is exhibited by RT enzymes whereby during cDNA (minus-strand) synthesis the enzyme jumps from one template to another leading to formation of chimeric cDNA molecules [205, 206]. The rate of retroviral minus-strand recombination is reported to be 20% [207, 208] and is highly dependent on two linked factors; i) temperature, and ii) RNA secondary structure – reduces with increase in temperature. The presence of intrinsic secondary structures in the RNA templates leads to pausing of the enzyme=nascent cDNA complex [209]. This has two negative effects; i) more degradation of RNA template due to RNase H activity resulting in formation of truncated cDNA [210], and ii) increased probability of mutations at pause sites [211]. Also, stoppage of RT at pause sites leads to dissociation of the enzyme-

nascent cDNA complex from the RNA template and increases the probability of it being transferred to other homologous template strands to synthesize chimeric cDNA [212-214].

Terminal nucleotidyl transferase-like activity (TdT) - RT enzymes adds a few extra non-template nucleotides beyond the 5' termini of both DNA and RNA templates [215, 216]. More than 30% of molecules have one to four extra nucleotides added beyond the template termini. The TdT activity is undesirable as it adds mutations to the nascent cDNA during the strand transfer process [217]. During in vitro reactions, the enzyme exhibits temperature and time dependent TdT-like activity specific to either the DNA or RNA strand of double-stranded substrates with preference for addition of purine nucleotides ($dA > dG \gg dT \geq dC$) [218].

3.1.3 Insights from reverse transcription-polymerase chain reaction (RT-PCR)

In vitro reverse transcription - polymerase chain reaction (RT-PCR) is the gold standard for detecting low and medium abundance mRNA over a wide dynamic range. It has permeated many area of research such as clinical diagnostics, food safety, biodefense, gene expression analysis to name a few. Besides the uncertainty associated with PCR step, It has also been observed that RT step is highly uncertain as it introduces errors resulting from RNA secondary and tertiary structure, priming strategy, abundance of gene detected, and the properties of reverse transcriptase [219-224].

The RT enzyme lacks 3' - 5' exonuclease proof reading ability leading to error prone DNA synthesis. These errors are then amplified exponentially during PCR. Amongst the three RT enzymes, HIV exhibits the highest error rate with misincorporation frequency of 1 in 6,900 on RNA template and 1 in 5,900 on DNA

template making it unsuitable for molecular biology applications [225, 226]. The misincorporation frequency of MMLV and AMV RT on DNA template is 1 in 29,000 and 1 in 17,000 respectively [225, 227]. Now, MMLV and MMLV RNase H- show similar misincorporation frequency on RNA templates of 1 in 37,000 [204, 225]. Recently, there is one commercially available RNase H- deficient MMLV RT mutant that has been engineered with error rate as low as 1.6×10^{-5} /base (i.e. misincorporation frequency of 1 in 62000).

RT enzymes do not have strand displacement activity and are unable to transcribe through secondary structure leading to formation of truncated cDNA. A number of methods have been reported to increase the processivity of RTs. These methods are based on either relaxation of the RNA template before starting the RT reaction (e.g. heat denaturation, pretreatment of RNA with methylmercury hydroxide, 1% DMSO) or carrying out the reaction at high temperature (above 45°C) in the presence of a stabilizer (e.g. trehalose) [228, 229]. Another method to improve cDNA yields particular at low temperature (37°C) is addition of accessory proteins (e.g. nucleic acid binding protein). The use of nucleic acid chaperones has been reported to increase the efficiency of HIV RT and commercially available superscript II MMLV H- RT, though they are not widely used in laboratory as their effect is highly variable with other commercial RNase H- RT [203].

Another drawback of reverse transcriptase, especially RNase H- mutants is that they inhibit PCR amplification. The enzyme remains bound to the RNA:DNA template making it inaccessible to any DNA polymerase and have to removed or denatured directly after the RT reaction to reduce inhibitory and/or nuclease activities [230-232]. Thermostable RTs create even more problems as they have increased heat resistant and retain their activity even after heat inactivation [221, 233]. One method that helps to minimize RT-PCR inhibition is the dilution of the

cDNA products, but imprecise result in imprecise quantification of low copy number templates [234]. The alkaline hydrolysis of RNA and subsequent phenol extraction has been shown to relieve RT enzyme inhibitory effects [223].

Finally, the sensitivity and reproducibility of many commercially available RT enzymes has been investigated with respect to RT-PCR based mRNA quantification. Most RT enzymes consistently detect medium and highly expressed genes. Their sensitivity to lowly expressed genes and rare transcripts is highly variable and is dependent on many factors not limited to source of sample, integrity of RNA, concentration of other RNA (carrier molecules), gene detected, RT and PCR reaction buffers and conditions, addition of carrier proteins (T4 gene 32, RNase inhibitors) and purification method [221-224, 235]. Even across the enzymes that demonstrate high sensitivity there is variation with respect to reproducibility and overall amplification efficiency [223].

3.2 Experimental methods and protocols

3.2.1 One round aRNA amplification with 5'-end dangling gene-specific primer

TargetAmp 1 –round aRNA amplification kit (Epicenter Biotechnologies Inc.) is used according to manufacturer's protocol to make aRNA starting from mRNA using gene-specific primer. Briefly, 4 pmol of ACT1 mRNA is primed with 10 pmol of 71 bp long primer (25 bp gene-specific sequence). The first strand cDNA synthesis is done at 55°C for 45 min (5 µl) followed by enzyme (superscript III) inactivation at 75°C for 15 min. Next, 5 µl of second strand cDNA synthesis enzyme mix is added to the first strand reaction. The reaction is done at 65°C for 10 min and the enzyme is inactivated at 80°C for 10 min. Excess unhybridized primers are removed by incubation with cDNA finishing solution (proprietary formulation) at 37°C for 1 h, followed by enzyme inactivation at 80°C for 10 min. 5 µl of double-stranded DNA is transcribed for 4 h at 42°C with T7 RNA polymerase (25 µl). The template DNA is degraded with 1 µl DNase I at 37°C for 15 min. The DNase I is inactivated with one round of 25:24:1 phenol: chloroform: isoamyl alcohol (PCIA), pH 4.7 extraction and after purification on column (RNeasy), the RNA is eluted in 50 µl nuclease water. 10 µl aliquot of aRNA is degraded with 0.43 M sodium hydroxide (65°C for 15 min). Also, 0.5 µl aliquot of first strand cDNA is amplified with 0.5 µM forward and reverse primers by 20 cycles of PCR (20 µl).

3.2.2 Removal of excess sequence-specific primers

Silica-based filter spin column purification

The protocol outlined above is used to test the removal of double-strand long primer with one modification, that is, the second strand cDNA synthesis reaction is

cleaned up on silica-based spin chromatography columns (PCR QiaQuick purification kit, Qiagen Inc.)

Selective digestion of single-strand DNA

2 pmol of ssDNA (66 bp) is digested with different concentrations of exonuclease I (20 U, 40 U; Epicenter Biotechnologies Inc.) and single strand binding (SSB) protein (0, 0.2, 4 and 6 µg/µl; Epicenter Biotechnologies Inc.). The reaction (50 µl) is done at 37°C for 1 h, followed by enzyme inactivation at 80°C for 15 min. The amount of undigested ssDNA is detected by real time PCR. 5 µl of exonuclease reaction product is amplified by real-time PCR (50 µl), followed by melt curve analysis.

Size-dependent fractionation of DNA on magnetic beads

1 µg of 100 bp DNA ladder (New England Biolabs) is fractionated with magnetic carboxylate beads (Agencourt AMPureXP bead technology) in solution containing 1.25 M sodium chloride and different concentrations of PEG 8000 (7 % - 11 %, 13 %). 10 µl of DNA ladder is mixed by pipetting (10 times) with 1.8 x volume (18 µl) of beads. The precipitation is carried out at room temperature (~25°C) for 10 min under constant shaking. The bead bound and unbound fractions are separated by placing the tubes on magnetic beads. The solution is aspirated to remove the unbound fraction. The beads are washed twice with 2x volume of freshly prepared 70 % ethanol. The remaining alcohol is evaporated by heating at 37°C for 2 – 5 min. The bound fraction is eluted from magnetic beads with Tris-HCl (pH 8.0) buffer and transferred to a new 1.6 ml tube.

Next, the 100 bp DNA ladder is suspended in 10 µl first strand synthesis buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 3 µg actinomycin D)

and fractionated with carboxylate beads. The fractionation is performed in six independent solutions with different concentrations of sodium chloride (0.9 M and 1.25 M) and PEG 8000 (9.5 %, 10 % and 10.5 %).

3.2.3 Optimize RT conditions when using sequence-specific primers

Effect of hybridization temperature on reverse transcription

0.26 fmol of Gal 7 RNA is reverse transcribed using 100x excess long primers (1.3 fmol). Transcript capture followed by reverse transcription is performed at 4 different temperatures (37°C, 45°C, 50°C and 55°C). The target-primer hybridization is performed in 300 mM NaCl concentration for 30 min (10 µl) followed by doubling of reaction volume with first strand cDNA buffer solution (50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 0.5 mM deoxynucleotides, 200 U superscript III) and incubated for 90 min. The enzyme is inactivated at 75°C for 15 min. 5 µl of first strand cDNA is amplified by 30 cycles of real-time PCR (50 µl), followed by melt curve analysis.

Effect of primer concentration on reverse transcription

0.26 fmol of Gal 7 RNA is reverse transcribed with 500x to 0.5x excess long primers (dilution factor 10). Transcript capture is done in 300 mM MgCl₂ buffer solution containing 20 U RNase inhibitor at 45°C for 1 h (5µl) followed by doubling of reaction volume with first strand cDNA buffer solution (50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 0.5 mM deoxynucleotides, 100 U superscript III) and incubated for 60 min. The enzyme is inactivated at 75°C for 15 min. 5 µl of first strand cDNA is amplified by 35 cycles of real-time PCR (50 µl), followed by melt curve analysis. The first two cycles are done at primer annealing temperature of T_m - 10°C.

3.2.4 Detect differential expression of transcripts with sequence-specific primers

Reverse transcription with mouse sequence specific primers

15 µg of mouse testicle total RNA (Ambion Inc.) and yeast galactose total RNA is hybridized with 50 pmol of mouse sequence-specific primers in nuclease free water containing amino-allyl dUTP. The hybridization solution (12 µl) is heated at 65°C for 5 min, and slowly cooled (-0.05°C/min) down to 37°C. Next, first strand synthesis buffers and superscript III enzyme is added and the reaction (20 µl) is incubated over 2 h at 37°C (0.5 h), 45°C (0.5 h) and 50°C (1 h) respectively. The reaction is heated at 95°C for 5 min and immediately put on ice (2 min) to denature the RNA|cDNA hybrids and inactive reverse transcriptase. 0.43 M sodium hydroxide is added to degrade the RNA at 65°C for 5 min. The solution is neutralized with 0.43 M HCl. The reaction volume is brought to 100 µl to which are added 0.3M sodium acetate and 2.5 volume 100% ethanol for DNA precipitation. The amino-allyl cDNA (aa-cDNA) is labeled with alexafluor dye in the dark at room temperature following manufacturer's protocol (Invitrogen Inc.). The labeled cDNA products are not purified from unbound dye to prevent loss of labeled cDNA.

The cDNA products are dynamically hybridized to oligonucleotide microarrays (5k probes) in hybridization oven for 20 h at 35°C (2 – 3 rpm rotational speed). The hybridization is done in 6x SSPE buffer (0.9 M NaCl), 0.1 µg/µl acetylated BSA, 0.1 µg/µl herring sperm DNA, 0.01 % Tween 20 and 15 % formamide solution (50 µl) using agilent gasket slides and hybridization chamber. Post hybridization the arrays are washed at room temperature with non-stringent (1x SSPE; two successive washes) and stringent (0.25x SSPE, one wash) buffers for 3 min each. The slide is briefly dipped in water to remove salts, spun dried for 10 s and scanned with Axon 4000B Fenepix scanner (Molecular devices Inc). The data is extracted using Genepix pro 4.0. software provided with the scanner.

Reverse transcription of single-strand DNA

Nuclease free water is substituted for total RNA in first strand cDNA synthesis reactions performed at three different temperatures (37°C, 45°C and 50°C) using the protocol above with few modifications, namely i) the aa-cDNA is purified with silica-based spin columns (Marligen Biosciences), ii) the 45°C cDNA column fractions are run on acrylamide gel and iii) the 50°C cDNA fractions are detected on oligonucleotide arrays.

10 pmol of single-stranded sequence specific primers are hydrolyzed with exonuclease I at 37°C (20 U enzyme) and 50°C (40 U enzyme) for 3 h (10 µl) and run with 45°C cDNA column fractions and 25 bp DNA ladder on 6 % denaturing acrylamide gel. The gel is run for 4 h at 120 V. The top plate is removed; the gel is gently washed with water and stained on bottom plate with 10x SYBR Gold (Invitrogen Inc.) in 10 ml Tris-HCl buffer for 20 min. The gel is visualized on UV illuminator and photographed with camera (orange filter).

3.2.5 Activity of reverse transcriptase in Actinomycin D

26 fmol of Gal7 transcript is hybridized with 10x long primer (Gal7_30bp_P4) in 0.3 M sodium phosphate buffer and 10 U of RNase inhibitor. The reaction (10 µl) is initially denatured at 65°C for 5 min, followed by 30 min incubation at 45°C and 5 min at 37°C. 5 µl aliquots of reaction are used for first strand cDNA synthesis with superscript II (200 U) in 50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂ buffer solution (20 µl) at 37°C for 2 h. The concentration of actinomycin D is varied from 0, 10, 50 – 3200 ng/µl (doubled after 50 ng/µl). Actinomycin D is removed by extraction with PCIA (pH 8.0). The cDNA is precipitated and the pellet suspended in 10 µl of nuclease free water. 5 µl of cDNA is amplified by real-time PCR (25 µl). 1/10th volume of amplicons (150 bp) are run on 1 % agarose gel for 45 min at 90 V.

Microarray study: Actinomycin D induced inhibition of DNA dependent DNA polymerase activity of reverse transcriptase

80 µg of yeast galactose total RNA is hybridized with 2 pmol of sequence-specific primer library (~4800 primers, 0.26 x excess primer). 20 µl of hybridization solution consists of 200 mM NaCl and 40 U Superase-In. The RNA is denatured at 65°C for 5 min, followed by target capture at 50°C for 60 min. The hybridization solution is equally split into 4 separate reactions. 3 of the aliquots (5 µl) are used directly for reverse transcription. The RT is done with superscript III enzyme in 20 µl volume in 1x first strand buffer, 10 mM DTT, 40 U RNasin inhibitor, 1x amino-allyl dUTP (0.5 mM of dA, dC,dG; 0.3 mM dT; 0.2 mM aa-dUTP) and 200 U superscript III for 90 min at 50°C. For reactions with actinomycin, 300 ng/µl ActD is added to the RT solution. After 90 min, the RT solution is added to directly to solution (30 µl) containing 40 U exonuclease I, 10 µg SSB protein, 40 U RNasin in 1x exonuclease I buffer (Epicentre biotechnologies Inc.) and incubated at 45°C for 60 min. The reaction is heated at 95°C for 5 min and immediately put on ice (2 min) to denature the RNA|cDNA hybrids and inactive reverse transcriptase. 0.43 M sodium hydroxide is added to degrade the RNA at 65°C for 5 min. The solution is neutralized with 0.43 M HCl. The reaction volume is brought to 100 µl to which are added 0.3M sodium acetate and 2.5 volume 100% ethanol for DNA precipitation. The cDNA pellet is suspended in 6 µl nuclease free water. 1 µl is used to detect GAL 1 gene by real-time PCR and the rest is labeled with alexafluor dye (AF555, Invitrogen Inc.) for detection on microarray.

PCR study: Actinomycin D effect on DNA dependent DNA polymerase activity of reverse transcriptase

20 µg of yeast glucose and galactose total RNA is hybridized with 0.5 pmol of yeast sequence-specific primers in solution containing 0.3 M NaCl and 10 U RNase

inhibitor (5 µl). The hybridization solution is heated at 65°C for 5 min, and cooled down to 45°C and incubated for 3 h, followed by brief incubation at 37°C for 20 min. Next, 40 U exonuclease I and 10 µg SSB protein is added to the reaction to hydrolyze unhybridized long primers in 67 mM glycine-KOH, 6.7 mM MgCl₂, 10 mM 2-mercaptoethanol buffer solution for 60 min (50 µl). For 45°C exonuclease digestion reaction, the 37°C hybridization step is omitted. 0.5 M NaCl is added to the solution and exonuclease is inactivated with PCIA, pH 8.0 extraction. Post precipitation, the RNA: long primer hybrids are suspended in 10 µl Tris-HCl, pH 8.0. To this we add 10 µl of first strand synthesis solution (50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂ and 10 mM DTT) containing 6 µg actinomycin D and 0.5 mM dNTP. The reaction is incubated at 37°C for 1.5 h, followed by PCIA, pH 8.0 extraction (to remove actinomycin D) and nucleic acid precipitation. The cDNA pellet is suspended in 25 µl nuclease free water.

Ten genes are detected by 35 cycles of real-time PCR with 0.25 µM unique forward primer and universal reverse primer. 2 µl of cDNA is used of each individual 25 µl reaction. 1/4th volume of amplicons are run on 1 % agarose gel for 45 min at 90 V.

3.2.6 Two step process to remove excess long primers

Step 1 - 100 fmol of single-stranded DNA (82 bp) is digested with 40 U exonuclease I and 5 µg SSB protein in 1x exonuclease I (67 mM glycine-KOH, 6.7 mM MgCl₂, 10 mM 2-mercaptoethanol), 1x cDNA synthesis (50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂), 10 mM DTT and 6 µg actinomycin D at 45°C for 60 min (50 µl). The enzyme is inactivated at 80°C for 15 min. 1/10th volume of the reaction is amplified by 35 cycles of real-time PCR (25µl).

Step 2 - The original buffer (1.25 M NaCl, 13 % PEG 8000) of magnetic carboxylate beads (Agencourt AMPureXP bead technology) is replaced with different buffers with constant 1.25 M sodium chloride and varying percent PEG 8000 concentration. The beads are separated on magnet stand and the original buffer is aspirated to replace with the new buffer. This step is performed twice. PCR amplicons of genes GAL1 (450 bp), GAL2 (554 bp) and XDJ1 (402 bp) are mixed together and suspended in 1x first strand cDNA buffer, 10 mM DTT and 0.1 M sodium phosphate buffer (10 μ l). The amplicons are added to the 1.8x volume (18 μ l) beads and DNA is precipitated at 4°C for 10 min. The bead bound and unbound fractions are separated by placing the tubes on magnetic beads. The solution is aspirated to remove the unbound fraction. The beads are washed twice with 2x volume of freshly prepared 70 % ethanol. The remaining alcohol is evaporated by heating at 37°C for 2 – 5 min. The bound fraction is eluted from magnetic beads with 10 μ l Tris-HCl, pH 8.0 and transferred to a new 1.6 ml tube. The products are run on 1.2 % agarose gel for 45 min at 90 V.

3.2.7 Reverse transcription with sequence-specific long primer library

0.5 pmol of yeast sequence specific long primer library (approx. 4800 primer) is annealed to 500 ng of yeast total RNA (glucose and galactose) in solution containing 300 mM NaCl, 20 mM sodium phosphate and 10 U superase-in (RNase inhibitor). The reaction is heated at 65°C for 5 min to denature RNA secondary structure, cooled down and incubated at 45°C for 12 h. 5 μ l of hybridization solution is added to 10 μ l of reverse transcriptase enzyme cocktail mix containing 1x first strand buffer, 10 mM DTT, 10 U superase-in, 0.5 mM dNTP, 300 ng/ μ l actinomycin D, and 150 U superscript III. First strand cDNA synthesis is incubated at 45°C for 0.5 h and then 50°C for 1 h. Next, 10 μ l of exonuclease reaction mix (2.5 μ l of 10x

exonuclease I buffer, 10 U superase-in, 4 µg SSB protein and 40 U exonuclease I) is added to reaction (total volume 25 µl) and incubated at 45°C for 30 min. The RNA:DNA hybrids and excess long primers are fractionated with 1.8x vol AMPure beads in 1.25 M NaCl, 10 % or 11 % PEG 8000 solution. The hybrid products are eluted with 10 µl Tris-HCl, pH 8.0, of which 0.5 µl per gene is used for PCR detection.

3.2.8 *Saccharomyces cerevisiae* growth conditions and total RNA extraction

Saccharomyces cerevisiae strain W303 is grown in 2 % glucose and 2% galactose rich media at 30°C under constant shaking 300 rpm. The cells are harvested between OD600 0.4 – 0.6 and total RNA is extracted using hot phenol total RNA extraction protocol [236]. The RNA concentration is quantified on nanodrop spectrophotometer. Any genomic DNA in the samples is digested with 0.2 U TurboDNase (Ambion Inc.) per 10 µg RNA at 37°C for 15 min. The DNase enzyme is inactivated with proprietary DNase inactivation slurry supplied with the kit (Ambion Inc.). 1 µg of total RNA before and after DNase I digestion is run on 0.8 % agarose gel to test for RNA integrity.

3.2.9 Polymerase chain reaction (PCR)

In general, 50 µl of PCR reaction mix consists of 0.5 µM forward primer, 0.5 µM reverse primer, 0.2mM dNTP, 2.5 mM MgCl₂ and 2.5 U GoTaq flexi polymerase in 1x concentration of proprietary buffer (Promega In.) PCR amplification is done in Mastercycler ep gradient S (Eppendorf Inc.). The amplification program consists of 30 cycles as follows: initial dsDNA denaturation at 94°C for 2 min; 30 cycles of dsDNA denaturation (Td) at 95°C for 20 s, primer annealing (Ta) at T_m-5°C for 30 s, primer extension (Te) at 72°C for 30 s; final elongation at 72°C for 5 min.

3.2.10 Real-time PCR and Melt Curve

In general, 25 µl of PCR reaction mix consists of 0.25 µM forward primer, 0.25 µM reverse primer, 0.2mM dNTP, 2 µM SYTO 13 dye, 2.5 mM MgCl₂ and 2.5 U GoTaq flexi polymerase in 1x concentration of proprietary buffer (Promega In.) PCR amplification is done in Chromo4-PTC 200 DNA engine (MJ Research Inc.). The amplification program consists of 30 cycles as follows: initial dsDNA denaturation at 94°C for 2 min; 30 cycles of dsDNA denaturation (Td) at 95°C for 10 s, primer annealing (Ta) at T_m-5°C for 20 s, primer extension (Te) at 72°C for 20 s; final elongation at 72°C for 5 min.

3.2.11 Microarray analysis

The samples are hybridized on overnight (18 h) at 45°C in 6x SSPE buffer, 10 % formamide, 0.01 µg/µl acetylated BSA and 0.01 % Tween 20 solution. The arrays are successively washed twice with non-stringent buffer (1x SSPE, room temperature and 45°C) and once with stringent buffer 0.25x SSPE (room temperature) for 3 min. They are scanned with Axon 4000B scanner.

The raw data of remaining arrays is processed to flag probes with high coefficient of variation (Std. deviation / Mean). The intensity of the flagged probes is replaced by the average intensities of replicate gene probes (5 replicate probes per gene). Variations in the multiple one-color arrays are removed using a two-step normalization procedure. First, the replicate arrays are normalized using quantile normalization (QNorm) algorithm such that they have absolutely identical distribution, based on the assumption that the cDNA populations hybridized to the arrays is the same. Second, global scaling normalization (Median scaled) is performed to adjust all arrays to an average baseline.

3.3 Experimental results and discussion

3.3.1 Interpretation of real-time PCR plots.

Amplification plot: For data analysis, the C_q value or threshold value between the exponential and linear phase of PCR is extracted using Opticon 3.0 software. C_q value gives the cycle number at which the fluorescence signal corresponding to dsDNA product is visible. The sooner fluorescence is visible, the more template is present, and vice versa.

Melt curve: The melt curve shows the melting temperature of PCR product in the form of a peak. A single sharp peak represents PCR products of the same size, whereas multiple sharp peaks represent different size DNA. In general, one DNA size is amplified per real-time PCR tube. Thus presence of multiple peaks represents formation of non-specific products.

3.3.2 One round aRNA amplification with long gene-specific primers

The reverse transcription of mRNA into cDNA is usually primed with either short random primers or (dT)_n oligonucleotides. In the case these cDNA have to be transcribed into aRNA, a promoter sequence is fused at the 5' end of the primer (only to oligo(dT) primers). Usually, this sequence is a 19-mer corresponding to the binding site of the T7 phage polymerase. During reverse transcription, this short promoter sequence is dangling at the extremity with little, if no effect on the hybridization of the 3' ends of the primer to its target and the yield of aRNA during in vitro transcription reaction. We investigated if a significantly larger dangling sequence can be safely used for one round aRNA amplification.

In order to fully control all parameters during one round RNA amplification, we produced a 0.78 kb long PCR amplicon of *S. Cerevisiae* actin gene (ACT1 – YFL039C) with T7 promoter sequence at the 5'-end of leading DNA strand. The PCR amplicon is transcribed to RNA followed by DNase digestion of template DNA, phenol: chloroform extraction and spin column purification of RNA. Next, we designed a long oligonucleotide (71-mer) made of the following sub-units; 26-mer T7 promoter sequence at the 5'-end (*italic*), 20-mer random sequence (underlined) and a 25-mer sequence complementary to the in vitro transcribed RNA (**bold**) (5'-TAATACGACTCACTATAGGGAGCAGACTGAAGCTTACCTAGCTGTTTAAACTGACTTGACCATCTGGAAGT-3'). When hybridized to the RNA molecules, the oligonucleotides has 51 nucleotides dangling at its 5'end.

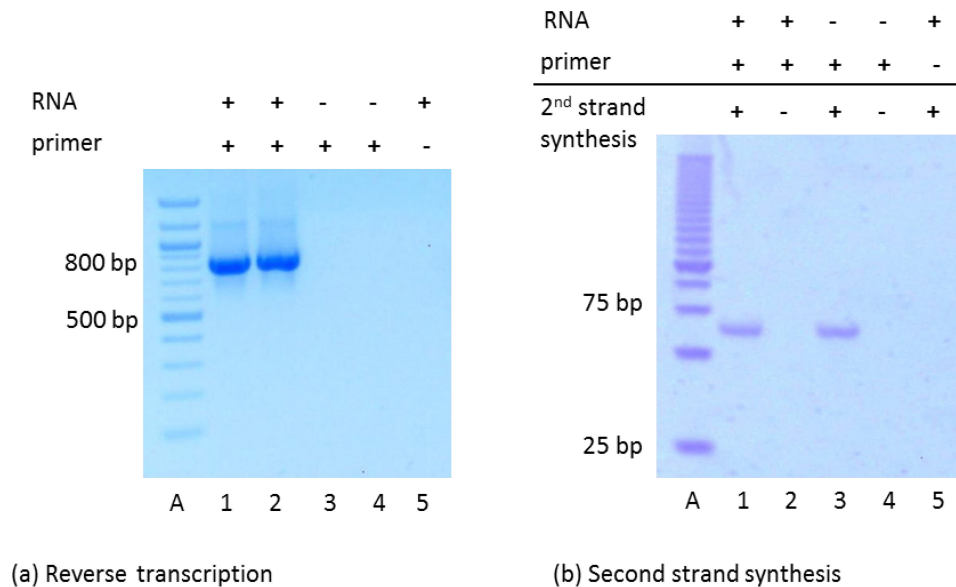


Figure 3-1 cDNA synthesis with sequence-specific primers 1

When the long primer is used for priming the reverse transcription reaction [Figure 3-1.a], we obtain cDNA (800 bp marker) with expected size (lane 1, 2). When one of the main reagents (long primers, RNA) is omitted as a negative control, we do not observe any cDNA (lane 3 – 5). From, Figure 3-1.b. the primers become double-stranded on addition of 2nd strand DNA polymerase irrespective of whether there is RNA in the reaction tube (lane 1 – mRNA present and lane 3 – minus RNA). There are no duplex primers when second strand synthesis polymerase (lane 2, 4) or primers (lane 5) are omitted as negative controls.

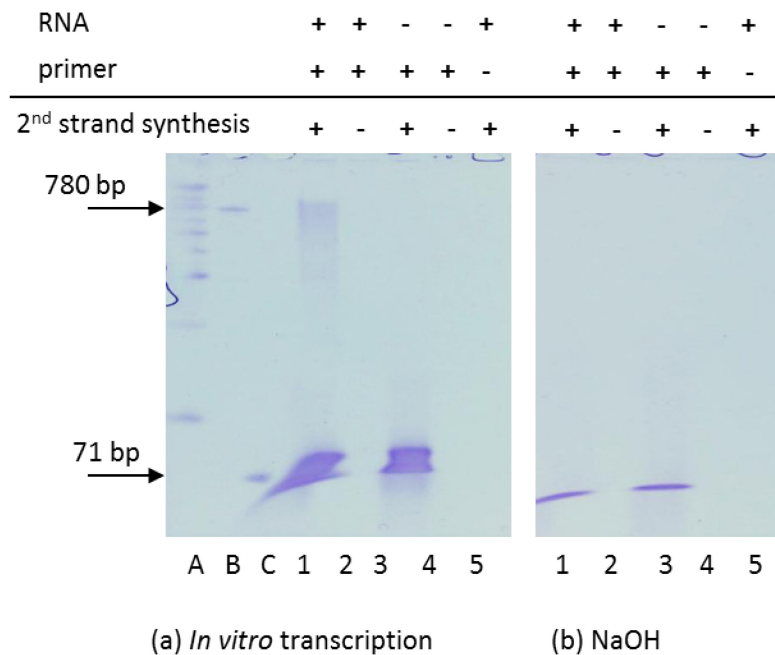


Figure 3-2 cDNA synthesis with sequence-specific primers 2

An aliquot of 2nd strand reaction is used to transcribe aRNA [Figure 3-2.a.] RNA ladder (lane A), template RNA (780 bp, lane B, top arrow) and long primer (71 bp, lane C, bottom arrow) are run as size markers. There is amplification of ACT1 mRNA (lane 1, RNA smear around 600 – 800 bp, top arrow) as well as amplification of undesired short RNA molecules (lane 1, bottom arrow). The yield of desired full length RNA is much lower than short RNA fragments transcribed from duplex primers (lane 1). This is due to high turnover rate of the T7 polymerase on the shorter ds-primer template. There is no IVT in negative second strand synthesis (lane 5, 7) and minus-primer (lane 8) controls. Formation of aRNA is confirmed by degradation with sodium hydroxide as shown in Figure 3-2.b.

This experiment demonstrates, i) the feasibility of using primers with significantly long 5' dangling ends to initiate the reverse transcription of RNA molecules, ii) duplexation of long primers in the absence of RNA during second strand DNA synthesis temperature, leading to transcription of short RNA molecules, and iii) the requirement of double-stranded DNA for successful T7 RNA polymerase IVT reaction.

3.3.3 Removal of excess sequence-specific primers

After reverse transcription reaction the excess sequence-specific primers have to be removed prior to IVT reaction to increase yield of desired RNA and most importantly eliminate detection of false positives by PCR, microarrays or 'next-generation' sequencing. Three possible methods to remove excess long primers are i) filter spin columns, ii) selective digestion of single-strand DNA and iii) bead-based purification. Each of the methods is discussed in detail under separate sub-headings

Silica-based filter spin column purification

Majority of silica-based filter spin columns available from manufacturer's use chaotropic salts and alcohol to enhance and influence the binding of nucleic acids to silica membrane. The percent ethanol and the volume effect binding capacity, too much leads to binding of degraded nucleic acids and small species that influence UV260 readings and too little make it difficult to wash away all of the salt from the membrane.

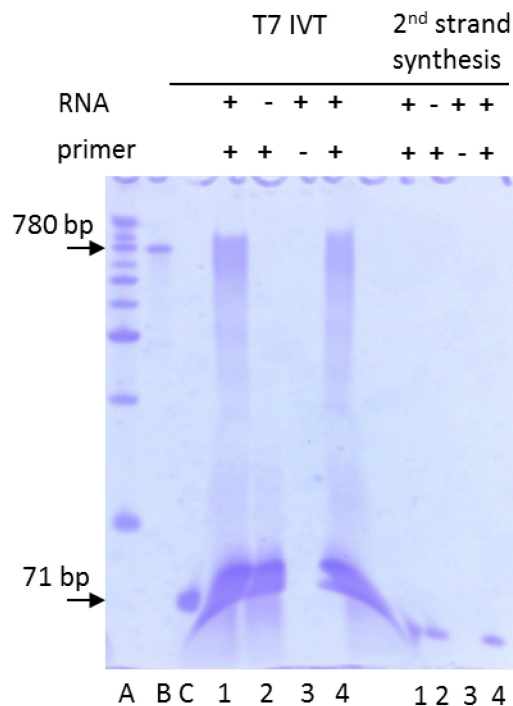


Figure 3-3 Spin column removal of double-stranded primer

We investigated the possibility of using these columns to remove ds-primer observed after second strand synthesis reaction when using one-round RNA amplification kit [Figure 3-3]. The gel is divided under two headings; i) T7-IVT and ii) 2nd strand synthesis. Lanes 1 – 4 in T7-IVT are the downstream products of lanes 1 – 4 of 2nd strand synthesis. 100bp ssRNA ladder (lane A), ACT1 mRNA (lane B, top arrow) and long primer (lane C, bottom arrow) are run as size markers. Similar to the previous gel, there is transcription of Act1 mRNA (RNA smear, lane 1, 4, top arrow), with majority of short RNA molecules.

The 2nd strand synthesis products corresponding to lane 4 are purified on silica-based spin column prior to loading on gel. The presence of dsDNA and short RNA molecules (lane 4) indicate that purification on spin column does not remove ds-primer (71bp). The DNA cutoff threshold of commercial purification kits is 100 bp for dsDNA and 200 bases for ssDNA. The size-cutoff threshold can be changed by varying the chaotropic salts and alcohol mix ratio. This direction is not considered as cutoff threshold depends on the concentration of short dsDNA and spin-columns are optimized for PCR reactions where, the concentration of amplicons (usually > 200 bp) is much greater than unreacted primers and dNTPs. We have observed presence of PCR-primers even after spin-column purification (data not shown).

Selective digestion of single-stranded DNA

One possibility to remove excess long primers is to selectively digest them (ssDNA) with single-strand specific exonucleases after reverse transcription. The use of a sequence-specific primer with 5' dangling tail, restrains the choice of single-strand exonuclease. The exonuclease should not degrade RNA:cDNA hybrids and have only 3' – 5' mode of action. Only exonuclease I meets these criteria.

Exonuclease I digestion of sequence-specific long primer

We investigated the amount of digestion achieved by exonuclease enzyme to digest ssDNA equivalent to the size of sequence-specific long primers (60 – 120 bp). Also, the input concentration of ssDNA is kept the same as long primers in the RT reaction. 2 pmol of ssDNA (66 bp) is digested with exonuclease I, of which 1/10th volume is amplified by real-time PCR. The C_q or cutoff threshold value between the ‘no digestion’ negative control and digestion reactions is used to determine the digestion fold- factor. There is a difference of approx. 7.5 cycles between positive (C_q = 11.88 ± 0.07) and minus (C_q = 4.41 ± 0.01) exonuclease I reactions [Figure 3-4]. A higher C_q value indicates lower template starting concentration. Assuming PCR of efficiency of 95%, this gives $1.957.5 = 150$ fold reduction in the concentration of single-stranded DNA upon exonuclease I digestion.

Exol (Units)	[SSB] ug	C _q	C _q (SD)	Fold digestion compared to (-) Exol (-) SSB
0	0	4.41	0.01	
20	0	11.88	0.07	1.47E+02
20	0.2	11.81	0.23	1.40E+02
20	2	18.69	0.10	1.39E+04
20	6	25.70	0.48	1.50E+06
40	2	19.01	0.27	1.72E+04
40	6	26.00	0.43	1.82E+06

Figure 3-4 Effect of SSB protein on amount of ssDNA digested

Effect of SSB protein concentration on exonucleolytic digestion of ssDNA

Different additives namely, DMSO, bovine serum albumin (BSA), CaCl₂, ATP and single strand binding (SSB) protein are added to the exonuclease I reaction to increase amount of ssDNA degraded. Of these, only SSB protein showed some promise and is studied in more detail. Once again 2 pmol of ssDNA (66 bp) is hydrolyzed at different concentrations of SSB protein and exonuclease I [Figure 3-4]. The amount of ssDNA digested remains unchanged at low concentrations of binding protein (0.2 μ g) and progressively increases with increase in binding protein concentration. At 6 μ g SSB concentration, there is a million fold digestion of ssDNA.

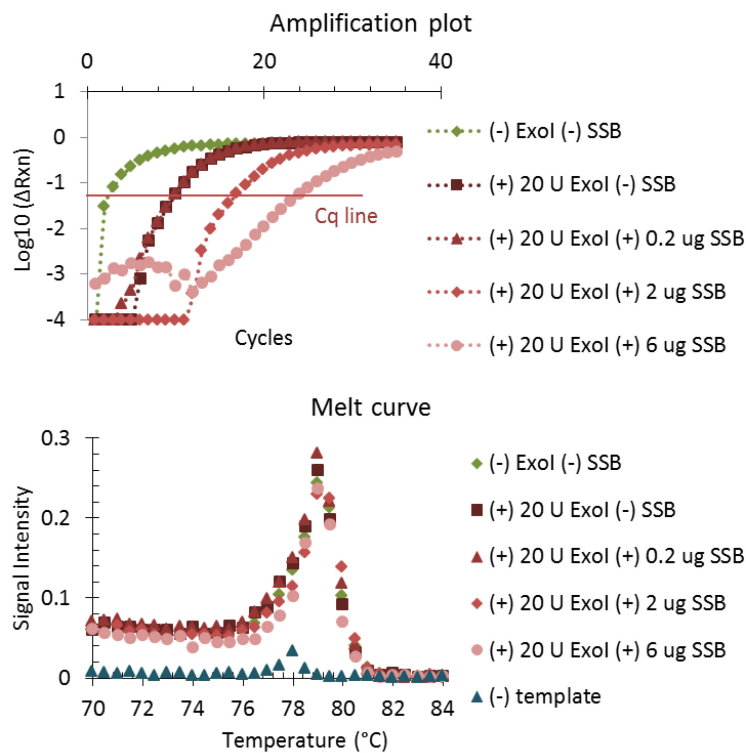


Figure 3-5 Real-time PCR - Effect of SSB protein on amount of ssDNA digested

Real-time PCR amplification and melt curve plots are displayed to show amount of oligonucleotides digested and presence of correct size DNA [Figure 3-5]. From the melt curve, we see that even after million fold digestion, real-time PCR is able to detect low concentrations of ssDNA (approx. 0.26 nmol). This will confound interpretation of gene-expression data discussed in the upcoming sections. Similar experiments are performed on a longer template (115 bp). In this case the concentration of SSB protein has to be increased to 10 μ g to get million fold digestion. This could be due to secondary structure or self-hybridization of DNA molecules, though none are found by upon structural analysis. SSB protein interacts with both exonuclease I and DNA molecules. At excess concentrations of binding protein, there is a 3- fold increase in rate of digestion catalyzed by exonuclease I [237]. Also, SSB protein unwinds some of secondary structure in ssDNA and opens partially duplexed strands depending on length of ssDNA extension or size of ssDNA loop [238].

Size-dependent fractionation of DNA on magnetic beads

Another method to remove excess long primers is to fractionate the DNA according to size. Although fractionation of nucleic acids on agarose or polyacrylamide gels gives sharp cut-off bands, the recovery of nucleic acids takes anywhere from 2 h (agarose gel) to >24 h (acrylamide gel) and with only 30 – 50 % yields [239]. Also, purification on gel is limited by the nucleic acid concentration required to visualize the DNA band. One method to precipitate DNA is to use polyethylene glycol (PEG). The size of DNA molecules precipitated depends on the molecular weight and concentration of PEG as well as the concentration of salt and incubation time [240, 241]. This method has been adapted to reversibly and non-specifically bind nucleic acids on carboxylated magnetic beads. The PEG and salt

(sodium chloride) concentration of the solution is adjusted to tune the size of DNA bound to the beads. The beads with bound nucleic acids are separated from solution (unbound DNA fragments) and then eluted of the beads. The unbound DNA fragments can be recovered by ethanol precipitation. Thus using different concentration of PEG, it is theoretically possible to get many size fractions of DNA.

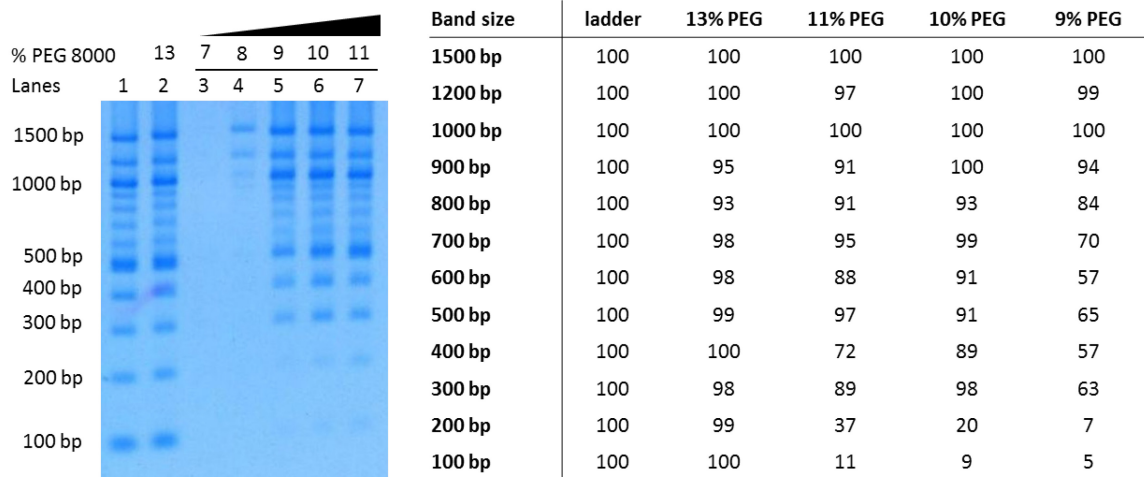


Figure 3-6 Size dependent fractionation - PEG concentration (1.25 M NaCl)

Commercially available, carboxylated magnetic bead kit (AMPure XP, Beckman Genomics Inc.) for DNA size fraction is studied to remove DNA less than 200 bp, that is, to remove long primers and capture cDNA (>200 bp). The beads are supplied in buffer (13% PEG 8000, 1.25 M NaCl) rated to remove DNA less than 100 bp. To remove DNA < 200 bp, the PEG concentration of the buffer is reduced and then used for fractionation of 100 bp ladder (range 100 bp – 1500 bp, New England Biolabs) [Figure 3-6]. We observe that under room temperature (30°C) binding conditions, reducing the PEG concentration leads to progressive loss of smaller sized DNA strands. Upon quantification, the desired result is achieved with 10% PEG 8000 concentration, which has a large difference in amount of DNA recovered between 300 bp (> 95%) and 200 bp (approx. 20%).

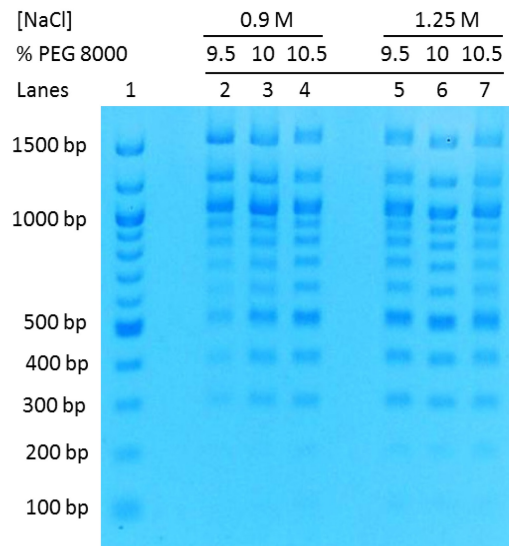


Figure 3-7 Size dependent fractionation PEG and NaCl concentration

The size selection is fine-tuned by varying the % PEG concentration in steps of 0.5 % (range: 9.5 % - 10.5 %), at two different salt concentrations (0.9 M NaCl and 1.25 M NaCl) [Figure 3-7]. For this experiment the 100 bp ladder is suspended in 1st strand DNA synthesis buffer to test removal of actinomycin D. Focusing on DNA sizes 100 – 400 bp, for 1.25 M NaCl there is not much qualitative difference in the cutoff threshold across % PEG concentration. There is near complete recovery of 300 bp and >50% loss of 200 bp DNA strand. In the 0.9 M NaCl, 9.5% PEG 8000 solution there is much greater loss of DNA between 300 bp – 700 bp as compared with other solutions. This gel is only qualitatively assessed as the high background levels result in greater than 25% uncertainty when quantifying the gel. Thus a combination of exonuclease I digestion and AMPure beads purification has the potential to remove excess long primers.

3.3.4 Optimize reverse transcription conditions

Sequence-specific primers are useful for targeted capture of transcripts from a population. The temperature at which the primers bind to their specific RNA is dependent on the length of the hybridization sequence and the monovalent and divalent salt concentration. It is important to have the optimal temperature of binding and reverse transcription, as too low a temperature will result in non-specific capture of transcripts (background priming) and at too high a temperature very few RNA molecules will be reverse transcribed. Also, there is a variation of 5 - 6 orders of magnitude in concentration of transcripts in yeast biological samples [242]. For unbiased results, it is desired that the sequence-specific primer be in excess of its corresponding target (RNA). We investigated the optimal hybridization

temperature for 30 bp binding sequence and the ratio of long primer to RNA for accurate detection of a specific RNA species.

Effect of temperature on reverse transcription

Constant molar concentration of GAL 7 mRNA is reverse transcribed using 100x excess long primers at 4 different temperatures (37°C, 45°C, 50°C and 55°C). The resultant cDNA is amplified by real-time PCR, followed by melt curve analysis [Figure 3-8]. The sharp band in melt curve (79°C) conveys that there are no non-specific products and the signal observed in amplification plot corresponds to desired product. From, Figure 3-8 assuming average 90% PCR efficiency, from Cq value the amount of RNA detected at 37°C and 45°C is approximately 10 times more than at 50°C and 55°C.

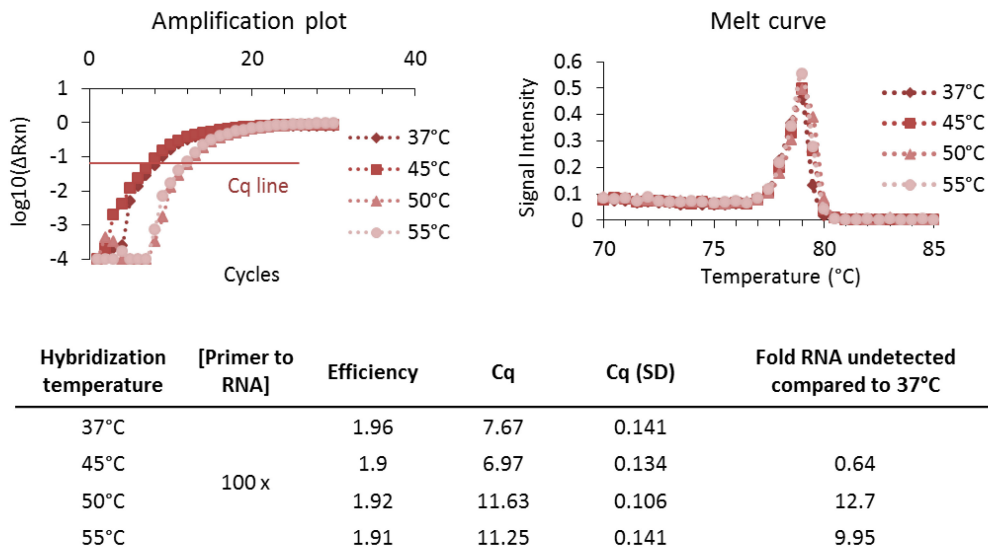
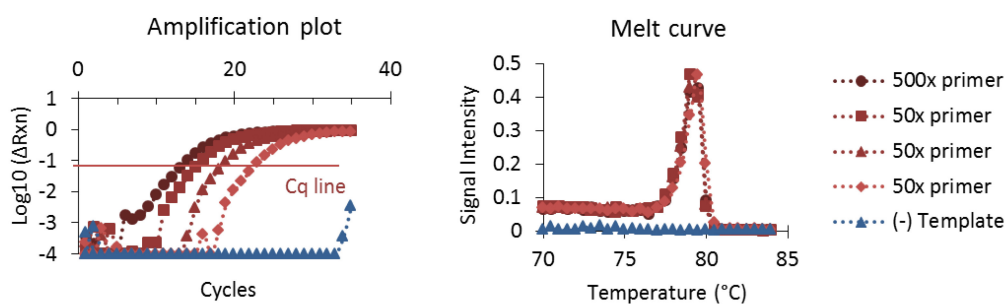


Figure 3-8 Effect of temperature on reverse transcription

Effect of primer concentration on reverse transcription

Keeping constant the yeast Gal7 RNA (0.26 fmol) concentration, the concentration of long primer concentration is varied from 500x – 0.5x excess. Both transcript capture and reverse transcription are performed at 45°C for 60 min. The amount of cDNA synthesized at a given primer concentration is determined by real-time PCR. From, Figure 3-9 assuming 85% PCR efficiency (Cq value) there is approximately a 10- fold reduction in amount of RNA detected between 0.5x and 50x excess primer concentration. Three times less RNA is detected between 50x than 500x excess primer indicating a possible saturation has been reached. Thus, for a 30 bp gene-specific primer, the hybridization temperature should be 45°C and with 50 x excess primer.



[RNA] fmol	[Primer]	Efficiency	Cq	Cq (SD)	Fold RNA undetected compared to 500 x primer
0.26	500 x	1.83	15.15	0	
	50 x	1.88	17	0.099	3.1
	5 x	1.82	20.26	0.163	23.2
	0.5 x	1.87	24.43	0.049	301.5

Figure 3-9 Effect of primer concentration on reverse transcription

3.3.5 Detect differential expression of transcripts with sequence-specific primers

Sequence-specific primers are used to determine gene expression differences in two different biological samples by targeting a small subset of the transcriptome population (approximately 100 genes). In both samples, amino-allyl dUTP is incorporated into the cDNA strand for subsequent fluorescent dye labeling and detection on microarray.

Expression of Mouse testicle and brain genes

Genes expressed in mouse tissue samples is obtained from the SymAtlas database maintained by the Novartis Institute for Functional Genomics, Inc (<http://symatlas.gnf.org/SymAtlas/>). Genes with high 'fold over median' values in either brain or testis tissues only or in both and with low 'fold over median' values in both tissues are extracted from the database. This list is cross referenced with the MGI – Mouse Genome Informatics [243] and Ensembl [244] databases to extract their coding sequence. Of these, the differential expression of 108 genes is studied with oligonucleotide microarrays. They are distributed of 108 genes is as follows; i) 11 gene over-expressed in testicle tissue, ii) 11 genes over-expressed in brain tissue, iii) 10 genes with high expression in both samples and iv) 76 with low or no expression in both samples.

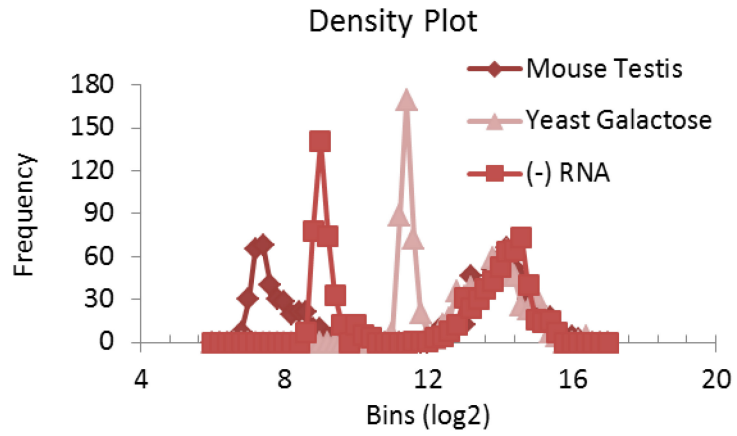


Figure 3-10 Density Plot – cDNA synthesis with mouse sequence-specific primer

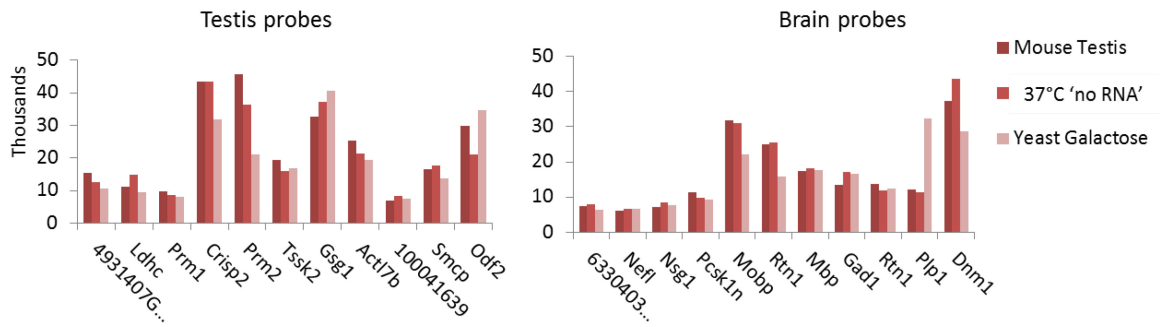


Figure 3-11 Differential gene expression – mouse sequence-specific primers

Long (mouse sequence-specific) primers are designed for each of the 108 genes where each primer has the following subunits, 5'-T7 promoter-DNA tag (barcode)-sequence specific sequence-3'. The sequence specific sequences (19 – 21 nt) are designed to target the selected genes (50mM NaCl, T_m range: 60 – 65 degree celsius, 1uM concentration) using OligoArray 2.1 software [245]. DNA tag library (20 nt) with at least 11 mismatches between probes and with constraints on sequence composition (<3 stretches of mono-,di- and tri-nucleotide repeats) and secondary structure, melting temperature (60°C - 65°C) and GC content (40 % – 60%), are designed using in-house developed software.

The dye labeled cDNA is hybridized to complementary probes randomly synthesized in a 4-5k format (that is, 4 arrays per slide with 5000 probes each) on oligonucleotide microarrays. There are 108 probe sets that target cDNA and 73 probes sets (to detect percent false positives) that target mouse gene and tag sequences not included in the long primer library. Three samples are reverse transcribed with mouse sequence-specific long primers, a) mouse testis total RNA, b) yeast galactose induced total RNA and c) No-RNA negative control.

Once the total intensity is normalized across the conditions, the probes intensities are log base2 transformed and displayed as density plot [Figure 3-10]. From the density plot, for each sample there are two distinct distributions, where the higher value distribution is the same across the three samples and represent the mouse probes. All the probes used to detect false-positives make up the lower value distribution which serves as the background cutoff threshold. The high background (approx. 4000 A.U.) in yeast total RNA sample is due to non-specific binding of non-DNA bound fluorescent dye.

Across all three samples, we are able to detect all 108 genes. The individual genes have the same signal intensity across three samples [Figure 3-11]. This is a

red flag as; a) transcripts expressed in the brain should not be detected in mouse testicle sample, b) the long primers are designed to target mouse and not yeast mRNA, and c) there is detection of cDNA in 'no RNA' control. Especially, in the 'no RNA' reaction, there should be no reverse transcription, but the incorporation of fluorescent dye and detection on microarray suggest the opposite.

Reverse transcription of ssDNA and Exonuclease I digestion of ssDNA library

In the above experiments, a three step temperature is used for cDNA synthesis starting at 37°C for 30 min followed by 45°C and 50°C incubation for 30 min each. Under the initial low temperature there may be significant mispriming events leading to false positive signals on microarray. We investigated the effect of RT temperature (45°C, 50°C) on double-strand formation of excess long primers as well as exonuclease I digestion at higher temperature [Figure 3-12]. These experiments are showed together as the substrate in lane 2 is used synthesize the products in lane 6 and 7. 25 bp DNA ladder is run as size marker (lane 1, 4).

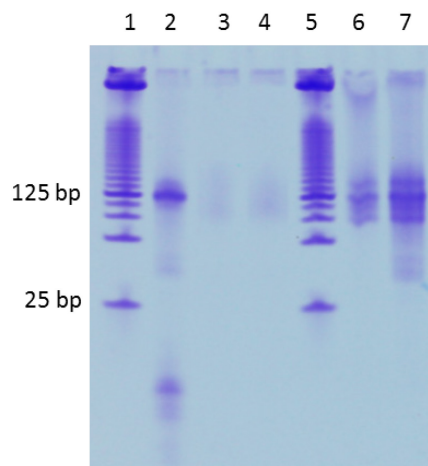


Figure 3-12 Exonuclease I digestion and reverse transcription of ssDNA

First, exonuclease I digests ssDNA (lane 2) at 37°C (lane 3) and 50°C (lane 4). The faint band (lane 2, lane 3, near 125bp) indicates that the exonuclease I is unable to completely digest the 10 pmol of ssDNA at either of the temperatures in this reaction. Also, the short oligonucleotides in the long primer sample (lane 2, below 25 bp band) if not removed will prime downstream DNA synthesis.

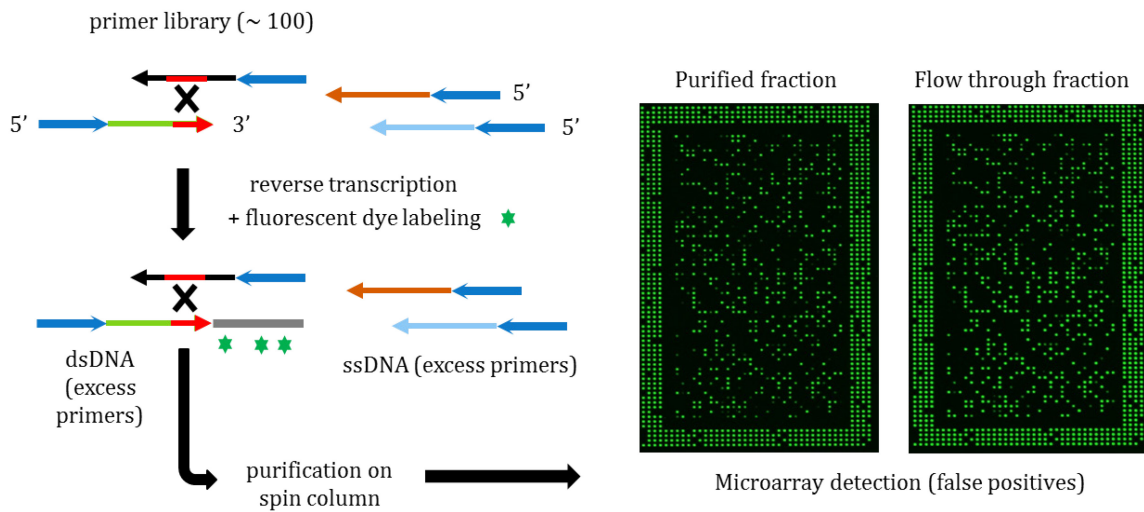


Figure 3-13 'no RNA' reverse transcription with sequence-specific primers

Second, the 45°C and 50°C 'no RNA' RT reaction products are purified on spin column prior to analysis on gel (45°C) or DNA microarray (50°C). Although not evident from the gel image, there is fluorescent DNA smear in elute (lane 6) and flow-through (lane 7) fraction of 'no RNA' RT reaction. Here, silica-based spin-columns only remove 40 - 50 % double-stranded long primers. When the 50°C 'no RNA' elute and flow-through fractions are hybridized on DNA microarray they light up all the probes with the same signal pattern as 37°C 'no RNA' cDNA synthesis control [Figure 3-13]. The figure also shows the presence of two contributing factors to false positive signal, single-stranded and double-stranded excess primers.

These experiments show that; i) reverse transcriptase is able to use DNA as a template for cDNA synthesis, i.e. they have RNA and DNA dependent DNA polymerase (RdDp and DdDp), ii) The duplexation of the long primers leads to confounding of data, iii) Exonuclease I alone and combination of exonucleaseI-spin column purification do not sufficiently remove labeled ds-primers and iv) increasing reverse transcription reaction does reduce signal intensity, though all 108 genes are detected (data not shown).

3.3.6 Activity of reverse transcriptase in Actinomycin D

Actinomycin D is an antibiotic that binds to DNA and inhibits RNA synthesis [246]. It has been reported that actinomycin D (ActD) at concentrations of 100 – 200 ng/ μ l inhibits spurious synthesis of second strand cDNA catalyzed by reverse transcriptases [247] and has been used to reduce detection of false-positives on microarrays [248].

In the next set of experiments, we investigate whether actinomycin D reduces double-stranded primer formation, such that they remain single-stranded and can be hydrolyzed by exonuclease I. An initial experiment is done to determine

the effect of actinomycin D on RNA-dependent DNA polymerase activity of reverse transcriptase. Reverse transcription of yeast GAL7 mRNA primed with long primer that has 30 bp (underlined) sequence specific sequence (48 bp, 5'-CTGCTGAACTTCAGATCGAGTCGCATTCAAAGGAGCCTGATGGATACC-3') is done with increasing concentration of actinomycin D (0 – 0.32 $\mu\text{g}/\mu\text{l}$). The resultant cDNA is amplified by real-time PCR and the products are analyzed on agarose gel [Figure 3-14]. There are PCR amplicons (150 bp) across all concentrations of actinomycin D. Analysis of Cq values does not show significant inhibition of RdDp activity of reverse transcriptase for actinomycin D concentration $<1.6 \mu\text{g}/\mu\text{l}$.

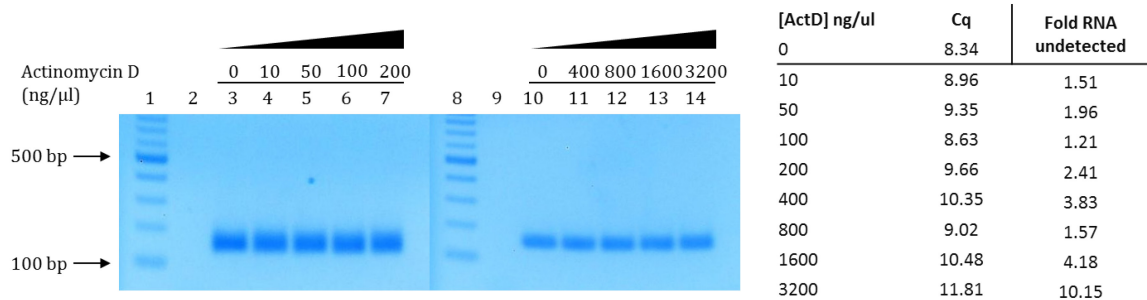


Figure 3-14 Activity of reverse transcriptase in presence of Actinomycin D

3.3.7 Differential expression of yeast genes induced in glucose and galactose media

Yeast messenger RNA sequences are extracted from the Saccharomyces Genome database [249]. Long primers are designed to target specific genes where each primer has the following subunits, 5'-T7 promoter-DNA tag (barcode)-sequence specific sequence-3'. The sequence specific sequences (29-31 nt) are designed to target the selected genes (50mM NaCl, T_m range: 60 – 65 degree celsius, 1uM concentration) using OligoArray 2.1 software [245]. DNA tag library (20 nt) with at least 11 mismatches between tags are designed using in-house developed software. Both the sequence-specific and tag region are detected by real-time PCR and microarray.

Of 5885 yeast protein encoding genes, 4800 pass the constraints set on sequence composition, secondary structure, and melting temperature. The primer library is synthesized on oligonucleotide microarrays and after series of molecular biology steps is obtained in single-stranded form, ready to primer reverse transcription reaction. The reverse transcription reaction is carried out in two steps a) hybridization of long primers to their transcripts, and b) first strand cDNA synthesis. To increase sensitivity, the primers are hybridized to their targets for extended period of time (2 – 12 h) followed by addition of the reverse transcription buffer-enzyme mix to initiate DNA strand synthesis. We investigated the removal of excess un-hybridized long primers either before or after cDNA synthesis.

Microarray study: Actinomycin D effect on DNA dependent DNA polymerase activity of reverse transcriptase

Yeast transcripts expressed in galactose media are primed with sequence specific primer library. The inhibition of DdDp activity of reverse transcriptase and removal of excess primers is studied by processing the first strand cDNA synthesis

reaction using four different methods. The dye labeled cDNA is detected on oligonucleotide microarrays.

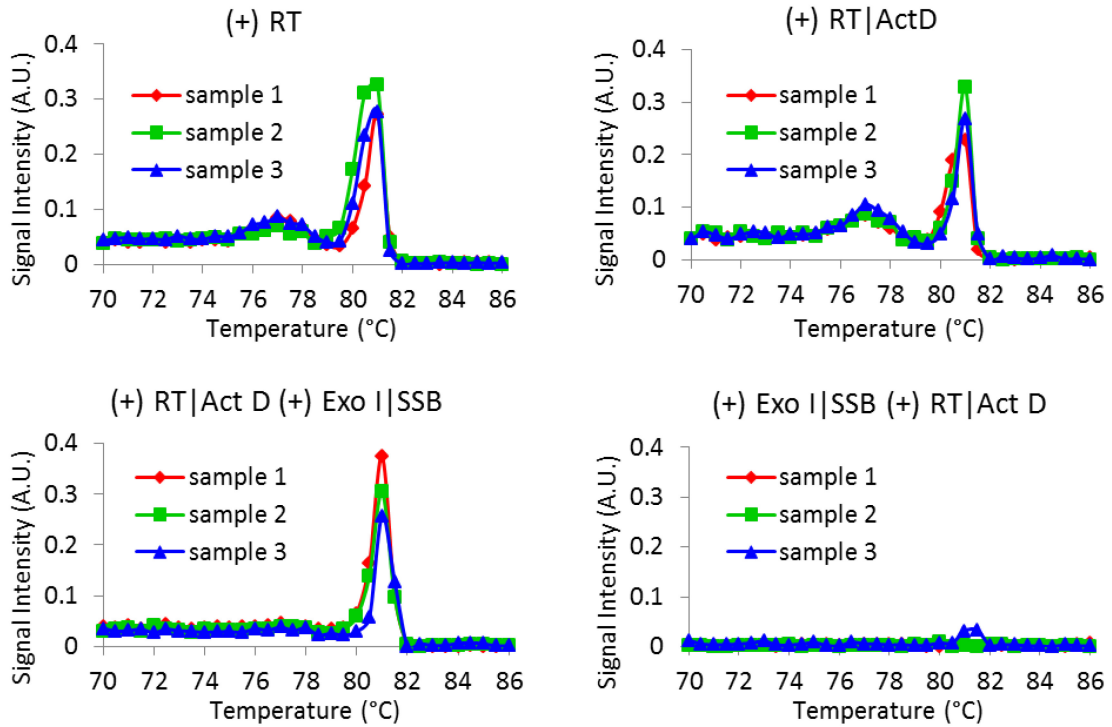


Figure 3-15 Melt curve of GAL1 gene - optimal sample processing method

Initially, in all four samples the sequence specific primers are annealed to their mRNA (GC) for 1 h at 50°C. Three of these are used for reverse transcription in the absence (RT; 1 sample) or presence of actinomycin D (RT|ActD; 2 samples). The excess primers in one of the RT|ActD samples are digested by exonucleolytic hydrolysis (ExoI|SSB). In the remaining fourth sample, the unhybridized long primers are digested from their 3' end, followed by first strand cDNA synthesis in presence of actinomycin D. The four methods are abbreviated as i) (+)RT, ii) (+)RT|ActD, iii) (+)RT|ActD (+)ExoI|SSB, and iv) (+)ExoI|SSB (+)RT|ActD. Each method is run in triplicate and the cDNA is labeled with fluorescent dye.

1/20th volume of cDNA is amplified by real-time PCR to detect highly expressed galactose induced GAL1 gene. The melting temperature (T_m) of Gal1 amplicon is 80.5°C and is detected in 3 of the 4 sample processing methods [Figure 3-15]. There is no GAL1 cDNA detected in any of the replicates where unhybridized primers are digested prior to first strand synthesis. Also, excess long primers are detected (amplicons with T_m = 77°C) in both samples not treated with exonuclease I, SSB protein enzyme cocktail.

Sample processing method	(+) RT	(+) RT ActD	(+) RT ActD (+) ExoI SSB
Array 1	416	157	157
Array 2	440	155	164
Array 3	429	157	165
Genes detected in all three methods		140	

Figure 3-16 Number of gene detected across sample processing methods

The rest of the cDNA is analyzed on DNA microarray. There is no signal detected in the three replicate arrays for sample where unhybridized long primers are removed prior to first strand synthesis (i.e. (+)ExoI|SSB (+)RT|ActD). The number of transcripts detected as present (cutoff threshold = 3x background intensity) on individual arrays across all three sample processing methods (that is, (+)GC (+)RT samples) is tabulated [Figure 3-16]. The amount of transcripts detected in RT reaction with actinomycin D is 2.8 times less than without actinomycin D. Also, 50% of the common genes (i.e. 70 out of 140) have 3 times less signal to minus actinomycin D control. This indicates that inclusion of ActD during RT reduces some of the false positive signal.

In addition, the low percentage of transcripts detected (only <10% of the probes on the array) may be due to contributions from three factors, i) too high hybridization and first strand cDNA synthesis temperature (50°C), ii) very low primer to mRNA concentration (0.2x primer per RNA) and iii) false negatives (hybrid/ization failure).

PCR study: Actinomycin D effect on DNA dependent DNA polymerase activity of reverse transcriptase

For the experiments below and rest of the theses, we compare the expression of yeast genes in glucose and galactose media and determine presence of excess primer by gel analysis. GAL genes (i.e GAL1, GAL2, GAL3, GAL7, GAL10 and PGM2) are expressed at high levels in galactose only. We used genes equally expressed in both glucose and galactose as controls or house-keeping genes. The list of genes is summarized in Figure 3-17 [250, 251]. Not all genes are compared across all experiments (only GAL2 – over-expressed in galactose media and PGI – equal expression in both media is kept constant).

Gene name	Galactose media	Glucose media
GAL 1	+	-
GAL 2	+	-
GAL 3	+	-
GAL 4	+	-
GAL 7	+	-
GAL 10	+	-
GAL 11	=	=
PGM 2	+	-
PGI 1	=	=
UGP 1	=	=
ALD5	=	=

Symbol	Expression Level
+	High
=	Equal
-	Low / Absent

Figure 3-17 List of yeast genes expressed in galactose - glucose

The experiment details are presented in Figure 3-18.a. The steps are as follows; i) total RNA is hybridized with 0.3 x of long primers, ii) followed by RT in the presence of actinomycin D. iii) removal of actinomycin D by phenol chloroform extraction, iv) exonuclease I hydrolysis of excess primers and v) detection of PCR products (i.e. cDNA) on gel [Figure 3-18.b]. The PCR primers are designed to target unique primer binding sequence (PBS) at 5' end of cDNA (grey arrow) and universal PBS at their 3' end (blue arrow). The universal PBS will amplify any leftover primers. As expected from literature data, the GAL genes (GAL 1, Gal 2, Gal 3, Gal 7, Gal 10, PGM2) are highly expressed in galactose over glucose growth conditions with equal expression in the control genes (GAL11, UGP1, PG1 and ALD5). This shows that a library of long primers with 5' dangling ends can be used to analyze gene expression.

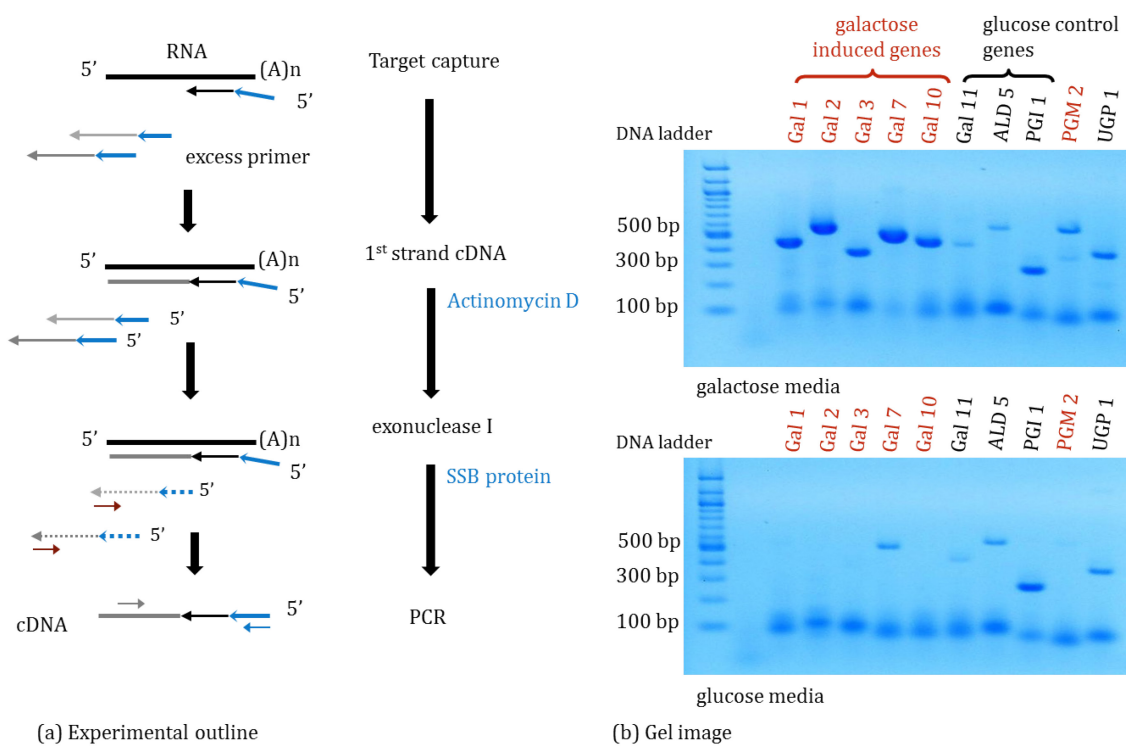


Figure 3-18 Yeast gene expression – Effect of actinomycin D

Unfortunately, there are present DNA bands between 100 bp and 200 bp DNA size markers. These bands correspond to undigested long primers (sequence-specific primer library). The earlier experiment shows that with actinomycin D, there is a reduction in number of false positives, but this experiment implies that actinomycin D does not completely inhibit formation of dsDNA. This is because actinomycin D inhibits second-strand DNA synthesis by binding to 3' of ssDNA, which prevents DNA-DNA hybridization required for strand extension. It does not inhibit DNA polymerase activity of reverse transcriptase enzyme [252]. As actinomycin D prevents DNA-DNA hybridization, we cannot add it to the preceding hybridization (target capture) step. The reduction in false positives is most likely seen due to inhibition of strand transfer [253].

3.3.8 Two step process to remove excess long primers

As actinomycin D does not completely prevent formation of double-strand primers (i.e. false positives), we investigated the use of a two-step removal process.

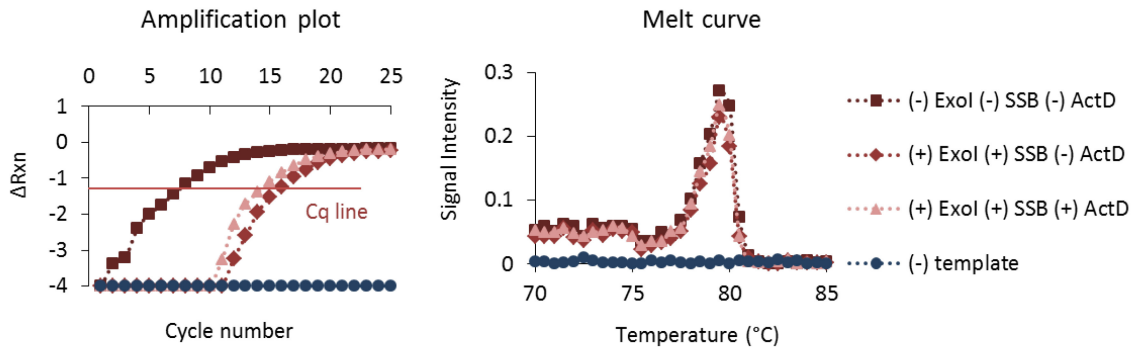


Figure 3-19 Exonuclease I digestion in presence of actinomycin D

Step 1 - To minimize sample processing steps, the efficiency of exonuclease I is studied in presence of actinomycin D. 82 bp ssDNA is hydrolyzed with exonuclease I enzyme cocktail in the presence or absence of actinomycin D (300 ng/ μ l). The amount of undigested ssDNA is determined by real-time PCR. Using no digestion control as reference there is 3.2 times less digestion of single-stranded DNA when actinomycin D is added to the reaction [Figure 3-19].

Step 2 - From earlier results, we know that exonuclease I alone does not digest ssDNA below the detection level of real-time PCR. Additional purification is required to remove the remaining long primers. The use AMPure beads after exonuclease I is studied to fractionate cDNA strands from un-hybridized primers. The size-fractionation of DNA is dependent on many factors; i) polyethylene glycol concentration, ii) salt concentration, iii) DNA concentration, iv) the incubation time, and v) the incubation temperature [240]. Three of the five parameters are kept constant (salt concentration, incubation time and temperature).

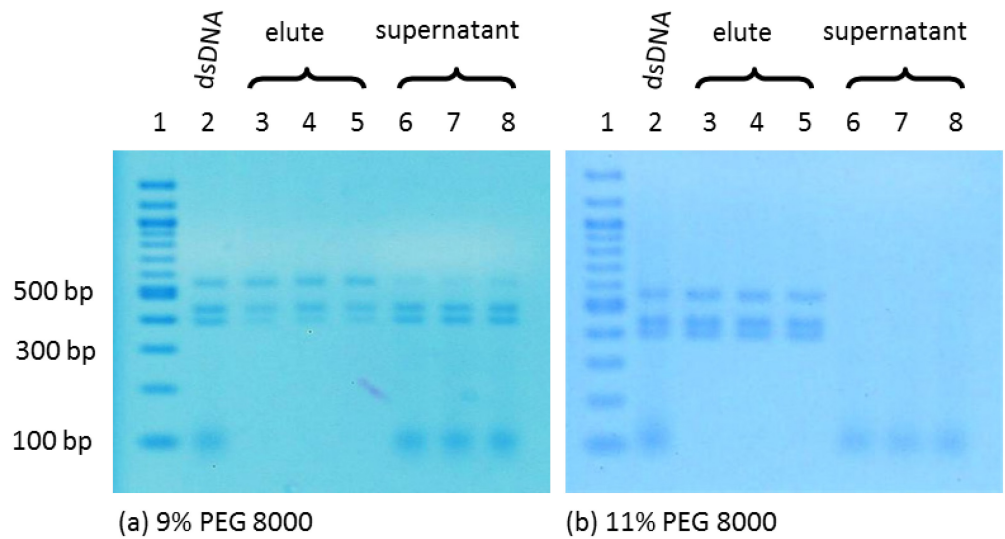


Figure 3-20 AMPure bead - Effect of PEG8000 concentration

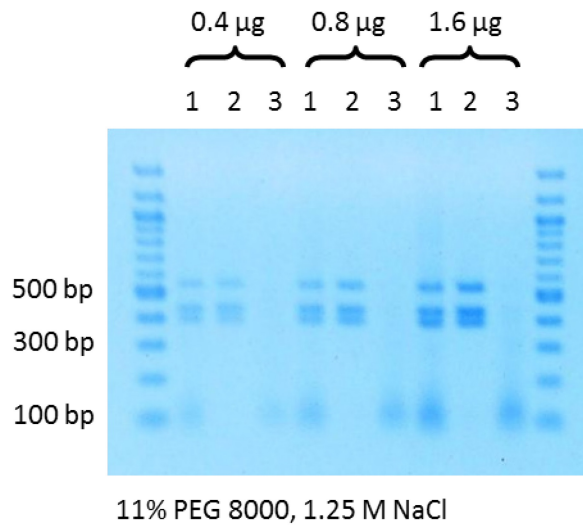


Figure 3-21 AMPure bead - Effect of DNA concentration

PCR amplicons of three genes (GAL1, GAL2 and XDJ1; 450bp, 554 bp, 402 bp respectively) exhibiting 100 bp ds-primer are co-purified with carboxylate beads at 9% and 11% PEG 8000 concentrations at constant salt concentration (1.25 M NaCl) [Figure 3-21]. The experiments are run in triplicate. The 100 bp ladder (lane 1) and non-purified amplicons samples (lane 2) are run as size markers. At 9% PEG solution, although there is complete removal of 100 bp ds-primer (lane 6 – 8, supernatant), there is approximately 50% loss of 400 bp and 450 bp PCR amplicons (compare elute and supernatant). Using 11% PEG solution, there is separation of 100 bp ds-primer (lane 6 – 8) with loss of any PCR amplicons (lane 3 – 5).

At very high input DNA concentrations there may not be enough beads (surface sites) for all DNA to bind, leading to some loss of cDNA strands. We investigated the effect of DNA concentration on purification efficiency of AMPure beads. Three concentrations of PCR amplicons (0.4 µg, 0.8 µg and 1.6 µg) are fractionated with 10 µl of AMPure beads [Figure 3-20]. At the concentrations tested there is complete separation of PCR amplicons (lane 2) from 100 bp ds-primer (lane 3). When compared with un-purified PCR amplicons sample (lane 1), there is no noticeable loss of cDNA PCR amplicons (lane 3). Thus, AMPure beads suspended in 11% PEG8000, 1.25 M NaCl buffer solution offer a solution to remove excess primers.

3.3.9 Reverse transcription with sequence-specific long primer library

The above process is implemented to study gene expression versus glucose and galactose growth conditions. The experiment details are presented in Figure 3-22.a. The steps are as follows; i) total RNA is hybridized with 10 x of long primers, ii) followed by RT in the presence of actinomycin D. iii) exonuclease I hydrolysis of excess primers, iv) size fractionation and v) detection of PCR products (i.e. cDNA) on

gel [Figure 3-22.b]. The cDNA product is detected by real-time PCR with forward primers specific to the gene and universal reverse primer. As the universal reverse primer has sequence common to all long primers, it will amplify any leftover excess primers.

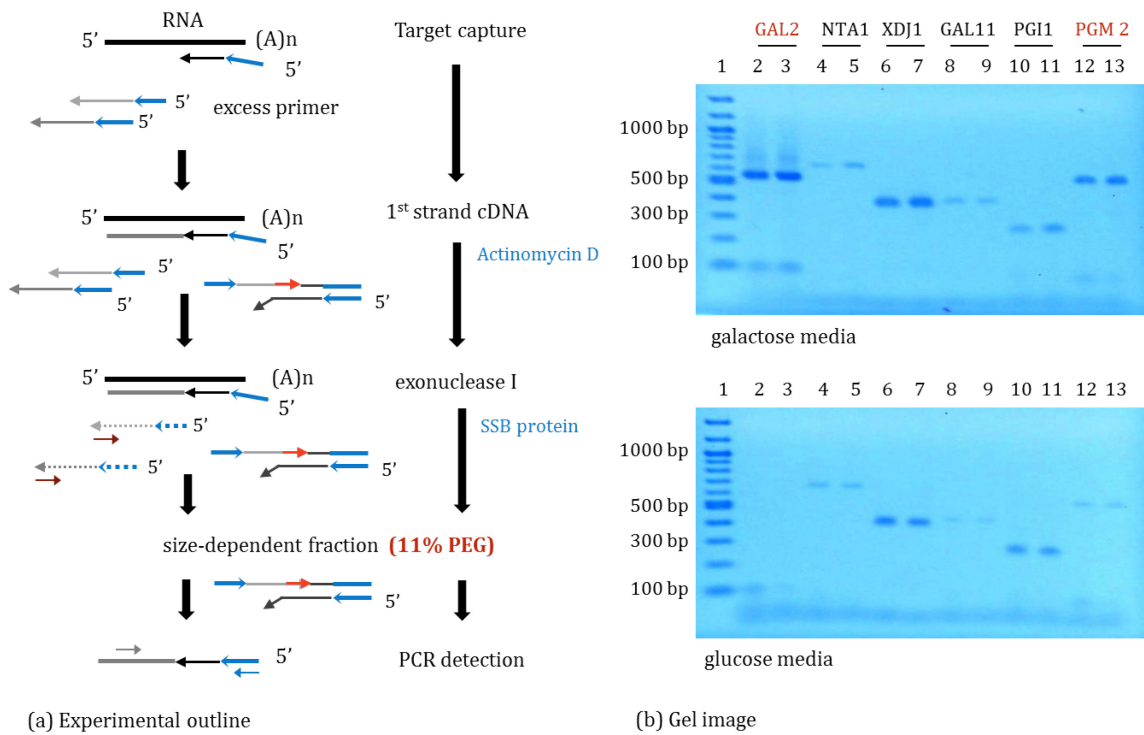


Figure 3-22 Gene expression - exonuclease I and 11 % PEG 8000

A total of six transcripts are analyzed across duplicate RT reactions for each condition. GAL2 and PGM2 (biologically relevant) are expressed at much higher levels in galactose than glucose media. Even, after size fractionation, ds-primer is present in 6 of the 24 samples (near 100 bp band). This is a vast improvement over use of exonuclease I alone or exonuclease with spin column purification to remove excess long primers.

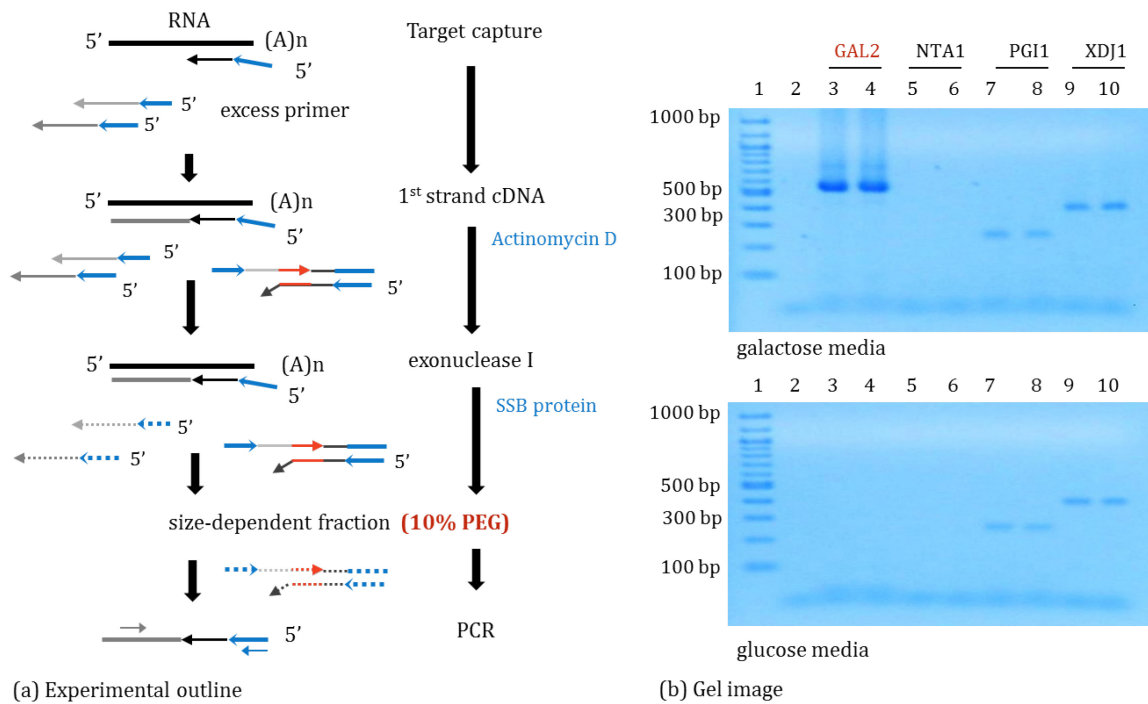


Figure 3-23 Gene expression – exonuclease I and 10 % PEG 8000

The above experiment is repeated with two modifications, i) target capture is done at 45°C for 6 h (instead of 12 h), and ii) fractionated cDNA and long primers with carboxylate beads suspended in 10% PEG 8000, 1.25 M NaCl solution. 1 µl of RT reaction is used for real-time PCR detection of 4 genes that are detected in previous experiment (GAL2, NTA1, PGI1 and XDJ1) [Figure 3-23]. Like above gel image GAL2 gene is detected in galactose media only (lane 3, 4) and controls genes XDJ1 (lane 9, 10) and PGI1 (lane 7, 8) have similar band intensity in galactose than glucose growth conditions. There is no ds-primer band (near 100 bp), although there is a smear band of PCR primer-dimer (not to be confused with ds-primer) below 100 bp DNA size marker. Primer-dimers are PCR artifacts and their formation is confirmed by amplification of 'no-template' control (lane 2).

We also did similar experiments, where exonuclease I digestion step is omitted from the reaction scheme. For all these experiments, we are able to detect gene expression difference between the two samples, but there are double-stranded primers. Two other methods to potentially remove double-stranded primers are, i) use of dsDNA specific exonuclease with 3'-5' or 5'-3' action in the exonuclease I mix, and / or ii) incorporating a restriction enzyme sequence immediately downstream of the universal primer sequence. For reference the construction of long primer is (5'-universal primer sequence-barcode-gene specific sequence-3').

Finally, the conditions and steps for first strand cDNA synthesis incorporating long primer library and reducing the factors that contribute to false positive signal are i) 100x long primer per mRNA species, ii) 1 µg of total RNA (30 pmol of primer), iii) 2 step hybridization temperature from high (50°C) to low (45°C) over 6 h, iv) 2 step first strand cDNA synthesis temperature (45°C, 50°C) over 2 h, v) addition of 300 ng/µl actinomycin D during reverse transcription, vi) high temperature (45°C) exonucleolytic hydrolysis of excess long primers (40 U

exonuclease I, 10 µg SSB protein, 2U antarctic phosphatase) for 30 min and vii) fractionation of RNA:DNA hybrids and excess primers with carboxylate beads in 10.0% PEG8000, 1.25 M NaCl.

3.4 Conclusion

In this chapter, we demonstrate the use of sequence-specific primer library to prime first strand cDNA synthesis of multiple transcripts. Two major challenges of using primer library are; i) the high concentrations of primers required to increase the detection sensitivity and accuracy, and ii) the removal of excess primer after the RT step.

The detection accuracy depends on the cDNA synthesis temperature of gene-specific sequence and concentration of primer to RNA molecule. The optimal hybridization and reverse temperature is 45°C for 30 nucleotide length sequence-specific primers. The concentration of primers must be at least 100 x excess to accurately quantify RNA levels.

The leftover excess primers show up as false positives that skew data analysis. Multiple methods are investigated to minimize false positives. The first method tested involved hydrolysis of single-stranded primers from 3' end by single-strand specific exonuclease (exonuclease I). Here, it is found that the amount of ssDNA digested by exonuclease I depends on the initial concentration of DNA, length and structure of DNA and the concentration of accessory SSB protein. Also, exonuclease I alone is only able to achieve 100- fold reduction in amount of ssDNA. Although this can be increased to million fold by addition of SSB protein, it is not enough to digest the high concentration of primers below detection levels. In addition the amount of ssDNA digested by exonuclease I depends on other parameters such as primer length, secondary structure and reaction conditions.

The high concentration of excess primers increases the probability of them being cross-hybridized. Any primers hybridized by their 3' end to other primers can be extended by reverse transcriptase to become double-stranded. The double-

stranded primers contribute to false-positive detection by both PCR and microarray technologies. These double-stranded primers are finally removed by size-fractionation of RT products on magnetic beads. The complete removal of ds-primer is dependent on the size-cutoff threshold which can be controlled by varying either one or both polyethylene glycol and sodium chloride concentration.

Finally, actinomycin D does not inhibit DNA-dependent DNA polymerase activity of reverse transcriptase. It possibly prevents hybridization of two DNA strands by binding at their 3' end and also prevents jumping of nascent DNA molecules to prime other RNA, thereby reducing false positives.

In conclusion, in experiment that use of sequence-specific primer libraries with upstream barcodes and universal primers necessitates additional steps to minimize factors that contribute to false positive signal for meaningful data interpretation and analysis.

CHAPTER 4

CONCLUSION AND RECOMMENDATIONS

4.1 Conclusion

In this dissertation, a robust method is presented to generate large quantities of single-stranded nucleic acids from DNA microarrays, and the application of these libraries to prime reverse transcription.

In chapter 2, we develop the process to improve low yields DNA libraries released from the microarray by processing them through a series of molecular biology steps. First, we observed that conventional PCR to amplify complex DNA library leads to formation of non-specific products. To solve that effect, DNA libraries are reproducibly amplified by separating them into micron sized water-in-oil (w/o) emulsion droplets. This helps to preserve the library complexity and sequence integrity. The power of emulsion PCR is realized by having 1 or less template per droplet. Towards this goal, the droplet size and stability after PCR cycling is assessed by confocal imaging and is found to be stable up to 35 cycles. The number of droplets is determined from droplet diameter and volume of aqueous phase and the amount of PCR amplified sequences is estimated by real-time PCR standard curve. Here, we demonstrated that the presence of multiple templates in emulsion droplets produces recombinant PCR products.

Second, we tackled the issue of removing the primer binding sequences (PBS) and formation of single-stranded DNA (ssDNA). For this, we tested many enzymatic methods to get highly-quality single-strand DNA. Only, *in vitro* transcription – reverse transcription method is conveniently able to produce large quantities of ssDNA to increase the sensitivity, reproducibility and perform significant number of downstream reactions. We demonstrated the use of two types of endonucleases that cut outside their recognition sequence to efficiently remove priming binding sequences. Restriction endonucleases cut dsDNA, whereas nicking endonucleases cuts ssDNA. These methods give a significant amount of flexibility to prepare functionally diverse ssDNA templates for many applications.

Chapter 3 describes an integrated method to use sequence-specific primers for cDNA library construction. The sequence-specific portion of the primer libraries are designed by OligoArray2.0 software [254], followed by a 5' upstream identification tag and universal primer binding sequence. The use of primer libraries leads to formation of RT artifacts that mask cDNA levels. There are 3 types of non-specific products after RT reaction, i) chimeric cDNA molecules, ii) double-stranded primers, and iii) excess single-stranded primers. The first two are formed due to catalytic activity of reverse transcriptase and the latter is to increase the sensitivity. Many enzymatic and purification methods and their combinations are tested to remove these non-specific processes. Of these, the most efficient and reliable method is the use of actinomycin D during reverse transcription followed by exonucleolytic hydrolysis of ssDNA and size-fractionation to remove double-stranded primers.

To summarize, the observations and results detailed in the dissertation provide a toolbox of methods to prepare large quantities of diverse single-stranded nucleic acid that, improve the sensitivity and reproducibility of assays and can be

utilized for a variety of applications such as synthetic biology, targeted sequencing, protein engineering, biosensors and understanding basic biology.

4.2 Recommendations for future work

4.2.1 Under generation of single-strand nucleic acid libraries

Library diversity

The process of making large amounts of single-stranded DNA libraries from microarrays consists of 4 steps, i) microarray synthesis, ii) emulsion PCR, iii) *in vitro* transcription and iv) reverse transcription. The latter 3 steps use polymerase enzymes with multiple purification steps. There is possibility that some percentage of the oligonucleotides maybe lost during each of these steps. The presence or absence of an individual oligonucleotide strand can be determined at each stage of the process on microarray. Further the products after each of the steps can be sequenced by 'next-generation' sequencing to determine preferential hybridization of signal over others.

Alternative amplification methods

One promising technique to make large quantities of single-stranded DNA is rolling circling amplification [255]. Here, the DNA molecules are circularized by ligase enzyme. They are amplified using a common primer to produce a long single-strand polymer chain. The primer sequences can be removed binding complementary primer sequence and nicking endonuclease. Any product biases in the above reactions (RNA amplification, reverse transcription, rolling circling amplification) can be minimized by performing them in an emulsion reaction.

Another method is to perform on-chip amplification. This would entail reducing the steric hindrance from the array surface and neighboring oligonucleotides for polymerase accessibility and primer extension. Also, the PCR reaction has to be enclosed to avoid evaporation of small volume during thermocycling. After 1-2 rounds of solid-phase PCR, the product could be collected for emulsion PCR (tube or chip). Also, chip based emulsion PCR has been developed by RainDance Technologies [256].

Process automation

Another option is to automate the entire process from emulsion PCR to single-strand generation. For now, the use flammable solvents (diethyl ether and ethyl acetate) to break the emulsions make the process unsafe for bench-top automation. Methods can be explored to change the oil-surfactant formulation to use solutions compatible with purification kits (isobutanol, isopropanol) or one can use magnetic beads in the emulsion reaction [257]. There are many robotic systems compatible with 96 well plates to do nucleic acid purification and magnetic bead selection. One such method is outlined here. Once the oil-surfactant and aqueous phase are pooled into the well, the emulsion can be formed by sonication. This followed thermal cycling and breaking of emulsion with isopropanol washes compatible with most 96 well purification systems yields dsDNA in the wells. To this is added the *in vitro* transcription reaction components, followed by purification on plates. The reverse transcription is done with biotinylated primer which is then captured on beads, followed by sequential on-bead digestions to first cut and remove the 3' end primer binding sequence (PBS) and then the desired ssDNA molecule.

Other Application of diverse oligonucleotide libraries

DNA nanotechnology – Here nucleic acids are used as structural materials as opposed to carriers of genetic information. The field of DNA nanotechnology first conceptualized by Seeman N in 1982 [258] has fostered many subareas such as DNA origami, DNA computing and DNA robotics. The areas are based on Watson-Crick base-pairing rules.

In DNA origami, the single-stranded genomic DNA is used as a scaffold to create and build wide variety of nanostructures. Multiple short DNA fragments are bound along the length of the scaffold to create bends and curves to build 2D and 3D shapes such as tiles, basic geometric shapes, octahedron, hollow nanodevices, balls, boxes and polyhedral structures as reviewed by Li et al [259]. Similar to DNA, RNA strands can be used as building materials, although the latter employs tertiary interactions between structural motifs rather than annealing of DNA strands. Also complex and hybrid structures with multiple properties can be made by combining multiple materials (DNA, RNA, proteins, molecules) [259]. The use of RNA scaffolds have been described to control spatial organization of enzymes, protein, molecules within the cell [260]. ssDNA scaffolds have been used for ordered distribution of proteins, biomolecules to increase biological and chemical reactions.

In 2000, DNA strands were used to make molecular tweezers based on the principle of sequence complementarity. This mechanism along with DNA nanocages can be used to target-specific delivery of drugs [261]. Since then other environmental stimuli such as pH change, ionic strength, electric fields, light, nucleic acid hydrolysis and polymerization can be used to control DNA nanomachines [259].

DNA sensors – Here, DNA and RNA oligonucleotides can be bound to biological (nucleic acids, peptides, antibodies) and non-biological (biotin, gold,

silver, other metals and fluorescent dyes) to realize electrochemical [262], optical [263] and fluorescent sensors [264]. These sensors can be used for numerous applications ranging from detection of infectious pathogens, harmful metal ions [265] to screening of genetic disorders and cancer. Combined with microfluidic technology, they can be used to make handheld field testing devices [266, 267].

DNA and Computers – Currently computer chips are made using a ‘top down’ approach that involves etching of electronic components from large structures. Researches at IBM describe the use of lithography and etching techniques to control the formation of DNA origami structures. Here, DNA strands are used to guide integration of silicon nanowires and carbon nanotubes with conventional microcircuit fabrication [268]. This ‘bottom up’ approach will enable 6 nm distance between components (current 45 nm) making devices smaller, faster and power efficient.

DNA vaccines – Here, DNA oligonucleotide libraries can be used to engineer plasmid constructs to encode pathogen (viral, bacterial or parasitic) antigens. The libraries provide flexibility to enhance immunogenicity from codon optimization. The plasmid constructs (DNA vaccines), when administered produce non-functional pathogen proteins within host cells leading to a range of immunological responses [269, 270].

4.2.2 Under sequence-specific primer libraries

The next step is to verify and validate the protocol to detect greater number of cDNA per reaction. Here, the sequence-specific libraries will be used to prime a small population (~100) of high and low abundance RNA and analyze them on ‘next-generation’ RNA sequencing platform [271-273]. The efficiency of the protocol will be determined on the basis of coverage of 3’ and 5’ end and transcript

abundance. The scale of these reactions will be increased progressively to reproducibly identify and detect multiple splicing variants and mRNA fusions. [274, 275]. Abnormal splicing variations and their functional consequences in cancer cells is an active area of research [276]. Here, sequence-specific primers to selectively capture transcripts of interest will increase the fold coverage to call genotypes with high confidence [277].

APPENDIX

EXPERIMENTAL PROTOCOLS

1.1 From oligonucleotide microarray to single-stranded DNA library

1.1.1 Quantification

The oligonucleotide library released from the microarray surface has truncated DNA fragments and makes the concentration amplifiable material measured by UV spectrophotometry (absorbance 260 nm) is inaccurate. To quantify the amount of PCR amplifiable template, duplicate real-time PCR reactions are run on 10x template dilution series and compared against real-time PCR standard curve.

1. Prepare following real-time PCR master solution (minus template). Scale the solution for duplicate runs of 3 template dilutions i.e. 6 runs + 1 run (compensate for pipetting errors).

x μ l 0.3 μ M forward universal primer 1

x μ l 0.3 μ M reverse universal primer 2

10 μ l 5x PCR buffer (1x)

5 μ l 25 mM MgCl₂ (2.5 mM)

1 μ l 10 mM dNTP (0.2 mM)

1 μ l 0.1 mM SYTO13 dye (0.2 μ M)

x μ l nuclease-free water

0.5 μ l 5 U/ μ l GoTaq Flexi-polymerase

49 μ l Total

2. In separate 0.2 ml real-time PCR tubes add 1 µl of template i.e. 100 pg to 1 pg.
Leave tubes on ice.
3. Add 49 µl of master solution to each tube and close lid of tube.
4. Start the program with desired cycling conditions on real-time PCR thermocycler. Pause the thermocycler when the lid and block reach denaturation temperature.
5. Transfer the tubes to the thermocycler and resume the program. This simulates Hot-Start conditions and minimizes non-specific or primer-dimer products formed during initial slow start-up (~3 min 30 s for MJ research DNA engine fitted with Chromo4 real-time detector)
6. Cycling conditions:
 - a. Initial denaturation time (Td) = 94°C for 2 min
 - b. 30 cycles at Td = 94°C for 10 s, Ta = Tm - 2°C for 20 s, Te = 72°C for 20 s, plate read
 - c. Final extension time (Te) = 72°C for 5 min
 - d. Note: Ta = annealing temperature, Tm = melting temperature of primers, amplicon size = 100 to 150 bp (adjust annealing and extension time for >150 bp)
7. Calculate C(t) value for each template concentration of x-y plot of no. of cycles versus log₁₀(dye intensity) and average across duplicates.
8. Calculate the actual concentration (moles) of amplifiable template using equation below from standard curve (valid for amplicons of size 80 – 150 bp) for concentration between 10 femtomoles (10⁻¹⁵; c(t) = 5.9) to 100 zeptomoles (10⁻²¹; c(t) = 26.75) i.e. 6 orders of magnitude.

$$\text{Template concentration (moles)} = 10^{\left(\frac{C_T - 11.425}{-3.688}\right)}$$

9. Run melt curve on thermocycler to check for non-specific product.

1.2.1 Emulsion PCR (adapted from Williams et al. Nature Methods.

The oligonucleotide pool library is amplified in water-in-oil (w/o) emulsion based PCR reaction, whereby each template is isolated in a single PCR droplet reactor. This prevents formation of chimeric non-specific products i.e. no PCR biases.

1. Prepare oil surfactant mix at RT. Vortex for 1 min and put on ice for 15 - 30 min.

80 μ l 100% ABIL EM90 surfactant (4 % v/v)

1 μ l 100% Triton X100 (0.05 % v/v)

1919 μ l mineral oil (96 % v/v)

2 ml Total

2. Prepare aqueous phase (PCR mix) of the emulsion. Pipette reagents in order.

x μ l nuclease-free water

20 μ l 5x GC rich buffer (1x buffer, 1.5 mM MgCl₂)

2 μ l 50 mM MgCl₂ (1 mM)

x μ l 0.5 μ M forward universal primer 1

x μ l 0.5 μ M reverse universal primer 2

2 μ l 10 mM dNTP (0.2 mM)

x μ l template concentration (1 template / droplet; See note below
OR 10 ng)

5 μ l 20 μ g/ μ l BSA (1 μ g/ μ l)

2 μ l 2 U/ μ l Phusion Hot-Start DNA polymerase

100 μ l Total

Note: Determine the number of template molecules / droplet

$$\text{Number of droplets } (n) = \frac{\text{PCR mix vol. (ul)} * 10^9 * 3 * 7}{4 * 22 * (\text{droplet size } (\mu\text{m}) / 2)^3}$$

$$\frac{\text{template}}{\text{droplet}} = \frac{\text{conc. } \left(\frac{\text{moles}}{\text{ul}}\right) * \text{vol (ul)} * NA}{n}$$

Where, NA = Avogadro's Number ($6.022 * 10^{23}$)

3. For 10 μl aqueous phase, transfer 400 μl of oil surfactant mixture to 2 ml glass vial and add a 3x8mm pivot stir bar. Begin stirring at 1000 rpm on the magnetic stirrer.
4. Notes:
 - Use minimum 100 μl PCR mix to get enough products for downstream IVT reaction.
 - Maintain the ratio of aqueous: oil phase as 1:4 (v/v). Keep the volume of emulsion between 200 – 600 μl to ensure good mixing and average droplet size of 8 μm .
 - Place the tube at the center of the magnetic plate in a container (we use a lid from a box of pipette tips filled with ice) to make sure the stir bar rotates freely.
 - Add aqueous phase to oil surfactant mixture in a drop-wise manner (10 μl) over a period of 30 s. Continue stirring for 15 min to create w/o emulsion.
 - Distribute 50 μl aliquots of emulsion into separate PCR tubes or wells of a PCR plate.
5. Thermocycling conditions
 - a. Initial denaturation time (Td) = 98°C for 1 min
 - b. 30 cycles at Td = 98°C for 10 sec, Ta = Tm for 20 sec, Te = 72°C for 20 sec
 - c. Final extension time (Te) = 72°C for 5 min

6. Pool the emulsified PCR reactions in a 1.6 ml micro-centrifuge tube. Add 2 μ l of loading buffer (indicator – orange G/xylene cyanol, to identify water droplet in organic solvents).
7. Add mineral oil to a final volume of 800 μ l, vortex 30 sec and centrifuge at 13,000 g for 10 min at RT. Dispose of the upper (oil) phase.
8. Several extractions with an organic solvent will remove the remaining oil and surfactant from the emulsion and cause it to break. Perform the following extraction: add 1 ml of water-saturated diethyl ether, vortex the tube to break emulsion (1 min), spin (30 sec), and dispose of the upper (solvent) phase. Add 1 ml of water-saturated ethyl acetate, and then one more extraction with diethyl ether. The extraction with ethyl acetate is necessary to remove the ‘gel-like’ phase, which is formed when diethyl ether is mixed with ABIL EM 90. The ‘gel-like’ phase may contain amplicons, do not discard.
9. Remove the residual solvent by making a hole at the top of tube and spin evaporate the solvent in 45°C incubator or leave tubes open on a 37°C heat block for 10 min.
10. Purify 100 μ l of PCR volume with commercial purification kits (Qiagen MinElute PCR Kit).
11. Notes:
 - If precipitation is required, add up to 10 μ g of linear acrylamide, 0.1 vol 5 M NaCl and 2.5 vol 100 % ethanol. Store at -20°C for atleast 30 min and precipitate DNA. Do not precipitate with sodium acetate.
 - To avoid precipitation, perform second wash on Qiagen column with 80% ethanol solution instead of buffer PE wash.

1.3.1 T7 *in vitro* transcription

The amount of double-stranded product (100 bp) from 100 μ l is roughly 1-2 μ g or 15 – 30 pmol. Use 2 pm template per 20 μ l T7 IVT reaction. This yields approx. 1700 pm of single-strand RNA.

1. Use either AmpliScribe™ T7-Flash™ Transcription Kit (Epicentre Biotechnologies) or MEGAscript™ T7 Kit (Invitrogen).
2. Incubate the reaction at 37°C for 4 h.
3. Add 1 μ l of TURBO DNase, and mix well. Incubate at 37°C for 15 min.
4. Inactivate TURBO DNase with inactivation reagent provided.
5. Purify RNA on 2 separate RNeasy columns.
6. Quantify RNA and analyze on bioanalyzer (if possible) or on 1.0 % agarose gel.

1.4.1 Reverse transcription

The amount of complementary cDNA formed per 100 pmol RNA is 50 pmol. Of this only about 30 % (15 pmol) ssDNA is recovered after primer removal and gel purification. Affinity purification increases ssDNA recovery to 60 % (30 pmol).

1. Prepare the following mix to hybridize RNA library with RT primer

x μ l *in vitro* transcribed RNA library (100 pm)

x μ l RT (150 pm)

2 μ l 10 mM dNTP (1 mM)

10 μ l Total

2. Incubate the reaction at 65°C for 5 min. Quick chill on ice for 2 min.
3. Add 10 μ l of the following RT mix

4 μ l 5 X First-strand buffer (1 X)

2 μ l 0.1 M DTT (10 mM)

x μ l RNase-Free water

0.6 μ l 10 μ g/ μ l Actinomycin D (0.3 μ g/ μ l)*

1 μ l 20 U/ μ l SUPERase-In (20 U)

1 μ l 200 U/ μ l SuperScript II Reverse Transcriptase

10 μ l Total

* The use of actinomycin D in the reaction is optional

4. Incubate the reaction at 37°C for 3 h.

5. Incubate at 95°C for 5 min. Quick chill on ice.

then add 0.43 μ l (12.9 μ l) 1 M NaOH (0.3 M),

then mix and incubate at 65°C for 15 min,

then add equal volume of 1 M HCl (12.9 μ l) to neutralize the solution,

then add 44.2 μ l of nuclease-free water to bring total reaction volume 100 μ l,

then add 5 μ l 3 M sodium acetate pH 5.0 (0.3 M),

then add 2 μ l 5 μ g/ μ l linear acrylamide (10 μ g),

then add 2.5 x vol 100 % ethanol, and store at - 80°C for > 30 min.

1.5.1 3' end primer binding removal

Some reverse transcriptase enzymes add extra nucleotides at the end of 3' end of cDNA strand. These few extra nucleotides have to be removed to free the 3'-hydroxyl end of the mRNA-specific sequence.

1. Precipitate single-strand DNA. Re-suspend pellet in 20 μ l nuclease free water.

Quantify amount of ssDNA.

2. Prepare 100 μ l of the following primer hybridization mix.

x μ l ssDNA library (500 pm RT product)

x μ l *Nb.BtsI primer (1.5 x excess; 5'-GGGAGA(N)₆GCAGTG-3')##

10 μ l 10x NEB buffer 4 (1x)

x μ l nuclease free water

100 μ l Total

* Replace enzyme with BspQI restriction sequence (5'-GGGAGA(N)₆GCTCTTCNNNN-3'; enzyme is active at 50°C)

If using affinity purification, use following primer sequence for hybridization (5'-(dA)₃₀GGGAGA(N)₆GCAGTG-3' or 5'-(dA)₃₀GGGAGA(N)₆GCTCTTCNNNN-3')

3. Incubate the reaction at 65°C for 5 min. Quick bring temperature to 37°C for 2 min.

4. Add 100 µl of the following enzyme cocktail mix.

10 µl 10x NEB buffer 4 (1x)

2 µl 100 x NEB buffer buffer (2x)

x µl nuclease free water

20 µl 10 U/µl of Nb.BtsI (200 U)

100 µl Total

5. Incubate the reaction at 37°C for 2 h (* 50°C for 30 min)&&

&& If using affinity purification method, do not heat inactivate and proceed to affinity purification.

6. Inactivate enzyme at 80°C for 20 min.

7. If desired, add 0.1 vol 5 M NaCl and 2.5 vol 100 % ethanol. Store at -20°C for atleast 30 min and precipitate DNA. Re-suspend in 50µl

8. Add 50 µl 2x RNA loading dye (New England Biolabs). Heat denature at 95°C for 2 min.

9. Run on 6 - 8 % denaturing polyacrylamide gel for 3 – 4 h at 150 V.

10. Also load 2 µl of Quick-Load® Low Molecular Weight DNA Ladder (New England Biolabs) processed as in step 8 above.

&& Do not heat inactivate. Proceed directly proceed to affinity purification.

1.6.1 Affinity purification

Use 200 μ l of Dynabeads® MyOne™ Streptavidin C1 (Invitrogen) for every 750 pmol of biotinylated primer. This protocol is only valid for removal of primer binding sequences bound to complementary biotinylated primer.

1. Prepare beads according to manufacturer's protocol. Re-suspend streptavidin coated magnetic beads in 200 μ l of 2x B&W buffer (2 M NaCl, 10 mM Tris-HCl, 1 mM EDTA)
2. Add 200 μ l of restriction digestion product.
3. Incubate at RT for 15 min. Collect beads on magnetic stand.
4. Remove supernatant and save (200 μ l)
5. Wash beads with 200 μ l of nuclease free water. Add 200 μ l of nuclease free water to beads and incubate at 70°C for 30 sec. Put on ice on 30 sec.
6. Add 200 μ l of supernatant to beads.
7. Repeat steps 3 – 6, 3 more times i.e. 4 rounds of binding and removal of primers. Also save beads in 200 μ l 2x B&W buffer for may be additional 1-2 rounds of binding (see step 10 below).
8. Add equal volume of nuclease free water to supernatant (step 4) after 4 rounds of binding and 2.5 vol 100 % ethanol. Store at -20°C for at least 30 min and precipitate DNA.
9. Re-suspend DNA pellet in 30 μ l nuclease free water and clean up on centrispin 20 column.
10. Measure concentration on spectrophotometer and run same aliquot on gel to ensure clean ssDNA with no primer contamination. If primer contamination, repeat additional rounds of bead binding.

2.1 Construction of cDNA library with sequence-specific primer library

Gene-specific primers enable high specificity capture of known RNA sequences. For reproducible detection across multiple samples, it is important to have a high concentration of primers, generally, 2 - 10 pm / gene. This is not possible due to limited amount of primer obtained post PCR-IVT-RT reaction scheme and preferable use of same primer batch across replicate experiments and conditions compared for accurate differential gene-expression analysis.

2.1.1 Target capture

1. Note:

- a. Extract RNA samples using phenol: chloroform pH 4.7. Only collect 1/4th volume of upper phase.
- b. Purify sample on appropriate RNA purification kit.
- c. Remove genomic DNA using TurboDNase free kit. Follow manufacture's protocol except use 0.2 U DNase per 10 µg of RNA

2. Anneal the sequence-specific-tag primers to total RNA sample.

1 µl Total RNA sample (1 - 10 µg)

3. Add 4 µl of Gene-specific-tag primer Master Mix.

1 µl 100 mM Sodium phosphate buffer (20 mM; pH 8.0)

1 µl 1.5 M NaCl (0.3 M)

1 µl Gene-specific-tag primer library (10 - 100 pmol)

1 µl 20 U/ µl SUPERrase-In (20 U)

4 µl Total

4. Briefly vortex (1 s) and spin down sample. Transfer to thermocycler.

Incubate at 65°C for 5 min (lid at +5°C),

then incubate at 50°C for 1 h (lid at +5°C),

finally incubate at 45°C for 6 h (lid at +5°C).

5. Note: Ratio of Gene-specific-tag primer to mRNA
 - a. 5 (1) μg yeast total RNA \sim 250 (50) ng mRNA (5% mRNA)
 - b. Average size of yeast mRNA = 1.6 Kb and M.W. of ssRNA = (# nucleotides \times 320.5) + 159.0 = 512959
 - c. moles of yeast mRNA (250 ng) = $\text{g/M.W.} = 250^{-9} (\text{g}) / 512959 (\text{g/mole}) = 0.49 \text{ pmol}$
 - d. Primer concentration = 50 pmol
 - e. Ratio of sequence-specific primer: mRNA \sim 100:1 (for accurate detection). Do not go below 50:1 primer: mRNA ratio, if accurate quantification is required.

2.2.1 Reverse transcription

Actinomycin D inhibits template switching activity of reverse transcriptase. T4 gene 32 increases the activity of reverse transcriptase and increases the yield of full length cDNA by opening RNA secondary structure.

6. Add 14 μl of the following 1st-Strand cDNA Synthesis Master Mix.

4 μl 5 X First-strand buffer (1 X)
2 μl 0.1 M DTT (10 mM)
1 μl 10 mM dNTP (0.5 mM)
x μl RNase-Free water
0.6 μl 10 $\mu\text{g}/\mu\text{l}$ Actinomycin D (0.3 $\mu\text{g}/\mu\text{l}$)
0.4 μl 5 $\mu\text{g}/\mu\text{l}$ T4 gene 32 protein (0.1 $\mu\text{g}/\mu\text{l}$)
1 μl 20 U/ μl SUPERase-In (20 U)
14 μl Total

7. Incubate at 45°C for 5 min (lid at +5°C),
8. Add 1 μl 200 U/ μl SuperScript III Reverse Transcriptase
Incubate at 45°C for 30 min (lid at +5°C),

finally incubate at 50°C for 30 min (lid at +5°C).

2.2.3 Removal of excess sequence-specific- tag primers

9. Bring block and lid temperature to 45°C. Immediately add 10 µl of the following Exonuclease I mix.

3 µl 10 X Exonuclease I buffer (1 X)

1 µl 20 U/ µl SUPERrase-In (20 U)

5 µl 2 µg/µl SSB protein (10 µg)

2 µl 20 U/µl Exonuclease I (40 U)

x µl RNase-Free water

10 µl Total

10. Incubate at 45°C for 30 min,

then add 1.5 µl 0.5 M EDTA pH 8.0 (25 mM)**

Inactivate enzymes at 80°C for 20 min.

** Step 10 and 11 below (tested protocol) may be replaced by step 12 and 13

11. Purify RNA: cDNA hybrid products with 1.8 x vol (55 µl) AMPure beads (follow manufacture's protocol except use suspend replace original bead buffer with 1.25 M NaCl, 10 % PEG 8000 solution and incubate at RT for 10 min). Elute with 30 µl nuclease-free TE buffer.

12. Incubate at 95°C for 5 min. Quick chill on ice.

then add 0.43 V (12.9 µl) 1 M NaOH (0.3 M),

then incubate at 65°C for 15 min,

then add equal volume of 1 M HCl (12.9 µl) to neutralize the solution,

then add 44.2 µl of nuclease-free water to bring total reaction volume 100 µl,

then add 5 µl 3 M sodium acetate pH 5.0 (0.3 M),

then add 2 µl 5 µg/µl linear acrylamide (10 µg),

then add 2.5 x vol 100 % ethanol, and store at - 20°C for > 30 min.

13. Inactivate enzymes at 95°C for 5 min. Quick chill on ice (1 min)
then add 0.9 µl 10 M NaOH (0.3 M; Sigma-aldrich),
then incubate at 65°C for 15 min,
then add 0.75 µl 12 N HCl (0.3 M; Sigma-aldrich) to neutralize the solution,
then add 3.35 µl 1 M Tris-HCl pH 7.5 – 8.0 to bring reaction volume 35 µl,
14. Purify cDNA with 1.8 x vol (63 µl) AMPure beads (follow manufacture's protocol except use suspend replace original bead buffer with 1.25 M NaCl, 10 % PEG 8000 solution and incubate at RT for 10 min). Elute with 5 µl warm nuclease-free TE buffer.

Methods and suggestion to process cDNA library for downstream applications

1. Keep the size of cDNA library constructed at least 5 times greater i.e. >500 bases than the size of sequence-specific primer library (100 bases).
2. The most popular method is to add poly (dA) tail, followed by second strand synthesis, limited number of PCR cycles and agarose gel purification to recover >400 base-pairs amplicons [278-280].
3. Another method for second strand cDNA synthesis would be to using a sequencing primer tailed with a pre-designed library of random nonamers to target cDNA library of interest.
4. It is critical to keep reaction times constants across multiple experiments for comparison.

BIBLIOGRAPHY

1. Tian, J., et al., *Accurate multiplex gene synthesis from programmable DNA microchips*. Nature, 2004. **432**(7020): p. 1050.
2. Tian, J.D., et al., *Accurate multiplex gene synthesis from programmable DNA microchips*. Nature, 2004. **432**(7020): p. 1050-1054.
3. Kong, D.S., et al., *Parallel gene synthesis in a microfluidic device*. Nucleic Acids Research, 2007. **35**(8): p. e61.
4. Gibson, D.G., et al., *Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome*. Science, 2008. **319**(5867): p. 1215-1220.
5. Gibson, D.G., et al., *Chemical synthesis of the mouse mitochondrial genome*. Nat Meth. **7**(11): p. 901.
6. Dahl, F., et al., *Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments*. Nucleic Acids Research. **33**(8): p. e71.
7. Gnirke, A., et al., *Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing*. Nat Biotech, 2009. **27**(2): p. 182.
8. Myllykangas, S., et al., *Efficient targeted resequencing of human germline and cancer genomes by oligonucleotide-selective sequencing*. Nat Biotech. **29**(11): p. 1024.
9. Bassik, M.C., et al., *Rapid creation and quantitative monitoring of high coverage shRNA libraries*. Nat Meth, 2009. **6**(6): p. 443.
10. Albayrak, S., *Construction and Screening of Custom Peptide Libraries Derived from Synthetic DNA Pools*, 2011.
11. Saboulard, D., et al., *High-throughput site-directed mutagenesis using oligonucleotides synthesized on DNA chips*. Biotechniques, 2005. **39**(3): p. 363-8.
12. Wu, C.-t.B., MA, US), Church, George M. (Brookline, MA, US), Williams, Benjamin Richard (Seattle, WA, US), *Oligonucleotide Paints*, 2010, President and Fellows of Harvard College (Cambridge, MA, US): United States.
13. Isalan, M., et al., *Evolvability and hierarchy in rewired bacterial gene networks*. Nature, 2008. **452**(7189): p. 840-5.

14. Quan, J., et al., *Parallel on-chip gene synthesis and application to optimization of protein expression*. Nat Biotechnol, 2011. **29**(5): p. 449-52.
15. Kosuri, S., et al., *Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips*. Nat Biotechnol, 2010. **28**(12): p. 1295-9.
16. Martin, V.J., et al., *Engineering a mevalonate pathway in Escherichia coli for production of terpenoids*. Nat Biotechnol, 2003. **21**(7): p. 796-802.
17. Cello, J., A.V. Paul, and E. Wimmer, *Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template*. Science, 2002. **297**(5583): p. 1016-1018.
18. Smith, H.O., et al., *Generating a synthetic genome by whole genome assembly: phi X174 bacteriophage from synthetic oligonucleotides*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(26): p. 15440-15445.
19. Mckay, C.D., et al., *Molecular Codebreaking and Double Encoding - Laboratory Experiments*. International Journal of Unconventional Computing, 2009. **5**(6): p. 547-564.
20. Elbaz, J., et al., *DNA computing circuits using libraries of DNAzyme subunits*. Nature Nanotechnology, 2010. **5**(6): p. 417-422.
21. Yang, D., et al., *Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(15): p. 9942-9947.
22. Cerutti, H. and J.A. Casas-Mollano, *On the origin and functions of RNA-mediated silencing: from protists to man*. Current Genetics, 2006. **50**(2): p. 81-99.
23. Cleary, M.A., et al., *Production of complex nucleic acid libraries using highly parallel in situ oligonucleotide synthesis*. Nature Methods, 2004. **1**(3): p. 241-248.
24. Oleinikov, A.V., J. Zhao, and M.D. Gray, *RNA interference by mixtures of siRNAs prepared using custom oligonucleotide arrays*. Nucleic Acids Research, 2005. **33**(10).

25. Bassik, M.C., et al., *Rapid creation and quantitative monitoring of high coverage shRNA libraries*. Nature Methods, 2009. **6**(6): p. 443-U62.
26. Noren, K.A. and C.J. Noren, *Construction of high-complexity combinatorial phage display peptide libraries*. Methods, 2001. **23**(2): p. 169-178.
27. Lindner, T., et al., *DNA Libraries for the Construction of Phage Libraries: Statistical and Structural Requirements and Synthetic Methods*. Molecules, 2011. **16**(2): p. 1625-1641.
28. Derda, R., et al., *Diversity of Phage-Displayed Libraries of Peptides during Panning and Amplification*. Molecules, 2011. **16**(2): p. 1776-1803.
29. Katz, C., et al., *Studying protein-protein interactions using peptide arrays*. Chemical Society Reviews, 2011. **40**(5): p. 2131-2145.
30. Frank, R., *Spot-Synthesis - an Easy Technique for the Positionally Addressable, Parallel Chemical Synthesis on a Membrane Support*. Tetrahedron, 1992. **48**(42): p. 9217-9232.
31. Sobreira, N.L.M., et al., *Whole-Genome Sequencing of a Single Proband Together with Linkage Analysis Identifies a Mendelian Disease Gene*. Plos Genetics, 2010. **6**(6).
32. Lupski, J.R., et al., *Whole-Genome Sequencing in a Patient with Charcot-Marie-Tooth Neuropathy*. New England Journal of Medicine, 2010. **362**(13): p. 1181-1191.
33. Walsh, T., et al., *Whole Exome Sequencing and Homozygosity Mapping Identify Mutation in the Cell Polarity Protein GPM2 as the Cause of Nonsyndromic Hearing Loss DFNB82*. American Journal of Human Genetics, 2010. **87**(1): p. 90-94.
34. Turner, E.H., et al., *Massively parallel exon capture and library-free resequencing across 16 genomes*. Nature Methods, 2009. **6**(5): p. 315-6.
35. Kahvejian, A., J. Quackenbush, and J.F. Thompson, *What would you do if you could sequence everything?* Nat Biotechnol, 2008. **26**(10): p. 1125-1133.
36. Porreca, G.J., et al., *Multiplex amplification of large sets of human exons*. Nature Methods, 2007. **4**(11): p. 931-6.

37. Dahl, F., et al., *Multigene amplification and massively parallel sequencing for cancer mutation discovery*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(22): p. 9387-9392.
38. Gnirke, A., et al., *Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing*. Nat Biotechnol, 2009. **27**(2): p. 182-189.
39. Turner, E.H., et al., *Massively parallel exon capture and library-free resequencing across 16 genomes*. Nat Methods, 2009. **6**(5): p. 315-6.
40. Mamanova, L., et al., *Target-enrichment strategies for next-generation sequencing (vol 7, pg 111, 2010)*. Nature Methods, 2010. **7**(6): p. 479-479.
41. Coleman, J.R., et al., *MRNA-targeted fluorescent in situ hybridization (FISH) of gram-negative bacteria without template amplification or tyramide signal amplification*. Journal of Microbiological Methods, 2007. **71**(3): p. 246-255.
42. Wyart, M., D. Botstein, and N.S. Wingreen, *Evaluating Gene Expression Dynamics Using Pairwise RNA FISH Data*. Plos Computational Biology, 2010. **6**(11).
43. Gao, X., et al., *A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids*. Nucleic Acids Res, 2001. **29**(22): p. 4744-50.
44. Wilk, A., et al., *Deoxyribonucleoside phosphoramidites*. Curr Protoc Nucleic Acid Chem, 2001. **Chapter 2**: p. Unit 2 7.
45. Lockhart, D.J., et al., *Expression monitoring by hybridization to high-density oligonucleotide arrays*. Nat Biotechnol, 1996. **14**(13): p. 1675-80.
46. McGall, G.H., et al., *The Efficiency of Light-Directed Synthesis of DNA Arrays on Glass Substrates*. J Am Chem Soc, 1997. **119**(22): p. 5081-5090.
47. Nuwaysir, E.F., et al., *Gene expression analysis using oligonucleotide arrays produced by maskless photolithography*. Genome Res, 2002. **12**(11): p. 1749-55.
48. Albert, T.J., et al., *Light-directed 5'-->3' synthesis of complex oligonucleotide microarrays*. Nucleic Acids Res, 2003. **31**(7): p. e35.

49. LeProust, E.M., et al., *Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process*. Nucleic Acids Res, 2010. **38**(8): p. 2522-40.
50. Hughes, T.R., et al., *Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer*. Nat Biotechnol, 2001. **19**(4): p. 342-7.
51. Egeland, R.D. and E.M. Southern, *Electrochemically directed synthesis of oligonucleotides for DNA microarray fabrication*. Nucleic Acids Res, 2005. **33**(14): p. e125.
52. Gao, X., E. Gulari, and X. Zhou, *In situ synthesis of oligonucleotide microarrays*. Biopolymers, 2004. **73**(5): p. 579-96.
53. Zhou, X., et al., *Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences*. Nucleic Acids Res, 2004. **32**(18): p. 5409-17.
54. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. Science, 1995. **270**(5235): p. 467-70.
55. Baker, M., *Microarrays, megasynthesis*. Nature Methods, 2011. **8**(6): p. 457-460.
56. Schulze, A. and J. Downward, *Navigating gene expression using microarrays - a technology review*. Nature Cell Biology, 2001. **3**(8): p. E190-E195.
57. Shinawi, M. and S.W. Cheung, *The array CGH and its clinical applications*. Drug Discovery Today, 2008. **13**(17-18): p. 760-770.
58. Wang, L., R. Luhm, and M. Lei, *SNP and mutation analysis*. Microarray Technology and Cancer Gene Profiling, 2007. **593**: p. 105-116.
59. Kim, S. and A. Misra, *SNP genotyping: Technologies and biomedical applications*. Annual Review of Biomedical Engineering, 2007. **9**: p. 289-320.
60. Zhou, X.C., et al., *Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences*. Nucleic Acids Research, 2004. **32**(18): p. 5409-5417.

61. Borovkov, A.Y., et al., *High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides*. Nucleic Acids Research, 2010. **38**(19).
62. Matzas, M., et al., *High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing*. Nat Biotechnol, 2010. **28**(12): p. 1291-U101.
63. Gnirke, A., et al., *Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing*. Nature Biotechnology, 2009. **27**(2): p. 182-189.
64. Myllykangas, S., et al., *Efficient targeted resequencing of human germline and cancer genomes by oligonucleotide-selective sequencing*. Nat Biotechnol, 2011. **29**(11): p. 1024-U95.
65. Lee, W.H., *Gene Synthesis from Oligonucleotide Mixtures by Solid Phase PCR and Assembly PCR in a Microfluidic Chip System*, 2010.
66. Olejnik, J., E. Krzymanska-Olejnik, and K.J. Rothschild, *Photocleavable biotin phosphoramidite for 5'-end-labeling, affinity purification and phosphorylation of synthetic oligonucleotides*. Nucleic Acids Research, 1996. **24**(2): p. 361-6.
67. Kataja, K., et al., *A highly sensitive and multiplexed method for focused transcript analysis*. J Microbiol Methods, 2006. **67**(1): p. 102-13.
68. Shchepinov, M.S., S.C. CaseGreen, and E.M. Southern, *Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays*. Nucleic Acids Research, 1997. **25**(6): p. 1155-1161.
69. Travis, P.M., *Theory of emulsions and emulsification*. Journal of the American Oil Chemists Society, 1926. **3**(1): p. 5.
70. Becher, P., *Emulsions : theory and practice*. 3rd ed2001, Washington, D.C. Oxford; New York: American Chemical Society; Oxford University Press. viii, 513 p.
71. Binks, P.B., *Modern Aspects of Emulsion*. 1 ed1998: Royal Society of Chemistry.
72. Bancroft, W.D., *The theory of emulsification, V*. Journal of Physical Chemistry, 1913. **17**(6): p. 501-519.

73. Glasstone, S. and D. Lewis, *Elements of physical chemistry*. [2d ed 1960, Princeton, N.J.,: Van Nostrand. 758 p.
74. Cosgrove, T. and Wiley online library., *Colloid science principles, methods and applications*, 2005, Blackwell Pub.: Oxford, UK ; Ames, Iowa.
75. Verwey, E.J.W., *Theory of the Stability of Lyophobic Colloids*. Journal of Physical and Colloid Chemistry, 1947. **51**(3): p. 631-636.
76. Park, C.I., W.G. Cho, and S.J. Lee, *Emulsion stability of cosmetic creams based on water-in-oil high internal phase emulsions*. Korea-Australia Rheology Journal, 2003. **15**(3): p. 125-130.
77. Weiss, J., *Emulsion Stability Determination*, in *Current Protocols in Food Analytical Chemistry* 2001, John Wiley & Sons, Inc.
78. Sjöblom, J., *Emulsions and emulsion stability*. 2nd ed. Surfactant science series 2006, Boca Raton: Taylor & Francis. 668 p.
79. Koroleva, M.Y. and E.V. Yurtov, *Water mass transfer in W/O emulsions*. Journal of Colloid and Interface Science, 2006. **297**(2): p. 778-784.
80. Cohen, S.N., et al., *Construction of Biologically Functional Bacterial Plasmids in-Vitro*. Proc Natl Acad Sci U S A, 1973. **70**(11): p. 3240-3244.
81. Mullis, K., et al., *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction*. Cold Spring Harb Symp Quant Biol, 1986. **51 Pt 1**: p. 263-73.
82. Saiki, R.K., et al., *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase*. Science, 1988. **239**(4839): p. 487-91.
83. Wagner, A., et al., *Surveys of Gene Families Using Polymerase Chain-Reaction - Pcr Selection and Pcr Drift*. Systematic Biology, 1994. **43**(2): p. 250-261.
84. Polz, M.F. and C.M. Cavanaugh, *Bias in template-to-product ratios in multitemplate PCR*. Applied and Environmental Microbiology, 1998. **64**(10): p. 3724-3730.
85. Ishii, K. and M. Fukui, *Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR*. Applied and Environmental Microbiology, 2001. **67**(8): p. 3753-3755.

86. Suzuki, M.T. and S.J. Giovannoni, *Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR*. Applied and Environmental Microbiology, 1996. **62**(2): p. 625-630.
87. Kanagawa, T., *Bias and artifacts in multitemplate polymerase chain reactions (PCR)*. Journal of Bioscience and Bioengineering, 2003. **96**(4): p. 317-323.
88. Meyerhans, A., J.P. Vartanian, and S. Wainhobson, *DNA Recombination during Pcr*. Nucleic Acids Research, 1990. **18**(7): p. 1687-1691.
89. Fang, G.W., et al., *Minimizing DNA recombination during long RT-PCR*. Journal of Virological Methods, 1998. **76**(1-2): p. 139-148.
90. Thompson, J.R., L.A. Marcelino, and M.F. Polz, *Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'*. Nucleic Acids Research, 2002. **30**(9): p. 2083-2088.
91. Qiu, X.Y., et al., *Evaluation of PCR-generated chimeras: Mutations, and heteroduplexes with 16S rRNA gene-based cloning*. Applied and Environmental Microbiology, 2001. **67**(2): p. 880-887.
92. Judo, M.S., A.B. Wedel, and C. Wilson, *Stimulation and suppression of PCR-mediated recombination*. Nucleic Acids Research, 1998. **26**(7): p. 1819-25.
93. Kurata, S., et al., *Reevaluation and reduction of a PCR bias caused by reannealing of templates*. Appl Environ Microbiol, 2004. **70**(12): p. 7545-9.
94. Eckert, K.A. and T.A. Kunkel, *DNA polymerase fidelity and the polymerase chain reaction*. PCR Methods Appl, 1991. **1**(1): p. 17-24.
95. Williams, R., et al., *Amplification of complex gene libraries by emulsion PCR*. Nature Methods, 2006. **3**(7): p. 545-50.
96. Nakano, M., et al., *Single-molecule PCR using water-in-oil emulsion*. Journal of Biotechnology, 2003. **102**(2): p. 117-124.
97. Nakano, M., et al., *Single-molecule reverse transcription polymerase chain reaction using water-in-oil emulsion*. Journal of Bioscience and Bioengineering, 2005. **99**(3): p. 293-295.

98. Griffiths, A.D. and D.S. Tawfik, *Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization*. *Embo Journal*, 2003. **22**(1): p. 24-35.
99. Griffiths, J.S., et al., *A bacterial selection for the directed evolution of pyruvate aldolases*. *Bioorganic & Medicinal Chemistry*, 2004. **12**(15): p. 4067-4074.
100. Mastrobattista, E., et al., *High-throughput screening of enzyme libraries: In vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions*. *Chemistry & Biology*, 2005. **12**(12): p. 1291-1300.
101. Tawfik, D.S. and A.D. Griffiths, *Man-made cell-like compartments for molecular evolution*. *Nat Biotechnol*, 1998. **16**(7): p. 652-656.
102. Ghadessy, F.J. and P. Holliger, *A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system*. *Protein Engineering Design & Selection*, 2004. **17**(3): p. 201-204.
103. Taly, V., B.T. Kelly, and A.D. Griffiths, *Droplets as microreactors for high-throughput biology*. *Chembiochem*, 2007. **8**(3): p. 263-272.
104. Kelly, B.T., et al., *Miniaturizing chemistry and biology in microdroplets*. *Chemical Communications*, 2007(18): p. 1773-1788.
105. Jensen, K., *Chemical kinetics - Smaller, faster chemistry*. *Nature*, 1998. **393**(6687): p. 735-+.
106. Diehl, F., et al., *BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions*. *Nature Methods*, 2006. **3**(7): p. 551-559.
107. Dressman, D., et al., *Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations*. *Proc Natl Acad Sci U S A*, 2003. **100**(15): p. 8817-8822.
108. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors*. *Nature*, 2005. **437**(7057): p. 376-380.
109. Pandey, V., R.C. Nutter, and E. Prediger, *Applied Biosystems SOLiD™ System: Ligation-Based Sequencing*, in *Next Generation Genome Sequencing 2008*, Wiley-VCH Verlag GmbH & Co. KGaA. p. 29-42.

110. Shendure, J., et al., *Accurate multiplex polony sequencing of an evolved bacterial genome*. Science, 2005. **309**(5741): p. 1728-1732.
111. Shi, G., et al., *Fabrication and optimization of the multiplex PCR-based oligonucleotide microarray for detection of Neisseria gonorrhoeae, Chlamydia trachomatis and Ureaplasma urealyticum*. J Microbiol Methods, 2005. **62**(2): p. 245-256.
112. Porreca, G.J., et al., *Multiplex amplification of large sets of human exons*. Nat Methods, 2007. **4**(11): p. 931-6.
113. Kiialainen, A., et al., *Performance of microarray and liquid based capture methods for target enrichment for massively parallel sequencing and SNP discovery*. PLoS One, 2011. **6**(2): p. e16486.
114. Gnirke, A., et al., *Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing*. Nat Biotechnol, 2009. **27**(2): p. 182-9.
115. Avci-Adali, M., et al., *Upgrading SELEX technology by using lambda exonuclease digestion for single-stranded DNA generation*. Molecules, 2010. **15**(1): p. 1-11.
116. Kerr, C. and P.D. Sadowski, *Gene 6 Exonuclease of Bacteriophage T7 .2. Mechanism of Reaction*. Journal of Biological Chemistry, 1972. **247**(1): p. 311-&.
117. Rogers, S.G. and B. Weiss, *Cloning of the Exonuclease-III Gene of Escherichia-Coli*. Gene, 1980. **11**(3-4): p. 187-195.
118. Nikiforov, T.T., et al., *The use of phosphorothioate primers and exonuclease hydrolysis for the preparation of single-stranded PCR products and their detection by solid-phase hybridization*. PCR Methods Appl, 1994. **3**(5): p. 285-91.
119. Clark, D. and S. Henikoff, *Ordered deletions using exonuclease III*. Methods Mol Biol, 1994. **31**: p. 47-55.
120. Little, J.W., *An Exonuclease Induced by Bacteriophage Lambda .2. Nature of Enzymatic Reaction*. Journal of Biological Chemistry, 1967. **242**(4): p. 679-&.

121. Higuchi, R.G. and H. Ochman, *Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction*. Nucleic Acids Research, 1989. **17**(14): p. 5865.
122. Subramanian, K., et al., *The enzymatic basis of processivity in lambda exonuclease*. Nucleic Acids Research, 2003. **31**(6): p. 1585-1596.
123. Hannon, K., et al., *Synthesis of PCR-derived, single-stranded DNA probes suitable for in situ hybridization*. Anal Biochem, 1993. **212**(2): p. 421-7.
124. Krishnakumar, S., et al., *A comprehensive assay for targeted multiplex amplification of human DNA sequences*. Proc Natl Acad Sci U S A, 2008. **105**(27): p. 9296-301.
125. Boissinot, K., et al., *Rapid exonuclease digestion of PCR-amplified targets for improved microarray hybridization*. Clin Chem, 2007. **53**(11): p. 2020-2023.
126. Hsieh, K., Y. Xiao, and H. Tom Soh, *Electrochemical DNA detection via exonuclease and target-catalyzed transformation of surface-bound probes*. Langmuir, 2010. **26**(12): p. 10392-6.
127. Pierce, K.E., et al., *Linear-After-The-Exponential (LATE)-PCR: primer design criteria for high yields of specific single-stranded DNA and improved real-time detection*. Proc Natl Acad Sci U S A, 2005. **102**(24): p. 8609-14.
128. Yuan, D.S., et al., *Improved microarray methods for profiling the Yeast Knockout strain collection*. Nucleic Acids Research, 2005. **33**(12): p. e103.
129. Zhu, L.X., et al., *Multiplex asymmetric PCR-based oligonucleotide microarray for detection of drug resistance genes containing single mutations in Enterobacteriaceae*. Antimicrob Agents Chemother, 2007. **51**(10): p. 3707-13.
130. Pierce, K.E. and L.J. Wangh, *Linear-after-the-exponential polymerase chain reaction and allied technologies. Real-time detection strategies for rapid, reliable diagnosis from single cells*. Methods Mol Med, 2007. **132**: p. 65-85.
131. Shepard, A.R. and J.L. Rae, *Magnetic bead capture of cDNAs from double-stranded plasmid cDNA libraries*. Nucleic Acids Research, 1997. **25**(15): p. 3183-5.

132. Pourmand, N., et al., *Multiplex Pyrosequencing*. Nucleic Acids Research, 2002. **30**(7): p. e31.
133. Hornes, E. and L. Korsnes, *Magnetic DNA Hybridization Properties of Oligonucleotide Probes Attached to Superparamagnetic Beads and Their Use in the Isolation of Poly(a) Messenger-Rna from Eukaryotic Cells*. Genetic Analysis-Biomolecular Engineering, 1990. **7**(6): p. 145-150.
134. Meyer, M., et al., *From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing*. Nucleic Acids Research, 2008. **36**(1).
135. Baltimore, D., *Viral Rna-Dependent DNA Polymerase - Rna-Dependent DNA Polymerase in Virions of Rna Tumour Viruses*. Nature, 1970. **226**(5252): p. 1209-&.
136. Temin, H.M. and S. Mizutani, *Viral Rna-Dependent DNA Polymerase - Rna-Dependent DNA Polymerase in Virions of Rous Sarcoma Virus*. Nature, 1970. **226**(5252): p. 1211-&.
137. Milligan, J.F., et al., *Oligoribonucleotide Synthesis Using T7 Rna-Polymerase and Synthetic DNA Templates*. Nucleic Acids Research, 1987. **15**(21): p. 8783-8798.
138. Gudnason, H., et al., *Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature*. Nucleic Acids Research, 2007. **35**(19).
139. Hoffman, C.S., *Preparation of Yeast DNA*, in *Current Protocols in Molecular Biology*2001, John Wiley & Sons, Inc.
140. Sambrook, J. and D.W. Russell, *Isolation of DNA Fragments from Polyacrylamide Gels by the Crush and Soak Method*. Cold Spring Harbor Protocols, 2006. **2006**(1): p. pdb.prot2936.
141. Moore, D. and D. Dowhan, *Purification and Concentration of DNA from Aqueous Solutions*, in *Current Protocols in Molecular Biology*2001, John Wiley & Sons, Inc.
142. Shoemaker, D.D., et al., *Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy*. Nat Genet, 1996. **14**(4): p. 450-456.

143. Winzeler, E.A., et al., *Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis*. Science, 1999. **285**(5429): p. 901-6.
144. Giaever, G., et al., *Functional profiling of the Saccharomyces cerevisiae genome*. Nature, 2002. **418**(6896): p. 387-91.
145. Knoth, A., I. Sdherze, and G. Muschiolik, *Effect of lipid type on water-in-oil-emulsions stabilized by phosphatidylcholine-depleted lecithin and polyglycerol polyricinoleate*. European Journal of Lipid Science and Technology, 2005. **107**(12): p. 857-863.
146. Musyanovych, A., V. Mailander, and K. Landfester, *Miniemulsion droplets as single molecule nanoreactors for polymerase chain reaction*. Biomacromolecules, 2005. **6**(4): p. 1824-1828.
147. Yang, Z., A.M. Sismour, and S.A. Benner, *Nucleoside alpha-thiotriphosphates, polymerases and the exonuclease III analysis of oligonucleotides containing phosphorothioate linkages*. Nucleic Acids Research, 2007. **35**(9): p. 3118-27.
148. Holloway, B., et al., *An Exonuclease-Amplification Coupled Capture Technique Improves Detection of Pcr Product (Vol 21, Pg 3905, 1993)*. Nucleic Acids Research, 1993. **21**(21): p. 4990-4990.
149. Boissinot, K., et al., *Rapid exonuclease digestion of PCR-amplified targets for improved microarray hybridization*. Clin Chem, 2007. **53**(11): p. 2020-3.
150. Mitsis, P.G. and J.G. Kwagh, *Characterization of the interaction of lambda exonuclease with the ends of DNA*. Nucleic Acids Research, 1999. **27**(15): p. 3057-3063.
151. Choi, M.H., et al., *Rheology of Decamethylcyclopentasiloxane (cyclomethicone) W/O Emulsion System*. Macromolecular Research, 2009. **17**(12): p. 943-949.
152. Farah, M.A., et al., *Viscosity of water-in-oil emulsions: Variation with temperature and water volume fraction*. Journal of Petroleum Science and Engineering, 2005. **48**(3-4): p. 169-184.
153. Rangsansarid, J. and K. Fukada, *Factors affecting the stability of O/W emulsion in BSA solution: Stabilization by electrically neutral protein at high ionic strength*. Journal of Colloid and Interface Science, 2007. **316**(2): p. 779-786.

154. Shendure, J., et al., *Accurate multiplex polony sequencing of an evolved bacterial genome*. *Science*, 2005. **309**(5741): p. 1728-32.
155. Zhou, X., et al., *Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences*. *Nucleic Acids Research*, 2004. **32**(18): p. 5409-17.
156. Borovkov, A.Y., et al., *High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides*. *Nucleic Acids Research*, 2010. **38**(19): p. e180.
157. Verma, I.M., *The reverse transcriptase*. *Biochim Biophys Acta*, 1977. **473**(1): p. 1-38.
158. Innis, M.A., *PCR protocols: a guide to methods and applications* 1990: Academic Press.
159. King, N., *RT-PCR protocols*. *Methods in molecular biology*, 1064-3745 ;6302010, New York: Humana. xi, 341 p.
160. Antonov, J., et al., *Reliable gene expression measurements from degraded RNA by quantitative real-time PCR depend on short amplicons and a proper normalization*. *Laboratory Investigation*, 2005. **85**(8): p. 1040-1050.
161. Vermeulen, J., et al., *Measurable impact of RNA quality on gene expression results from quantitative PCR*. *Nucleic Acids Research*, 2011. **39**(9).
162. Popova, T., et al., *Effect of RNA quality on transcript intensity levels in microarray analysis of human post-mortem brain tissues*. *BMC Genomics*, 2008. **9**.
163. Levesque-Sergerie, J.P., et al., *Detection limits of several commercial reverse transcriptase enzymes: impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR*. *Bmc Molecular Biology*, 2007. **8**.
164. Tian, B., et al., *A large-scale analysis of mRNA polyadenylation of human and mouse genes*. *Nucleic Acids Research*, 2005. **33**(1): p. 201-12.
165. von Ahlfen, S., et al., *Determinants of RNA quality from FFPE samples*. *PLoS One*, 2007. **2**(12): p. e1261.

166. Nam, D.K., et al., *Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription*. Proc Natl Acad Sci U S A, 2002. **99**(9): p. 6152-6156.
167. Das, M., et al., *Full-length cDNAs: more than just reaching the ends*. Physiological Genomics, 2001. **6**(2): p. 57-80.
168. Vanness, J., I.H. Maxwell, and W.E. Hahn, *Complex Population of Nonpolyadenylated Messenger-Rna in Mouse-Brain*. Cell, 1979. **18**(4): p. 1341-1349.
169. Wang, S.M., et al., *Screening poly(dA/dT)(-) cDNAs for gene identification*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(8): p. 4162-4167.
170. Zhang, J.L. and C.D. Byrne, *Differential priming of RNA templates during cDNA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR*. Biochemical Journal, 1999. **337**: p. 231-241.
171. Stangegaard, M., I.H. Dufva, and M. Dufva, *Reverse transcription using random pentadecamer primers increases yield and quality of resulting cDNA*. Biotechniques, 2006. **40**(5): p. 649-657.
172. Srivastava, G.P., et al., *Homolog-specific PCR primer design for profiling splice variants*. Nucleic Acids Research, 2011. **39**(10): p. e69.
173. Vandenbroucke, II, et al., *Quantification of splice variants using real-time PCR*. Nucleic Acids Research, 2001. **29**(13): p. E68-8.
174. Frohman, M.A., M.K. Dush, and G.R. Martin, *Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer*. Proc Natl Acad Sci U S A, 1988. **85**(23): p. 8998-9002.
175. Costa, V., et al., *Uncovering the complexity of transcriptomes with RNA-Seq*. J Biomed Biotechnol, 2010. **2010**: p. 853916.
176. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*. Nat Rev Genet, 2009. **10**(1): p. 57-63.
177. Levin, J.Z., et al., *Comprehensive comparative analysis of strand-specific RNA sequencing methods*. Nature Methods, 2010. **7**(9): p. 709-15.

178. Jayaprakash, A.D., et al., *Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing*. Nucleic Acids Research, 2011. **39**(21): p. e141.
179. So, A.P., R.F. Turner, and C.A. Haynes, *Increasing the efficiency of SAGE adaptor ligation by directed ligation chemistry*. Nucleic Acids Research, 2004. **32**(12): p. e96.
180. Aird, D., et al., *Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries*. Genome Biol, 2011. **12**(2): p. R18.
181. Hansen, K.D., S.E. Brenner, and S. Dudoit, *Biases in Illumina transcriptome sequencing caused by random hexamer priming*. Nucleic Acids Research, 2010. **38**(12): p. e131.
182. Leparc, G.G. and R.D. Mitra, *A sensitive procedure to detect alternatively spliced mRNA in pooled-tissue samples*. Nucleic Acids Res, 2007. **35**(21): p. e146.
183. Mamanova, L., et al., *Target-enrichment strategies for next-generation sequencing*. Nature Methods, 2010. **7**(2): p. 111-8.
184. Katz, R.A. and A.M. Skalka, *The retroviral enzymes*. Annu Rev Biochem, 1994. **63**: p. 133-73.
185. Marquet, R., et al., *tRNAs as primer of reverse transcriptases*. Biochimie, 1995. **77**(1-2): p. 113-24.
186. Mak, J. and L. Kleiman, *Primer tRNAs for reverse transcription*. Journal of Virology, 1997. **71**(11): p. 8087-95.
187. Bowman, E.H., V.K. Pathak, and W.S. Hu, *Efficient initiation and strand transfer of polypurine tract-primed plus-strand DNA prevent strand transfer of internally initiated plus-strand DNAs*. Journal of Virology, 1996. **70**(3): p. 1687-1694.
188. DeStefano, J.J., et al., *Polymerization and RNase H activities of the reverse transcriptases from avian myeloblastosis, human immunodeficiency, and Moloney murine leukemia viruses are functionally uncoupled*. Journal of Biological Chemistry, 1991. **266**(12): p. 7423-31.

189. Temin, H.M., *Reverse Transcription in the Eukaryotic Genome - Retroviruses, Pararetroviruses, Retrotransposons, and Retrotranscripts*. Molecular Biology and Evolution, 1985. **2**(6): p. 455-468.
190. Mourier, T., *Reverse transcription in genome evolution*. Cytogenetic and Genome Research, 2005. **110**(1-4): p. 56-62.
191. VanGuilder, H.D., K.E. Vrana, and W.M. Freeman, *Twenty-five years of quantitative PCR for gene expression analysis*. Biotechniques, 2008. **44**(5): p. 619-26.
192. Huber, H.E., et al., *Human immunodeficiency virus 1 reverse transcriptase. Template binding, processivity, strand displacement synthesis, and template switching*. Journal of Biological Chemistry, 1989. **264**(8): p. 4669-78.
193. Gerard, G.F. and J.M. D'Alessio, *Reverse Transcriptase (EC 2.7.7.49) Enzymes of Molecular Biology*, M.M. Burrell, Editor 1993, Humana Press. p. 73-93.
194. Spiegelman, S., et al., *Synthetic DNA-RNA hybrids and RNA-RNA duplexes as templates for the polymerases of the oncogenic RNA viruses*. Nature, 1970. **228**(5270): p. 430-2.
195. Spiegelman, S., et al., *DNA-directed DNA polymerase activity in oncogenic RNA viruses*. Nature, 1970. **227**(5262): p. 1029-31.
196. Spiegelman, S., et al., *Characterization of the products of DNA-directed DNA polymerases in oncogenic RNA viruses*. Nature, 1970. **227**(5258): p. 563-7.
197. Mizutani, S., D. Boettiger, and H.M. Temin, *A DNA-dependent DNA polymerase and a DNA endonuclease in virions of Rous sarcoma virus*. Nature, 1970. **228**(5270): p. 424-7.
198. Shaharabany, M. and A. Hizi, *The DNA-dependent and RNA-dependent DNA polymerase activities of the reverse transcriptases of human immunodeficiency viruses types 1 and 2*. AIDS Res Hum Retroviruses, 1991. **7**(11): p. 883-8.
199. Fisher, T.S., T. Darden, and V.R. Prasad, *Mutations proximal to the minor groove-binding track of human immunodeficiency virus type 1 reverse transcriptase differentially affect utilization of RNA versus DNA as template*. Journal of Virology, 2003. **77**(10): p. 5837-45.

200. Champoux, J.J. and S.J. Schultz, *Ribonuclease H: properties, substrate specificity and roles in retroviral reverse transcription*. FEBS J, 2009. **276**(6): p. 1506-16.
201. Schultz, S.J. and J.J. Champoux, *RNase H activity: structure, specificity, and function in reverse transcription*. Virus Res, 2008. **134**(1-2): p. 86-103.
202. Berger, S.L., et al., *Reverse transcriptase and its associated ribonuclease H: interplay of two enzyme activities controls the yield of single-stranded complementary deoxyribonucleic acid*. Biochemistry, 1983. **22**(10): p. 2365-72.
203. Das, M., et al., *Full-length cDNAs: more than just reaching the ends*. Physiol Genomics, 2001. **6**(2): p. 57-80.
204. Kotewicz, M.L., et al., *Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity*. Nucleic Acids Research, 1988. **16**(1): p. 265-77.
205. Coffin, J.M., S.H. Hughes, and H. Varmus, *Retroviruses* 1997, Plainview, N.Y.: Cold Spring Harbor Laboratory Press. xv, 843 p.
206. Zhang, J.Y., et al., *Most retroviral recombinations occur during minus-strand DNA synthesis*. Journal of Virology, 2000. **74**(5): p. 2313-2322.
207. Hu, W.S. and H.M. Temin, *Retroviral Recombination and Reverse Transcription*. Science, 1990. **250**(4985): p. 1227-1233.
208. Goodrich, D.W. and P.H. Duesberg, *Retroviral Recombination during Reverse Transcription*. Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(6): p. 2052-2056.
209. Destefano, J.J., et al., *Parameters That Influence Processive Synthesis and Site-Specific Termination by Human-Immunodeficiency-Virus Reverse-Transcriptase on Rna and DNA Templates*. Biochim Biophys Acta, 1992. **1131**(3): p. 270-280.
210. Gerard, G.F. and J.M. D'Alessio, *Reverse transcriptase (EC 2.7.7.49): the use of cloned maloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA*. Methods Mol Biol, 1993. **16**: p. 73-93.

211. Fry, M. and L.A. Loeb, *A DNA Polymerase-Alpha Pause Site Is a Hot-Spot for Nucleotide Misinsertion*. Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(2): p. 763-767.
212. Wu, W.M., et al., *Strand Transfer Mediated by Human-Immunodeficiency-Virus Reverse-Transcriptase in-Vitro Is Promoted by Pausing and Results in Misincorporation*. Journal of Biological Chemistry, 1995. **270**(1): p. 325-332.
213. Mikkelsen, J.G., et al., *Mutations of the kissing-loop dimerization sequence influence the site specificity of murine leukemia virus recombination in vivo*. Journal of Virology, 2000. **74**(2): p. 600-610.
214. Kim, J.K., et al., *Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1*. Journal of Biological Chemistry, 1997. **272**(27): p. 16769-16777.
215. Patel, P.H. and B.D. Preston, *Generation of Highly Variant DNA Termini during Polymerization by Hiv-1 Reverse-Transcriptase*. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 1991. **4**(3): p. 344-344.
216. Peliska, J.A. and S.J. Benkovic, *Mechanism of DNA Strand Transfer-Reactions Catalyzed by Hiv-1 Reverse-Transcriptase*. Science, 1992. **258**(5085): p. 1112-1118.
217. Patel, P.H. and B.D. Preston, *Marked Infidelity of Human-Immunodeficiency-Virus Type-1 Reverse-Transcriptase at Rna and DNA-Template Ends*. Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(2): p. 549-553.
218. Chen, D. and J.T. Patton, *Reverse transcriptase adds nontemplated nucleotides to cDNAs during 5'-RACE and primer extension*. Biotechniques, 2001. **30**(3): p. 574-80, 582.
219. Brooks, E.M., L.G. Sheflin, and S.W. Spaulding, *Secondary structure in the 3' UTR of EGF and the choice of reverse transcriptases affect the detection of message diversity by RT-PCR*. Biotechniques, 1995. **19**(5): p. 806-12, 814-5.
220. Stahlberg, A., et al., *Properties of the reverse transcription reaction in mRNA quantification*. Clinical Chemistry, 2004. **50**(3): p. 509-15.
221. Levesque-Sergerie, J.P., et al., *Detection limits of several commercial reverse transcriptase enzymes: impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR*. BMC Mol Biol, 2007. **8**.

222. Bustin, S.A. and T. Nolan, *Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction*. J Biomol Tech, 2004. **15**(3): p. 155-66.
223. Okello, J.B.A., et al., *Quantitative Assessment of the Sensitivity of Various Commercial Reverse Transcriptases Based on Armored HIV RNA*. PLoS One, 2010. **5**(11).
224. Stahlberg, A., M. Kubista, and M. Pfaffl, *Comparison of reverse transcriptases in gene expression analysis*. Clinical Chemistry, 2004. **50**(9): p. 1678-1680.
225. Ji, J.P. and L.A. Loeb, *Fidelity of HIV-1 reverse transcriptase copying RNA in vitro*. Biochemistry, 1992. **31**(4): p. 954-8.
226. Roberts, J.D., K. Bebenek, and T.A. Kunkel, *The accuracy of reverse transcriptase from HIV-1*. Science, 1988. **242**(4882): p. 1171-3.
227. Roberts, J.D., et al., *Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro*. Mol Cell Biol, 1989. **9**(2): p. 469-76.
228. Carninci, P., et al., *Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(2): p. 520-524.
229. Malboeuf, C.M., et al., *Thermal effects on reverse transcription: Improvement of accuracy and processivity in cDNA synthesis*. Biotechniques, 2001. **30**(5): p. 1074-+.
230. Chandler, D.P., C.A. Wagon, and H. Bolton, *Reverse transcriptase (RT) inhibition of PCR at low concentrations of template and its implications for quantitative RT-PCR*. Applied and Environmental Microbiology, 1998. **64**(2): p. 669-677.
231. Sellner, L.N., R.J. Coelen, and J.S. Mackenzie, *Reverse-Transcriptase Inhibits Taq Polymerase-Activity*. Nucleic Acids Research, 1992. **20**(7): p. 1487-1490.
232. Gerard, G.F., et al., *The role of template-primer in protection of reverse transcriptase from thermal inactivation*. Nucleic Acids Research, 2002. **30**(14): p. 3118-3129.

233. Arezi, B. and H. Hogrefe, *Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer*. *Nucleic Acids Research*, 2009. **37**(2): p. 473-81.
234. Bustin, S.A., *Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays*. *Journal of Molecular Endocrinology*, 2000. **25**(2): p. 169-193.
235. Piche, C. and J.P. Scherthaner, *Optimization of in vitro transcription and full-length cDNA synthesis using the T4 bacteriophage gene 32 protein*. *J Biomol Tech*, 2005. **16**(3): p. 239-47.
236. Collart, M.A. and S. Oliviero, *Preparation of Yeast RNA*, in *Current Protocols in Molecular Biology* 2001, John Wiley & Sons, Inc.
237. Molineux, I.J. and M.L. Gefter, *Properties of Escherichia-Coli DNA-Binding (Unwinding) Protein Interaction with Nucleolytic Enzymes and DNA*. *Journal of Molecular Biology*, 1975. **98**(4): p. 811-825.
238. Eggington, J.M., et al., *Polar destabilization of DNA duplexes with single-stranded overhangs by the Deinococcus radiodurans SSB protein*. *Biochemistry*, 2006. **45**(48): p. 14490-14502.
239. Ellington, A. and J. Chory, *Deprotection of oligonucleotides and purification using denaturing PAGE*. *Curr Protoc Immunol*, 2001. **Chapter 10**: p. Unit 10 7.
240. Lis, J.T. and R. Schleif, *Size fractionation of double-stranded DNA by precipitation with polyethylene glycol*. *Nucleic Acids Research*, 1975. **2**(3): p. 383-9.
241. Paithankar, K.R. and K.S. Prasad, *Precipitation of DNA by polyethylene glycol and ethanol*. *Nucleic Acids Research*, 1991. **19**(6): p. 1346.
242. Holland, M.J., *Transcript abundance in yeast varies over six orders of magnitude*. *Journal of Biological Chemistry*, 2002. **277**(17): p. 14363-14366.
243. Ringwald, M., et al., *The Mouse Gene Expression Database (GXD)*. *Nucleic Acids Research*, 2001. **29**(1): p. 98-101.
244. Flicek, P., et al., *Ensembl 2011*. *Nucleic Acids Research*, 2011. **39**(Database issue): p. D800-6.

245. Rouillard, J.M., M. Zuker, and E. Gulari, *OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach*. Nucleic Acids Research, 2003. **31**(12): p. 3057-3062.
246. Sobell, H.M., *Actinomycin and DNA transcription*. Proc Natl Acad Sci U S A, 1985. **82**(16): p. 5328-31.
247. McDonnell, J.P., et al., *DNA polymerases of Rous sarcoma virus: delineation of two reactions with actinomycin*. Nature, 1970. **228**(5270): p. 433-5.
248. Perocchi, F., et al., *Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D*. Nucleic Acids Research, 2007. **35**(19): p. e128.
249. Cherry, J.M., et al., *Genetic and physical maps of Saccharomyces cerevisiae*. Nature, 1997. **387**(6632 Suppl): p. 67-73.
250. Gasch, A.P., et al., *Genomic expression programs in the response of yeast cells to environmental changes*. Mol Biol Cell, 2000. **11**(12): p. 4241-57.
251. Hibbs, M.A., et al., *Exploring the functional landscape of gene expression: directed search of large microarray compendia*. Bioinformatics, 2007. **23**(20): p. 2692-9.
252. Bunte, T., et al., *Effect of actinomycin D on nucleic acid hybridization: the cause of erroneous DNA elongation during DNA synthesis of RNA tumor viruses in vitro*. Biochim Biophys Acta, 1980. **610**(2): p. 241-7.
253. Guo, J., et al., *Actinomycin D inhibits human immunodeficiency virus type 1 minus-strand transfer in in vitro and endogenous reverse transcriptase assays*. J Virol, 1998. **72**(8): p. 6716-24.
254. Rouillard, J.M., M. Zuker, and E. Gulari, *OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach*. Nucleic Acids Res, 2003. **31**(12): p. 3057-62.
255. Dean, F.B., et al., *Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification*. Genome Res, 2001. **11**(6): p. 1095-9.
256. Tewhey, R., et al., *Microdroplet-based PCR enrichment for large-scale targeted sequencing*. Nat Biotechnol, 2009. **27**(11): p. 1025-31.

257. Zhang, C. and D. Xing, *Single-molecule DNA amplification and analysis using microfluidics*. Chem Rev, 2010. **110**(8): p. 4910-47.
258. Seeman, N.C., *Nucleic acid junctions and lattices*. J Theor Biol, 1982. **99**(2): p. 237-47.
259. Li, H.Y., T.H. LaBean, and K.W. Leong, *Nucleic acid-based nanoengineering: novel structures for biomedical applications*. Interface Focus, 2011. **1**(5): p. 702-724.
260. Thodey, K. and C.D. Smolke, *Bringing It Together with RNA*. Science, 2011. **333**(6041): p. 412-413.
261. Yurke, B., et al., *A DNA-fuelled molecular machine made of DNA*. Nature, 2000. **406**(6796): p. 605-608.
262. Zhang, Q.D., et al., *Applications of carbon nanotubes to electrochemical DNA sensors: a new strategy to make direct and selective hybridization detection from SWNTs*. Advances in Natural Sciences: Nanoscience and Nanotechnology, 2010. **1**(4): p. 045011.
263. Li, M., et al., *Electrochemical and optical biosensors based on nanomaterials and nanostructures: a review*. Front Biosci (Schol Ed), 2011. **3**: p. 1308-31.
264. Dai, N. and E.T. Kool, *Fluorescent DNA-based enzyme sensors*. Chemical Society Reviews, 2011. **40**(12): p. 5756-5770.
265. Zhang, X.B., R.M. Kong, and Y. Lu, *Metal ion sensors based on DNazymes and related DNA molecules*. Annu Rev Anal Chem (Palo Alto Calif), 2011. **4**: p. 105-28.
266. Dutse, S.W. and N.A. Yusof, *Microfluidics-Based Lab-on-Chip Systems in DNA-Based Biosensing: An Overview*. Sensors, 2011. **11**(6): p. 5754-5768.
267. Wei, F., P.B. Lillehoj, and C.M. Ho, *DNA diagnostics: nanotechnology-enhanced electrochemical detection of nucleic acids*. Pediatr Res, 2010. **67**(5): p. 458-68.
268. Kershner, R.J., et al., *Placement and orientation of individual DNA shapes on lithographically patterned surfaces*. Nature Nanotechnology, 2009. **4**(9): p. 557-61.

269. Ingolotti, M., et al., *DNA vaccines for targeting bacterial infections*. Expert Rev Vaccines, 2010. **9**(7): p. 747-63.
270. Carvalho, J.A., et al., *DNA vaccines: a rational design against parasitic diseases*. Expert Rev Vaccines, 2010. **9**(2): p. 175-91.
271. Holland, M.J., *Transcript abundance in yeast varies over six orders of magnitude*. J Biol Chem, 2002. **277**(17): p. 14363-6.
272. Lee, A., et al., *Novel low abundance and transient RNAs in yeast revealed by tiling microarrays and ultra high-throughput sequencing are not conserved across closely related yeast species*. Plos Genetics, 2008. **4**(12): p. e1000299.
273. Mora, J.R. and R.C. Getts, *High-sensitivity detection methods for low-abundance RNA species: applications for functional genomics research*. Expert Rev Mol Diagn, 2007. **7**(6): p. 775-85.
274. Wang, L., et al., *Observations on novel splice junctions from RNA sequencing data*. Biochem Biophys Res Commun, 2011. **409**(2): p. 299-303.
275. Leparc, G.G. and R.D. Mitra, *A sensitive procedure to detect alternatively spliced mRNA in pooled-tissue samples*. Nucleic Acids Research, 2007. **35**(21): p. e146.
276. Levin, J.Z., et al., *Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts*. Genome Biol, 2009. **10**(10): p. R115.
277. Maher, C.A., et al., *Transcriptome sequencing to detect gene fusions in cancer*. Nature, 2009. **458**(7234): p. 97-101.
278. Tang, F., et al., *RNA-Seq analysis to capture the transcriptome landscape of a single cell*. Nat Protoc, 2010. **5**(3): p. 516-35.
279. Kurimoto, K., et al., *Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis*. Nat Protoc, 2007. **2**(3): p. 739-52.
280. Kurimoto, K. and M. Saitou, *A global single-cell cDNA amplification method for quantitative microarray analysis*. Methods Mol Biol, 2011. **687**: p. 91-111.