**Paired-End diTagging (PET) analysis for transcriptome characterization, genome annotation, and transcription factor binding site analysis**

Patrick Ng, Chia-Lin Wei and Yijun Ruan  
Genome Technology and Biology Group, Genome Institute of Singapore, 60 Biopolis Street #02-01,  
Singapore 138672, SINGAPORE.

Contact:  
1. Yijun Ruan, Ph.D.  
Email: ruanyj@gis.a-star.edu.sg  
2. Patrick Ng, Ph.D.  
Email: ngwp@gis.a-star.edu.sg

**Key Words**: PET, ditag, paired-end, mate-pair, SAGE, DNA sequencing, GIS-PET, ChIP-PET

**Introduction**

“Paired-End diTagging” (PET) analysis revolves around the concept of extracting 18-base “signatures” or “tags” from each of the 5’ and 3’ termini of any contiguous DNA sequence, and ligation them into Paired-End diTags (PETs) that are concatenated for enhanced sequencing efficiency. Each PET can subsequently be mapped onto the appropriate genome assembly to accurately define the location of the original fragment from which it was derived.

**Choice of protocol depending on application:**

(A) GIS-PET analysis

The Gene Identification Signature (GIS)-PET analysis method was developed to facilitate high-efficiency transcriptome analysis and genome annotation. It captures the 5’ and 3’ terminal signatures of full-length transcripts in the form of covalently-linked DNA paired-end ditags or “PETs”. Essentially, GIS-PET analysis is a two-part procedure: the first is the construction of a GIS full-length cDNA (flcDNA) library in a unique cloning vector (pGIS4) using a modified biotinylated CapTrapper approach, and the second is the construction of a GIS PET library containing concatenated PETs, for DNA sequencing. Concatenated PETs in bacterial clones are sequenced, and the PET sequences extracted and mapped onto assembled genomes for further analysis using a customized software suite (PET-Tool, contact Dr Yijun Ruan (ruanyj@gis.a-star.edu.sg) for access details). For ease of visualization, the PETs are displayed on a genome browser we developed in-house called the “T2G Browser”). Using GIS-PET, one can verify existing transcripts, validate gene predictions, identify splice variants and novel and/or unusual transcripts, and compare gene expression between different libraries. A major advantage of this procedure is the ready availability of the flcDNA plasmid library, from which transcripts of interest can be readily retrieved for downstream functional studies by simple PCR, using primers designed based on the PET sequences. The original published GIS-PET method involved the separate ligation of a 5’ adapter-mix and a 3’ adapter onto full-length cDNA, before cloning into vector pGIS1. This current variation of the protocol replaces pGIS1 with vector pGIS4, that already has the 3’ adapter sequences built into one of its “arms”, thereby reducing the number of steps required (Figure 1).

(B) ChIP-PET analysis

The same principle of paired-end ditagging can be applied to the efficient mapping of transcription factor binding sites (TFBS) identified by chromatin immunoprecipitation (ChIP). Here, randomly-sheared genomic DNA fragments that are first enriched for the TFBS of interest by ChIP are inserted into a specific cloning vector (pGIS3) and subjected to the same ditagging approach as in GIS-PET. We call this procedure ChIP-PET. Genuine TFBS are indicated by the overlapping of multiple distinct PET sequences on defined chromosomal loci.
Section (A): The Gene Identification Signature-Paired End diTagging (GIS-PET) analysis method

Materials list

GsuI-oligo dT primer (2 µg/µl)
RNasin-PLUS inhibitor (Promega)
Sample of interest: PolyA RNA (10-20µg)
Nuclease-free water
2x GC-I buffer (Takara) (available as part of the LA Taq kit)
10mM dNTP (with 5-Me-dCTP instead of dCTP) (see REAGENTS AND SOLUTIONS)
4.9M sorbitol
Superscript II reverse transcriptase (Invitrogen)
Saturated trehalose
Phenol/chloroform/IAA
3M NaOAc pH5.2
Absolute EtOH
10mM biotin hydrazide (long arm) (Vector Laboratories)
100mM sodium periodate, NaIO₄
1.1M NaOAc pH4.5
1M NaOAc pH 6.1
10% SDS
5M NaCl
Isopropanol
10x RNaseONE buffer (Promega)
RNaseONE (10U/ul) (Promega)
Yeast tRNA
M280 streptavidin Dynabeads (Dynal)
0.5M EDTA solution pH8.0
GlycoBlue (Ambion)
0.4ug/ul GIS-(N)5 adapter
0.4ug/ul GIS-(N)6 adapter
Takara Solution II (from Takara Ligation Kit 2.1)
Takara Solution I
T4 DNA ligase (5U/ul) (Invitrogen)
5x T4 DNA ligase buffer with PEG (Invitrogen)
Proteinase K (20mg/ml)
Deionized water
10x ExTaq buffer with Mg²⁺ (Takara)
2.5mM dNTP (Takara)
ExTaq polymerase (Takara)
10x buffer TANGO (Fermentas)
S-Adenosyl Methionine (SAM) (NEB; supplied with MmeI enzyme)
GsuI (5U/ul) (Fermentas)
NotI (10 U/ul) (NEB)
BseRI (4U/ul) (NEB)
BamHI (20U/ul) (NEB)
10x buffer B+ (Fermentas)
0.4ug/ul GIS-Sal adapter
10x NEBuffer 2 (NEB)
10x NEBuffer 3 (NEB)
10x unique BamHI buffer (NEB)
LoTE buffer (3mM Tris-HCl pH 7.5/ 0.2mM EDTA) (see REAGENTS AND SOLUTIONS)
Binding Buffer (see REAGENTS AND SOLUTIONS)
Blocking Buffer (see REAGENTS AND SOLUTIONS)
Wash Buffer (see REAGENTS AND SOLUTIONS)
TEN buffer for cDNA size fractionation equilibration and elution (1x buffer is: 10mM Tris-HCl pH 8.0; 0.1mM EDTA pH 8.0; 25mM NaCl) (see REAGENTS AND SOLUTIONS)
TEN buffer for oligonucleotide adapter annealing (2x buffer is: 20mM Tris-HCl pH8.0; 0.2mM EDTA; 100mM NaCl) (see REAGENTS AND SOLUTIONS)
SYBR Green I (Molecular Probes)
GelStar (Cambrex BioWhittaker)
pGIS4 cloning vector (obtain from authors)
20uM PMR011 PCR primer
20uM PMR012 PCR primer
HotStarTaq DNA polymerase system (Qiagen)
OneShot electrocompetent TOP10 cells (Invitrogen)
LB media, Lennox formulation with 5g/L NaCl
LB agar, Lennox formulation with 5g/L NaCl
Ampicillin
Zeocin (Invitrogen)
pZErO-1 cloning vector (Invitrogen)

Special equipment and expendables
Phase Lock Gel Light tubes (Eppendorf)
Precast 4-20% PAGE gels
Q-trays (Genetix) (or other large plates for bacterial plating)
1.5ml Siliconized microcentrifuge tubes
Heat/chill block (optional; thermal cycler can be substituted)
Benchtop hot/cold shaking incubator (Eppendorf)
Benchtop rotator (Grant-bio)
Notes and technical tips before starting

(i) The GIS-PET method described herein is for transcriptome analysis via fECDNA characterization. The application of the “Paired-End diTagging” (PET) concept in transcription-factor binding site (TFBS) identification and characterization via Chromatin Immunoprecipitation-PET (ChIP-PET) analysis is called “ChIP-PET” and is described in Section (B).

(ii) The cloning vector pGIS4 was derived from pGEM-3z (Promega) using site-directed mutagenesis and adapter ligation, to eliminate unwanted MmeI sites and to facilitate vector purification and usage. It is available upon request from the authors.

(iii) The terms “ditag” and “PET” in the context of this procedure are synonymous, and are used interchangeably within the text.

(iv) Do not use glycogen at any stage where it is not specifically mentioned, as this will interfere with the CapTrapper selection process.

(v) For all steps involving single-stranded DNA or RNA, it is preferable to use a siliconized or “lo-binding” 1.5ml or 1.7ml microfuge tube to avoid loss of nucleic acid.

(vi) We have found that using Eppendorf Phase Lock Gel tubes (2ml) is very convenient for the phenol/chloroform extraction steps, and reduces losses of aqueous sample. IMPORTANT! Phase Lock Gel should not be used after the reverse-transcription step, as the presence of trehalose and sorbitol increases the solution density to the point that the phases form below the gel layer, making recovery difficult.

(vii) PolyA RNA can be purified using any suitable method; we routinely purify polyA RNA from total RNA (extracted using Trizol reagent) using the µMACS mRNA Isolation Kit from Miltenyi Biotech as in our hand this gives the best combination of throughput and purity.

(viii) The original (published) GIS-PET method involves the separate ligation of a 5’ adapter-mix and a 3’ adapter onto full-length cDNA, before cloning into vector pGIS1 (Ng et al., 2005). This current variation of the protocol replaces pGIS1 with vector pGIS4, that already has the 3’ adapter sequences built into one of its “arms”, thereby reducing the number of steps required.
Figure 1. The current GIS-PET procedure, where the use of vector pGIS4 obviates 3'-adapter ligation.
GIS-PET Analysis- Part I of II (Construction of flcDNA library)

1. GIS full-length cDNA library construction

1-1 Reverse transcription
The following were mixed in a 0.2ml thin-walled PCR tube:
- 7µg GsuI-oligo dT primer (2 µg/µl)  
- RNasin-PLUS inhibitor (Promega)  
- PolyA RNA (10-20µg)  
- Nuclease-free water  

If necessary, ethanol precipitate the polyA RNA solution (without glycogen) to concentrate it prior to use. **Note**: Do not over-dry the RNA pellet, as it will be exceedingly difficult to resuspend.

The reaction mixture was heated to 65°C for 10min and cooled to 37°C for 1min, then held at 42°C while waiting for the other components to be prepared.

Set up on ice (to ensure no RT activity), in a 0.2ml thin-walled PCR tube:
- 2X GC-I buffer (Takara)  
- RNasin-PLUS inhibitor (Promega)  
- 10mM dNTP (with 5-Me-dCTP instead of dCTP)  
- 4.9M sorbitol (Sigma)  
- Superscript II reverse transcriptase (Invitrogen)  

**Note**: RNasin-PLUS inhibitor is preferred because of its wide-spectrum properties, and also its thermal stability.

Put saturated trehalose (made from trehalose powder (Sigma)) into another 0.2ml thin-walled PCR tube, and leave this warming at 42°C in a PCR block:
- Saturated trehalose  

When the oligo-dT/mRNA annealing step is complete, place the 2 pre-reaction mixtures above (in the 0.2ml PCR tubes) into the 42°C PCR block for 2min to preheat; mix the warm trehalose together with the other components (combined volume now = 131ul), and quickly transfer the entire reaction mix into the tube containing the annealed primer/mRNA (final volume = 151ul), and immediately start the incubation:
- 42°C for 40min; 50°C for 20 min; and 55°C for 20min.

Next, add 2 µl of proteinase K (20mg/ml stock; Ambion) to degrade all enzymes. The obtained solution is incubated at 45°C for 15 min, followed by phenol/chloroform (important: pH6.6, not 7.9) extraction, then re-extracted with another 150ul of nuclease-free water, and ethanol precipitated (Do NOT add glycogen):
- Aqueous layer containing RNA and DNA  
- 3M NaOAc pH5.2  
- Abs EtOH  

Keep at -80°C, 30min, then spin in Eppendorf microfuge at 13K RPM, 4°C, 30min; wash 1x with 75% EtOH. You should see a sizeable whitish pellet that has to be resuspended in 44.5ul nuclease-free water.
1-2 Oxidation of diol structures
Prepare the following stocks FRESH each time, using 1.7ml tubes:
10mM biotin hydrazide (long arm) (Vector Laboratories)
100mM sodium periodate, NaIO₄ (Sigma)

In a 1.7ml siliconized tube, add the following:
(-)DNA/RNA heteroduplex 44.5µl
1.1M NaOAc pH4.5 3ul
fresh 100mM NaIO₄ 2.5ul

Incubate 45min in the dark, in ice.

Then add the following to the 50ul reaction to precipitate the (-)DNA/RNA heteroduplex:
10% SDS 0.5ul
5M NaCl 11ul
Isopropanol 61ul
Keep at -80°C, 30min, then spin 13K RPM, 4°C, 30min; wash 1x with 75% EtOH. Resuspend the small whitish pellet in 50ul nuclease-free water.

1-3 Biotinylation
To the 50ul oxidized (-)DNA/RNA, add the following:
1M NaOAc pH 6.1 5ul
10% SDS 5ul
fresh 10mM biotin hydrazide 150ul
Incubate at room temperature, overnight (12-16hrs), in the dark.

1-4 RNaseONE selection
Precipitate the biotinylated (-)DNA/RNA heteroduplex: 210ul
5M NaCl 5ul
1M NaOAc pH 6.1 75ul
Abs EtOH 750ul
Keep at -80°C, 30min, then spin in Eppendorf microfuge at 13K RPM, 4°C, 30min; wash 1x with 75% EtOH. Resuspend the small whitish pellet in 170ul nuclease-free water.

Use approximately 5 units of RNaseONE (Promega) per µg starting mRNA:
biotinylated (-)DNA/RNA heteroduplex 170ul
10x RNaseONE buffer 20ul
RNaseONE (10U/ul) 4.5ul
Nuclease-free water 5.5ul
Incubate at 37°C, 30min, then quench the reaction by adding:
10mg/ml yeast tRNA 4ul
5M NaCl 50ul

Meanwhile, prepare Dynal M280 streptavidin Dynabeads. Plan to use 200ul of the beads suspension per RNA sample. You should be familiar with the use of the magnetic stand. Use siliconized or lo-binding tubes for all steps involving M280 beads.

Wash 1-3 200ul 1x Binding Buffer (BB; 2M NaCl, 50mM EDTA pH 8.0) ;Repeat 2x, room temp
Wash 4 200ul 1x BB + yeast tRNA at final conc. 0.25ug/ul ;4°C, 30min, using Eppendorf shaker, 800rpm
Wash 5-7 200ul 1x Binding Buffer (BB) ;Repeat 2x, room temp
1-5 Binding of full-length biotinylated (-)DNA/RNA heteroduplex
Remove supernatant from beads, add in the approx. 254ul RNaseONE-treated (-)DNA/RNA;
Rotate 30min at RT for binding to occur using an appropriate rotater.

| Wash 1-2  | 200ul | 1x Binding Buffer (BB) ;Repeat 1x, room temp |
| Wash 3    | 200ul | 1x Blocking Buffer (0.4% SDS + 50ug/ml yeast tRNA) |
| Wash 4    | 200ul | 1x Wash Buffer (10mM Tris-HCl pH 7.5, 0.2mM EDTA, 10mM NaCl, 20% glycerol, 40ug/ml yeast tRNA) |
| Wash 5    | 200ul | 50ug/ml yeast tRNA |

1-6 Hydrolytic degradation of bound RNA to release minus (-) strand flcDNA
**Note**: Use freshly made alkaline hydrolysis buffer (50mM NaOH/ 5mM EDTA pH 8.0)

Remove the supernatant (wash solution) from the beads, and add 50ul alkaline hydrolysis buffer;
Shake at 65°C for 10min using the Eppendorf shaker unit at 1400 rpm.
Collect the supernatant containing full-length (-)cDNA into a tube containing 50ul of 1M Tris-HCl pH 7.5
for neutralization.
Repeat the hydrolysis and collection steps another 2x, for a final volume of 300ul.

Do a phenol/chloroform (pH 7.9) extraction and ethanol precipitation to recover the (-) flcDNA:

| full-length (-) cDNA  | 300ul |
| 3M NaOAc pH 5.2       | 30ul  |
| GlycoBlue (Ambion)    | 4ul   |
| Abs EtOH              | 800ul |

Keep at -80°C, 30min, then spin 13K RPM, 4°C, 30min; wash 1x with 75% EtOH. The pellet is barely visible. Resuspend in 5ul LoTE (3mM Tris-HCl pH 7.5/ 0.2mM EDTA) or simply use 10mM Tris pH8.5.

1-7 Synthesis of double-stranded cDNA ready for cloning
In this series of steps, the (-) flcDNA will be converted to double-stranded (ds) cDNA by 5’ adapter ligation (the 5’ adapter is actually a mix of 2 very similar adapters, GIS-(N)5 and GIS-(N)6) followed by primer extension. GsuI is used to remove the residual polyA tail, leaving an AA dinucleotide residue; this is followed by a NotI digestion to form the 5’ cohesive site, resulting in ds flcDNA that can be directionally inserted into prepared pGIS4 (see step 1-7-3 and Appendix (C)).

**Notes:**
(i) Although BpmI is an isoschizomer of GsuI, it should NOT be used because it is insensitive to methylation (and may therefore cut within the insert).
(ii) Size-fractionation using gel-filtration columns is performed to remove all excess adapters, enzymes and small digestion products prior to insertion into the selected cloning vector.
(iii) To minimize sample losses, there are no intermediate adapter-removal steps until the cDNA size-fractionation at step 1-7-4.
1-7-1 Single-stranded linker (SSL) ligation (of mixed 5’ adapters)

Set up on ice using a 1.7ml microfuge tube:
- Full-length (-) cDNA 5ul
- 0.4ug/ul GIS-(N)5 adapter 4ul
- 0.4ug/ul GIS-(N)6 adapter 1ul
- Takara Solution II 10ul
- Takara Solution I (ligase) 20ul

Incubate 16 °C overnight (12-16hrs)

1-7-2 Primer extension

Set up on ice, using thin-walled PCR tube:
- The overnight ligation reaction 40ul
- Deionized water 20ul
- 10x ExTaq buffer with Mg²⁺ (Takara) 8ul
- 2.5mM dNTP (Takara) 8ul
- ExTaq polymerase (Takara) 4ul

Straight from ice, to 65°C preheated thermal block, and start incubation:
65°C, 5min; 68°C, 30min; 72°C, 10min --- and hold at 4°C

Add 2ul Proteinase K (20mg/ml) to remove any remaining DNA polymerase, and incubate 45°C, 15min.

Adjust volume to 200ul with deionized water, perform phenol/chloroform extraction (pH 7.9) and ethanol precipitate (with GlycoBlue) at -80°C for 30min. Spin, wash and resuspend ds cDNA pellet in 70ul water—keep 5ul of to run on an agarose gel to check if ds cDNA synthesis was successful.

1-7-3. Removal of polyA tail and formation of NotI cohesive 5’ terminal site

Set up using 1.7ml microfuge tube:
- Resuspended FL ds cDNA 65ul (5ul removed for control gel run)
- 10x buffer TANGO (Fermentas) 8.6ul
- 10x SAM 8.6ul (optional, but stimulates GsuI activity)
- GsuI (5U/ul) (Fermentas) 2ul
- Water 1.8

**Final volume 86ul**

Digest at 30°C (not 37°C) 4 hrs to overnight.

**Note**: Incomplete GsuI digest will lead to problems; colleagues have observed the entire oligo-dT stretch (from the oligo-dT primer used for reverse transcription) ligated to the 3’ adapter/ vector 3’ arm. This is probably due to a combination of the following: failed GsuI digest (and hence no removal of polyA tail), A-overhang after ExTaq extension, and degraded TT overhang from the 3’ adapter/ pGIS4 vector 3’ arm. It is critical to aliquot the validated, processed pGIS4 cloning vector and store each aliquot frozen.

Next, add more TANGO buffer to the reaction; this brings the buffer concentration to approx. 2x final (high salt). Use an excess of NotI (NotI is 20-50% active in 2x TANGO buffer):
- 10x buffer TANGO (Fermentas) 11.4ul
- NotI (10 U/ul) (NEB) 2.6ul

**Final volume 100ul**

Incubate at 37°C, 4 hr; then heat inactivate the enzymes at 65°C, 15min. Place tube on ice.
Proceed to step 1-7-4.

(Alternatively, do a sequential digestion with each enzyme, first with GsuI in Buffer B (Fermentas), followed by a spin-purification step to remove enzyme and buffer salts, then digest with NotI in NEBuffer 3 (NEB). This is immediately followed by cDNA size fractionation (Step 1-7-4 below)).
1-7-4 cDNA size fractionation prior to cloning
While doing the restriction digest, prepare Invitrogen cDNA size fractionation column (1 per sample) exactly as per the Invitrogen protocol.

Use column equilibrated to room temperature;
Remove top cap, then bottom, and allow to drain;
Add 0.8ml TEN (10mM Tris-HCl pH 8.0; 0.1mM EDTA pH 8.0; 25mM NaCl) buffer, allow to drain completely. Repeat 3x (this process will take an hour).
Label 20 x 1.7ml microfuge tubes to be used for fraction collection.

Add the 100ul digestion reaction (adjust volume with TEN buffer) into the prepared column;
Collect entire flowthru in collection tube 1.
Add 100ul TEN buffer, collect flowthru in tube 2.
Now add 100ul TEN buffer, and start collecting single drops, 1 drop (approx. 35ul) per tube. Allow column to drain completely before adding another 100ul TEN buffer each time.

At the end of the process, you will have 20 aliquots. Measure the absorbance of each fraction (use 10ul per fraction, in a 7ul submicro-volume cell (Amersham). Better yet, use a Nanodrop instrument, which only requires 1ul of sample, or use PicoGreen-based fluorimetry (Molecular Probes) (see Appendix (E)).

OPTIONAL (but useful): Remove 4ul from each of the 20 fractions, and run these on an agarose gel to assist in determining which fractions are suitable for cloning. Depending on yields and technique, ethidium bromide staining may be sufficient to observe the fractions, otherwise SYBR Green I (Molecular Probes) or GelStar (Cambrex Bio-Whittaker) stain can be used instead. We usually find that fractions 7-10 are suitable.

Selected fractions are pooled if desired, and ethanol precipitated (with GlycoBlue) to concentrate the ds flcDNA for ligation.

1-8 Ligation to cloning vector pGIS4 (See Appendix (C) for vector information)
Set up on ice, using 1.7ml microfuge tubes:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>40ng/ul Nol/BseRI cut pGIS4</td>
<td>1ul</td>
</tr>
<tr>
<td>Full-length ds cDNA fraction(s)</td>
<td>6ul</td>
</tr>
<tr>
<td>(minimum of 100ng, but preferably 200ng or more)</td>
<td></td>
</tr>
<tr>
<td>5x ligase buffer with PEG (Invitrogen)</td>
<td>2ul</td>
</tr>
<tr>
<td>5U/ul T4 DNA ligase</td>
<td>1ul</td>
</tr>
<tr>
<td><strong>Final volume</strong></td>
<td><strong>10ul</strong></td>
</tr>
</tbody>
</table>

Also set up a vector self-ligation control.

Incubate overnight (12-16hrs) at 16°C

1-9 Transformation by electroporation
Purify the ligation reaction to remove salts before electroporation: adjust volume to 200ul with deionized water, perform phenol/chloroform extraction (pH 7.9) and ethanol precipitate (with GlycoBlue) at -80°C for 30min. Spin, wash the pellet at least twice with 70% ethanol, and resuspend in 20ul Qiagen EB buffer.

Add 1ul of the purified ligation reaction into 25ul electrocompetent cells (we routinely use electrocompetent TOP10 (Invitrogen)) in pre-chilled 1.7ml microfuge tubes. Do not pipette up and down to mix; instead, gently stir with the pipette tip. Stand on ice for 5min, then transfer to pre-chilled Biorad electroporation cuvettes (0.1cm gap).

Add 1ml room temperature SOC media within 10sec of pulsing to maximize efficiency; transfer to 15ml Falcon tube (or equivalent), and recover at 37°C, 1hr, 200 rpm shaking.
Plate between 20-50ul (out of 1ml) on LBamp agar for quality control (QC) screening and library efficiency calculations. Incubate overnight at 37°C.

1-10 Library QC
Count the numbers of colonies and determine library efficiency taking into consideration the self-ligation background. Pick colonies (24-48 is sufficient and convenient) for PCR screening using primers PMR011 and PMR012, and agarose gel electrophoresis.

If the PCR shows a satisfactory range of insert sizes, pick 1 to 4x 96-well plates of colonies for overnight culture, plasmid purification and sequencing to determine full-length efficiency (by BLAST alignment of sequences against GenBank nr database). At this stage, the library can be stored in the form of purified ligation mix at -20°C, until one wants to proceed on to the second part of GIS analysis.

Optional: the quality of the flcDNA library can be assayed using a commercial cDNA integrity kit (KPL, Inc.). This set contains sets of primers that are designed to amplify a range of inserts from human and rodent species: abundant vs. rare, long vs. short. It is also important to sequence a sufficient number clones to get an idea of the full-length success rate in the library; it is critical that the full-length rate is high, in order to successfully perform the subsequent Part II of the GIS analysis method.
GIS-PET Analysis, Part II (ditagging)

2. GIS single-PET library construction

Overview:
The flcDNA clones constructed in Part I contain an MmeI site (TCCGAC) flanking the 5’ terminus of the cDNA insert, and another MmeI site on the minus strand, flanking the 3’ end of the insert. MmeI restriction enzyme will be used to cleave these clones 20bp/18bp into the cDNA fragments from both their 5’ and 3’ ends, thereby forming the 5’ and 3’ tags. Consequently, despite the variable sizes of the cDNA insert, the vector-plus-40bp cDNA signature tags (linear single-PET plasmids) will be of a constant size (approx. 2,800bp) that can be easily recognized upon agarose gel electrophoresis, and can be purified from the unwanted cDNA fragments that are a byproduct of the MmeI digestion.

In the next step, gel-purified linear single-PET plasmids will be blunted (to remove any incompatible 3’ overhangs), and self-ligated to give “single-PET plasmids” containing the 5’ and 3’ GIS signature tags now linked in the form of PETs. The entire pool of these single-ditag plasmids thus represents the GIS single-PET library.

2-1. Plasmid purification
The GIS flcDNA library is constructed by transforming all the remaining purified ligation mix, and amplified once by plating an appropriate number of clones (between 60,000 to 100,000 cfu/ tray; these numbers allow growth of each colony without excessive overcrowding) on large (22 x 22cm) agar plates (Genetix Q-trays). Use a maximum of 700ul of culture per Q-tray.

Notes:
(i) The number of colonies required is determined by the estimated transcriptome size; we routinely target 1x10^6 cfu as a convenient benchmark. (Basis of calculation: this number is estimated to give a 10x over-coverage of a transcriptome population of 100,000. The transcriptome size itself is estimated as 100,000 because 35% to 59% of the estimated 30,000 to 40,000 human genes undergo alternative splicing to give an average of 3 alternative transcripts per gene). Although only a small fraction of the expected milligram quantities of plasmid DNA obtained is subsequently used for MmeI digestion, it is still critical for the sake of proper representation/ complexity to first obtain the benchmark 1x10^6 cfu.
(ii) Growth on agar (solid-phase plasmid amplification) is important, as this maintains representation better than growth in liquid culture. This is because, on a solid substrate, each clone can grow without competition from other bacterial clones.
(iii) 10 Q-trays of bacteria produce about 1mg of plasmid DNA. Therefore, in theory 2 Qiagen Maxi tips (per library) should be sufficient. However, to avoid clogging due to the large volumes involved, we usually use 3-4 tips per library. Suggested volumes of buffers to use with bacteria from 10 Q-trays: 20 to 50ml each of P1, P2 and P3.

After an overnight 37°C incubation, the resulting bacterial colonies are harvested by manually scraping into LB media (20-30ml per Q-tray; a disposable plastic L-spreader is convenient), transferring to 500ml plastic centrifuge bottles and pelleted by centrifugation at ~4,000g for 20min in a Sorvall floor-standing ultracentrifuge (rotor SLA3000 is convenient). Plasmid DNA preparation is performed using the Qiagen HiSpeed Plasmid Maxi kit (or any other preferred system).
2-2. MmeI digestion (tagging)
Approximately 10µg of plasmid DNA was digested using MmeI as per the manufacturer’s conditions (NEB), ensuring that the number of units of enzyme used is always in less than 4-fold excess to prevent methylation-induced inhibition.

SAM should be prepared fresh, as it is unstable. The final SAM concentration should be about 50uM.

Suggested reaction conditions are:

<table>
<thead>
<tr>
<th>10µg plasmid DNA</th>
<th>100ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x NEBuffer 4</td>
<td>20ul</td>
</tr>
<tr>
<td>10x SAM</td>
<td>20ul</td>
</tr>
<tr>
<td>MmeI (2U/ul)</td>
<td>12ul</td>
</tr>
<tr>
<td>Deionized water</td>
<td>48ul</td>
</tr>
<tr>
<td><strong>Final volume</strong></td>
<td>200ul</td>
</tr>
</tbody>
</table>

Incubate at 37°C for 4 hrs to overnight. Usually, we find that an overnight digest works better. The main band of about 2,800bp is the desired linear single-PET plasmid.

2-3. Linear single-PET plasmid DNA agarose gel purification
The entire digestion reaction was electrophoresed on 0.7% agarose, loading the digestion products in as few lanes as possible, to facilitate excision. Excise the 2,800bp linear single-PET plasmid DNA band and purify using the Qiagen agarose gel extraction kit. It is advisable to run controls on the same gel, comprising uncut as well as linearized pGIS4 cloning vector, to ensure that the correct band is excised. The number of Qiaquick-spin columns to use depends on the amount of DNA that was digested: remember that each column has a maximum capacity of about 10ug DNA. Quantitate the amount of DNA recovered from the gel by spectrometry; our recovery rate is consistently about 50%.

**Note**: An approx. 2,800bp band corresponding to linear single-PET plasmid DNA should be the dominant species in the gel, when visualized after electrophoresis. Unwanted digestion products will be present as a faint smear behind the 2,800bp band.

2-4. Linear single-PET plasmid DNA self-ligation to create circular single-PET plasmids
MmeI digestion (TCC(A/G)AC 20/18) results in a random 2-base 3’-overhang at the ends of both 5’ and 3’ signature tags. These must be removed by end-polishing using T4 DNA polymerase, leaving behind 18bp blunt-ended tags:

<table>
<thead>
<tr>
<th>(0.5-2.0ug) DNA</th>
<th>50ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x TANGO buffer (Fermentas)</td>
<td>6.0ul</td>
</tr>
<tr>
<td>0.1M DTT</td>
<td>0.3ul</td>
</tr>
<tr>
<td>T4 DNA polymerase (Promega)</td>
<td>use 5U/µg DNA</td>
</tr>
<tr>
<td>10mM dNTP</td>
<td>0.6ul</td>
</tr>
<tr>
<td>Deionized water</td>
<td>to 60.0ul</td>
</tr>
</tbody>
</table>

Incubate at 37°C for 5min, then inactivate at 75°C for 10min.

**Alternative**: use the Epicentre End-It blunting kit, which does not require any optimization, and works well for up to 5ug per reaction.
The purified, blunted DNA is then adjusted to 200ul, phenol/chloroform extracted, ethanol precipitated and the DNA pellet resuspended, and (an aliquot) diluted to a concentration of approximately 2 ng/µl, assuming an arbitrary 100% recovery. Here, we are aiming for a very dilute DNA solution (final DNA concentration 1-2 ng/ul) to favour intramolecular ligation.

Self-ligation (intramolecular recircularization) is carried out as follows (set up the reaction on ice):

| Approx. 100ng DNA | 50ul |
| Ligation Solution I (Takara Ligation Kit ver.2) | 50ul |

**Alternative:** The use of 5x Invitrogen ligation buffer (with PEG) and 5 Units of T4 DNA ligase can be substituted. The final DNA concentration should be adjusted to 1-2 ng/ul.

Incubate at 16°C, overnight (12-16hrs)

2-5 Transformation by electroporation
Purify the ligation reaction to remove salts before electroporation: adjust volume to 200ul with deionized water, perform phenol/chloroform extraction (pH 7.9) and ethanol precipitate (with GlycoBlue) at -80°C for 30min. Spin, wash the pellet at least twice with 70% ethanol, and resuspend in 20ul Qiagen EB buffer.

Add 1ul of the purified ligation reaction into 25ul electrocompetent cells (we routinely use electrocompetent TOP10 (Invitrogen)) in pre-chilled 1.7ml microfuge tubes. Do not pipette up and down to mix; instead, gently stir with the pipette tip. Stand on ice for 5min, then transfer to pre-chilled Biorad electroporation cuvettes (0.1cm gap).

Add 1ml room temperature SOC media within 10sec of pulsing to maximize efficiency; transfer to 15ml Falcon tube (or equivalent), and recover at 37°C, 1hr, 200 rpm shaking.

Plate between 20-50ul (out of 1ml) on LBamp agar for quality control (QC) screening and library efficiency calculations. Incubate overnight at 37°C.

For QC counting, plate dilutions of between 1:100 to 1:1000, in a final 100ul volume made up in SOC media. Incubate overnight at 37°C. The following day, examine the colonies growing on agar: usually, even the 1:1000 plating is almost confluent (i.e. several thousand colonies).

**Note:** It is important that there are many colonies growing at this stage, as this indicates that the self-circularization step was efficient, and also that the resulting single-PET library will be of sufficient complexity.

2-6. Large scale plating of GIS single-PET library
Based on the number of colonies seen growing on the small LBamp agar plates, spread enough of the remaining transformed bacterial culture on Q-trays to obtain at least the same number (or greater) of colonies as in the original flcDNA library. This is to ensure that the library remains representative of the original sample. Incubate overnight at 37°C.
3. Creating the final GIS PET library comprising clones of concatenated PETs

In this final section, we will extract plasmid DNA from the single-PET library created previously, and perform a BamHI restriction digest to obtain the actual GIS PETs. These are gel-purified, concatenated, size-fractionated, and ligated to prepared pZErO-1 vector.

3-1. Single PET release using BamHI

PETs are excised from the single-PET plasmids using BamHI. After the overnight 37°C incubation, the resulting bacterial colonies are harvested by scraping into LB media (20-30ml per Q-tray), and pelleted by centrifugation at about 4000g for 20min in a Sorvall floor-standing ultracentrifuge, rotor SLA3000. Plasmid DNA preparation is performed using the Qiagen HiSpeed Plasmid Maxi kit (or any other kit that you may prefer).

Approximately 100ug - 500ug of plasmid DNA was digested using BamHI. Note: There is no upper limit to the amount of DNA being cut here - it is advisable to cut as much plasmid DNA as possible, to obtain a large quantity of PETs. Suggested reaction conditions are:

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>500ug plasmid DNA</td>
<td>x ul</td>
</tr>
<tr>
<td>10x unique BamHI Buffer</td>
<td>100ul</td>
</tr>
<tr>
<td>BSA</td>
<td>10ul</td>
</tr>
<tr>
<td>BamHI (20U/ul)</td>
<td>50ul</td>
</tr>
<tr>
<td>Deionized water</td>
<td>to 1ml</td>
</tr>
</tbody>
</table>

Aliquot as 100 ul fractions for enhanced digestion efficiency, and incubate at 37°C, overnight.

Notes:
(i) A two-fold excess of BamHI is used; in our hands and under these digestion conditions, we obtain better results downstream compared with a five-fold excess of enzyme.

3-2. Purification of BamHI-cohesive single PETs

After digest, perform phenol-chloroform extraction and ethanol precipitation. This step removes buffer salts which would otherwise distort the subsequent gel (especially PAGE, see below) run and purification of the 50bp PETs.

Note: For convenience, most people in the lab now run a 2% agarose gel followed by electroelution for PET purification. This is a lower-resolution method, and does not purify PETs as efficiently as PAGE, which was used when developing the original procedure (See step 3-3-1b). The effect of these impurities is manifested as damaged sequences that “poison” the concatenation step, leading to shorter than expected concatemers. Nonetheless, agarose gel electrophoresis is easier to set up and faster to run, and suitable if there are no problems with the single-PET library (which therefore enables one to obtain pure PETs). Nonetheless, it is critical to examine the quality of the agarose-gel purified PETs (e.g. by running a mini PAGE gel) before concatenation. The agarose gel purification approach is detailed in step 3-3-1a; the older PAGE purification procedure is presented below (step 3-2-1b) for reference.

3-2-1a. Agarose gel purification of GIS PETs

The digestion reaction is ethanol precipitated with GlycoBlue to reduce the volume for loading:

To every 200ul digest, add

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>3M NaOAc pH 5.2</td>
<td>20ul</td>
</tr>
<tr>
<td>1M MgCl₂</td>
<td>4.5ul</td>
</tr>
<tr>
<td>GlycoBlue</td>
<td>2.2ul</td>
</tr>
<tr>
<td>Abs. EtOH</td>
<td>800ul</td>
</tr>
</tbody>
</table>

Keep at -80°C, 30min, then spin 13K RPM, 4°C, 30min; wash 1x with 75% EtOH. Resuspend the pellets from all the tubes in a total of 100ul of LoTE buffer or Qiagen EB buffer, and loaded on a 2% agarose gel. Run at 80 volts, approx. 1.5 hrs and visualize at 365nm UV to confirm good separation of PETs from other DNA. It is convenient to run 20ul of Takara Wide Range DNA ladder for sizing.

The PET band is then excised from the agarose gel, and electroeluted using disposable Fermentas ElutaTubes (or equivalent). It is important to follow the manufacturer’s recommendations for maximum gel
size per tube. Electroelution is performed for 1 to 1.5 hrs at 90 Volts, and the harvested PETs are precipitated thus:

- To every 200ul eluate, add
- 3M NaOAc pH 5.2 20ul
- 1M MgCl₂ 4.5ul
- GlycoBlue 2.2ul
- Abs. EtOH 800ul

Resuspend all the precipitated PET DNA in 20ul LoTE buffer, and run 0.5ul and 1ul on a 4-20% PAGE minigel together with Low Mass Ladders (Invitrogen), and 25bp DNA ladder (Invitrogen) for purity check and visual quantitation.

**Optional:** PicoGreen quantitation of PETs (see Appendix (E)). Some workers have found that visual estimation of PET DNA quantity results in subjective variability, and hence uncertainty in subsequent enzymatic digestion steps. Although not critical, PicoGreen quantitation (which is not affected by the presence of glycogen/ GlycoBlue) can be useful in providing a more accurate quantitation of DNA amounts.
ALTERNATIVE
3-2-1b. PAGE purification of PETs
Resuspend the DNA pellet in LoTE buffer, at an estimated concentration of 10ug per 20ul, which is suitable for gel-loading. Cast a large (Hoefer Ruby 600 (Amersham/GE Healthcare), 15 x 15cm, 1.5mm thick, 10-well) 10% polyacrylamide gel according to this recipe:

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>40% acrylamide/bis (29:1) solution (Biorad)</td>
<td>10ml</td>
</tr>
<tr>
<td>5x TBE buffer</td>
<td>8ml</td>
</tr>
<tr>
<td>TEMED</td>
<td>16ml</td>
</tr>
<tr>
<td>Deionized water</td>
<td>21.6ml</td>
</tr>
<tr>
<td>When ready to pour, add 10% APS</td>
<td>0.4ml</td>
</tr>
</tbody>
</table>

When loading the BamHI-digestion products into the polyacrylamide gel, do not load more than approximately 10-15ug per well, as excessive DNA results in self-quenching of fluorescence. Load 1ul (=1ug) of 25bp DNA ladder into one well. Electrophorese at 200V until the Bromophenol Blue (standard tracking dye) band almost reaches the bottom of the gel. The gel is then stained in SYBR Green I (Molecular Probes, Inc.) for 30min before visualization and excision of the 50bp cohesive PETs.

**Note:**
It is preferable to use a Dark Reader (Clare Chemical) blue-light transilluminator for visualization, as exposure to UV light (especially short wavelength UV) will damage DNA. If this is not possible, then at least ensure that long wavelength (365nm) UV light is used.

**Elution by gel-crush method**
The 50bp GIS PETs are excised and collected into 0.6ml microfuge tubes (DNA from 2 lanes, per tube) which have been pierced at the bottom with a 21G needle. This pierced tube is placed inside a 1.7ml microfuge tube, and centrifuged at 13K rpm, 4°C for 5min. The gel pieces are thus conveniently shredded and collected in the bottom of each 1.7ml tube. Add 150µl of LoTE:NH₄OAc (125:25) to each tube and elute thus: overnight (12-16 hrs) at 4°C, then 37°C for 30 min to 2 hr.

Separate the supernatant (containing the eluted 50bp PETs) from the gel pieces with the aid of microspin filter units (SpinX, Costar or Mermaid spin columns, Bio101): using a 1ml pipet tip, aspirate as much liquid+gel into each column, then spin at 13K rpm, 10min, 4°C.

Pool the collected supernatant, and precipitate (a phenol-chloroform extraction step prior to precipitation is optional):

<table>
<thead>
<tr>
<th>Solution</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>3M NaOAc pH 5.2</td>
<td>20ul</td>
</tr>
<tr>
<td>1M MgCl₂</td>
<td>4.5ul</td>
</tr>
<tr>
<td>GlycoBlue</td>
<td>2.2ul</td>
</tr>
<tr>
<td>Abs. EtOH</td>
<td>800ul</td>
</tr>
</tbody>
</table>

Keep at -80°C, 30min, then spin 13K RPM, 4°C, 30min; wash 1x with 75% EtOH. Resuspend the pellets from all the tubes in a total of 20ul LoTE buffer and proceed to quantitation by eyeballing against mass ladders or by PicoGreen fluorimetry (see Appendix (E)). The presence of GlycoBlue (or glycogen) at this stage precludes spectrophotometric quantitation.
3-3. PET concatenation
Purified PETs obtained by BamHI digestion are concatenated, then partially redigested again with BamHI (step 3-4) to ensure the presence of cohesive termini suitable for insertion into BamHI/pZErO-1.

Set up the following ligation reaction:

- 50bp cohesive PETs: 200-1000ng
- 10x ligase buffer (with spermidine): 1ul
- (IMPORTANT! Spermidine enhances ligase activity, and favors linear concatenation)
- T4 DNA ligase (5U/ul): 1ul
- Deionized water: to 10ul

(or adjust to 20ul if PETs are too dilute)

Incubate at 16°C 2hrs to overnight (may require optimization, see Note below).

Note: Overnight incubation does not appear to affect the concatenation efficiency of agarose-gel purified PETs, which is somewhat disconcerting; one would properly expect over-concatenation. Apparently there is a limit to concatenation efficiency due to impurities “poisoning” the ligation reaction. In the case of PAGE-purified PETs, some optimization must be done with regard to ligation time, as over-concatenation can occur (due to high purity of recovered PETs), resulting in excessive high molecular-weight DNA that is difficult to clone.

3-4. Partial BamHI redigestion of PET concatemers
(Note: This BamHI redigestion step is performed ONLY on BamHI cohesive-ended concatemers)

This step is critical to the success of concatemer cloning. A partial digestion apparently ensures the re-creation of “correct” BamHI cohesive ends on the concatemer termini, otherwise, it appears that a significant fraction of the concatemers would fail to insert into the vector in the next step.

Purify the concatemer DNA using the Qiagen PCR purification Quickspin kit, keeping in mind the 10ug DNA capacity of each spin-column. Then quantitate the DNA (by Nanodrop if available), and do a short BamHI digest:

- Concatemers: 20ul
- 10x BamHI buffer: 3ul
- BamHI (diluted to 1U/ul): use from 1U/ug DNA to a MAXIMUM of 3U/ug DNA
- BSA: 0.5ul
- Deionized water: to 30ul

Important: Incubate at 37°C for 30 min, not any longer or you will destroy the concatemers. It is helpful initially to visualize the digestions at several time points on a PAGE minigel.

Quickly add 6ul of loading dye, heat at 65°C for 15 min and chill on ice. This step apparently minimizes non-specific aggregation/annealing of DNA which might otherwise give a false picture of concatenation efficiency.

3-5. PAGE purification of concatenated PETs
Load the entire sample preferably into 1 well (to facilitate recovery) of a 4-20% gradient polyacrylamide minigel, flanked by suitable DNA ladders to allow sizing. Suggested markers are the 25bp and 100bp ladders from Invitrogen, and the Takara Wide-Range DNA ladder.

Note: The use of multiple DNA ladders facilitates accurate sizing.

Electrophorese at 200V for about 1hr, or until the bromophenol blue tracking dye is at the bottom of the gel. Stain for 15-30min in SYBR Green I, and visualize on the Dark Reader transilluminator for gel excision.

3-6. Extraction of concatemers by gel-crush method
Excise the concatenated DNA in 3 separate fractions, low (400-1000bp); medium (1000-2000bp) and high (>2000bp).

**Important:** It is important when excising the high molecular-weight fraction, to avoid the collecting the DNA trapped in the wells: this debris appears to “poison” the subsequent ligation (or is itself favorably inserted), such that no high-molecular weight concatemers are cloned.

Place the gel slice of each excised size-fraction into a 0.6ml microfuge tube that has been pierced at the bottom with a 21G needle. This pierced tube is placed inside a 1.7ml microfuge tube, and centrifuged at 13K rpm, 4°C for 5min. The gel pieces are thus conveniently shredded and collected in the bottom of each 1.7ml tube. Add 200µl of LoTE:NH₄OAc (167:33) to each tube and elute by heating at 65°C for 2hrs.

Separate the supernatant (containing the eluted concatenated DNA) away from the gel pieces with the aid of microspin filter units as before, by spinning at 13K rpm, 10min, 4°C. Perform phenol/chloroform extraction on each eluted size-fraction, then ethanol precipitate:

- Eluted DNA fraction 200ul
- 3M NaOAc pH 5.2 20ul
- GlycoBlue 2.2ul
- Abs. EtOH 800ul

Keep at -80°C, 30min, then spin 13K RPM, 4°C, 30min; wash 1x with 75% EtOH. Resuspend the pellet in 6ul of LoTE buffer.

**Alternative:** The excised DNA can be recovered by electroelution using ElutaTubes.

3-7 Ligation to pZErO-1 vector
Prior to use, the cloning vector is prepared by digesting 2ug of pZErO-1 plasmid DNA (Invitrogen) with 10 units of BamHI for 2 hours at 37°C. The digested plasmid DNA is phenol-chloroform extracted and ethanol precipitated, then resuspended in LoTE at a concentration of 33ng/µl (Agarose gel purification is not necessary). As always, ensure that the vector preparation is validated before use (see Appendix (C)).

Set up the ligation as follows:

- Concatemer DNA fraction 6ul
- BamHI/pZErO-1 1ul
- 5x ligase buffer (with PEG) 2ul
- T4 DNA ligase (5U/ul) 1ul
- **Final volume** 10ul

Incubate at 16°C overnight. Do not heat inactivate.
Also set up vector self-ligation in parallel as a control.

3-8 Transformation by electroporation
Purify the ligation reaction to remove salts before electroporation: adjust volume to 200ul with deionized water, perform phenol/chloroform extraction (pH 7.9) and ethanol precipitate (with GlycoBlue) at -80°C for 30min. Spin, wash the pellet at least twice with 70% ethanol, and resuspend in 20ul Qiagen EB buffer.

Add 1ul of the purified ligation reaction into 25ul electrocompetent cells (we routinely use electrocompetent TOP10 (Invitrogen)) in pre-chilled 1.7ml microfuge tubes. Do not pipette up and down to mix; instead, gently stir with the pipette tip. Stand on ice for 5min, then transfer to pre-chilled Biorad electroporation cuvettes (0.1cm gap).

Add 1ml room temperature LB media (not SOC) within 10sec of pulsing to maximize efficiency; transfer to 15ml Falcon tube (or equivalent), and recover at 37°C, 1hr, 200 rpm shaking.

**Note:** We have confirmed that using LB instead of SOC, surprisingly, results in better titers.
Plate between 20-50ul (out of 1ml) on a small agar plate containing LB agar (Ensure that this is the Lennox formulation, with 5g of NaCl per liter media) plus Zeocin (25µg/ml final concentration) and incubate overnight at 37°C.

3-9 Library QC
Count the numbers of colonies and determine library efficiency after eliminating the self-ligation background. Pick colonies (24-48 is sufficient) for PCR screening using primers PMR011 and PMR012. Remember to do a control PCR on the pZErO-1 vector itself. If the PCR shows a good range of inserts, pick 1 to 4x 96-well plates of colonies for overnight culture (in Low Salt LB + Zeocin), plasmid purification and sequencing to determine the average number of PETs per insert.

At this stage, the library can be stored in the form of purified ligation mix at -20°C, until one wishes to perform large scale transformations, plasmid extractions and sequencing of GIS PETs.
Section (B): The Chromatin Immunoprecipitation-Paired End diTagging (ChIP-PET) method

Overview
ChIP-PET analysis (Figure 2) is a minor modification of the GIS-PET analysis procedure described in Section (A) above. Basically, ChIP-enriched genomic DNA (gDNA) fragments are end-polished and inserted into another unique cloning vector (pGIS3) that has the requisite restriction enzyme recognition sites already built into the vector “arms” flanking the inserted DNA. This results in the formation of a ChIP-DNA library (analogous to the flcDNA library in Section (A)). The ChIP-DNA library is then ditagged as in Part II of Section (A).
Figure 2. An overview of the ChIP-PET analysis method for mapping ChIP-enriched genomic DNA fragments.
Materials list

- Sample of interest: ChIP-enriched DNA fragments (at least 100ng)
- End-It kit for end-polishing (Epicentre)
- GlycoBlue (Ambion)
- T4 DNA ligase (5U/ul) (Invitrogen)
- 5x T4 DNA ligase buffer with PEG (Invitrogen)
- Deionized water
- S-Adenosyl Methionine (SAM) (NEB; supplied with MmeI enzyme)
- Mung Bean nuclease (Promega)
- BamHI (20U/ul) (NEB)
- 10x unique BamHI buffer (NEB)
- TEN buffer for cDNA size fractionation equilibration and elution (1x buffer is: 10mM Tris-Cl pH 8.0; 0.1mM EDTA pH 8.0; 25mM NaCl) (See REAGENTS AND SOLUTIONS)
- SYBR Green I (Molecular Probes)
- GelStar (Cambrex BioWhittaker)
- pGIS3 cloning vector (obtained from authors)
- 20uM PMR011 PCR primer
- 20uM PMR012 PCR primer
- HotStarTaq DNA polymerase system (Qiagen)
- OneShot electrocompetent TOP10 cells (Invitrogen)
- LB media, Lennox formulation with 5g/L NaCl
- LB agar, Lennox formulation with 5g/L NaCl
- Ampicillin
- Zeocin (Invitrogen)
- pZErO-1 vector (Invitrogen)

Special equipment and expendables

- Phase Lock Gel Light tubes (Eppendorf)
- Precast 4-20% PAGE gels (Invitrogen)
- Q-trays (Genetix) (or other large plates for bacterial plating)
- Heat/chill block (optional; thermal cycler can be substituted)

Notes before starting

(i) This procedure assumes familiarity with the GIS-PET procedure in the previous Section (A), especially with regard to basic molecular biology procedures.
(ii) The method of chromatin immunoprecipitation (ChIP) is not covered in this protocol.
ChIP-PET Analysis - Part I of II (Construction of ChIP-DNA library)

1. Construction of ChIP-DNA plasmid library

1-1. Pre-screening of ChIP-enriched gDNA
For best results, the de-linked, fragmented ChIP-enriched gDNA should show a smear of between 100 to 2000 bp, centering around 500 bp. These parameters can be met by adjusting sonication conditions. In addition, the minimum starting quantity of DNA should be 100ng. Due to the interfering presence of glycogen in the ChIP-enriched gDNA fragments, quantitation must be done by PicoGreen fluorimetry (See Appendix (E)).

1-2. End-polishing of DNA sample
The End-It kit (Epicentre) is convenient as it simultaneously blunts and 5’-phosphorylates the DNA termini. A typical setup is:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>100-200ng DNA</td>
<td>34ul</td>
</tr>
<tr>
<td>10x buffer</td>
<td>5ul</td>
</tr>
<tr>
<td>2.5mM dNTP mix</td>
<td>5ul</td>
</tr>
<tr>
<td>10mM ATP</td>
<td>5ul</td>
</tr>
<tr>
<td>Enzyme mix</td>
<td>1ul</td>
</tr>
<tr>
<td>Total volume</td>
<td>50ul</td>
</tr>
</tbody>
</table>

Incubate at 25°C for 45 min, then inactivate at 70°C for 10 min.
Adjust volume to 200ul with deionized water, perform phenol/chloroform extraction (pH 7.9) and ethanol precipitate (with GlycoBlue) at -80°C for at least 30min. As in Section A, the use of Phase Lock Gel tubes (Eppendorf) greatly facilitates this procedure.
Resuspend the DNA pellet in 100ul of TEN buffer (10mM Tris-HCl pH 8.0; 0.1mM EDTA pH 8.0; 25mM NaCl) in preparation for size fractionation.

1-3. Size-fractionation of DNA
While doing the ethanol precipitation, prepare Invitrogen cDNA size fractionation column (1 per sample) exactly as per the manufacturer’s protocol.

Use column equilibrated to room temperature:
Remove top cap, then bottom, and allow to drain;
Add 0.8ml TEN (10mM Tris-HCl pH 8.0; 0.1mM EDTA pH 8.0; 25mM NaCl) buffer, allow to drain completely. Repeat 3x (this process will take an hour).
Label 20 x 1.7ml microfuge tubes to be used for fraction collection.

Add the 100ul purified, blunted and phosphorylated DNA sample (adjust volume with TEN buffer) into the prepared column;
Collect entire flowthru in collection tube 1.
Add 100ul TEN buffer, collect flowthru in tube 2.
Now add 100ul TEN buffer, and start collecting single drops, 1 drop (approx. 35ul) per tube.
Allow column to drain completely before adding another 100ul TEN buffer each time.

At the end of the process, you will have 20 aliquots. Measure the absorbance of each fraction (use 10ul per fraction, in a 7ul submicro-volume cell (Amersham). Better yet, use a Nanodrop instrument, which only requires 1ul of sample, or use PicoGreen-based fluorimetry (Molecular Probes) (see Appendix (E)).

OPTIONAL (but useful): Remove 4ul from each of the 20 fractions, and run these on an agarose gel to assist in determining which fractions are suitable for cloning. Depending on yields and technique, ethidium bromide staining may be sufficient to observe the fractions, otherwise SYBR Green I (Molecular Probes) or GelStar (Cambrex Bio-Whittaker) stain can be used instead.
The objective of this size-fractionation step is to remove any DNA fragments that are too small (<100bp) as these will be preferentially cloned, yet relatively uninformative relative to the bulk of the fragmented DNA (which, as mentioned earlier, should be centered between 200-500 bp). We usually find that fractions 7-12 are suitable.

Selected fractions are pooled as appropriate, and ethanol precipitated (with GlycoBlue) to concentrate the selected DNA for ligation.

1-4. Ligation to cloning vector pGIS3 (See Appendix (C) for vector information and preparation)
Set up on ice, using 1.7ml microfuge tubes:
- 40ng/ul prepared pGIS3, 1ul
- Selected ChIP DNA fractions, 6ul
- 5x ligase buffer with PEG (Invitrogen), 2ul
- 5U/ul T4 DNA ligase, 1ul
- **Final volume 10ul**

Incubate overnight (12-16hrs) at 16°C
Also set up a vector self-ligation control.

1-5. Transformation by electroporation
Purify the ligation reaction to remove salts before electroporation: adjust volume to 200ul with deionized water, perform phenol/chloroform extraction (pH 7.9) and ethanol precipitate (with GlycoBlue) at -80°C for 30min. Spin, wash the pellet at least twice with 70% ethanol, and resuspend in 12ul Qiagen EB buffer.

Add 1ul of the purified ligation reaction into 25ul electrocompetent cells (we routinely use electrocompetent TOP10 (Invitrogen)) in pre-chilled 1.7ml microfuge tubes. Do not pipette up and down to mix; instead, gently stir with the pipette tip.
Stand on ice for 5min, then transfer to pre-chilled Biorad electroporation cuvettes (0.1cm gap).

Add 1ml room temperature SOC media within 10sec of pulsing to maximize efficiency; transfer to 15ml Falcon tube (or equivalent), and recover at 37°C, 1hr, 200 rpm shaking.

Plate between 20-50ul (out of 1ml) on LBamp agar for quality control (QC) screening and library efficiency calculations. Incubate overnight at 37°C.

1-6. Library QC
Count the numbers of colonies and determine library efficiency taking into consideration the self-ligation background. Pick colonies (24-48 is sufficient and convenient) for PCR screening using primers PMR011 and PMR012, and agarose gel electrophoresis. A range of different-sized PCR products should be observed, corresponding in size to the inserted DNA fractions.

At this stage, the library can be stored in the form of purified ligation mix at -20°C, until one wishes to proceed on to the second part of ChIP-PET analysis.

**ChIP-PET Analysis- Part II of II (ditagging)**
This procedure for this section is identical to the ditagging section in GIS-PET analysis described in Section (A).
REAGENTS AND SOLUTIONS

LoTE buffer
3mM Tris-HCl pH 7.5
0.2mM EDTA

Saturated trehalose (RNase-free)
Because the solubility of trehalose increases with temperature, it is important to maintain the temperature at not more than 42°C (the temperature at which the trehalose is used in the protocol) when making this solution.

Place magnetic stir bar into small beaker and heat 10ml of water to 42°C.
Slowly add in 8g of trehalose powder and let it dissolve by stirring.
While carefully maintaining temperature at 42°C, continue adding trehalose until saturation is reached.
Upon cooling down this solution to room temperature, trehalose crystals will form.
At this stage, add in enough DEPC to make a 0.1% DEPC final concentration, shake vigorously for 5-10 min (or stir overnight), then autoclave the saturated trehalose solution to deactivate the DEPC.

Modified 10mM dNTP with 5-Me-dCTP
For 1ml of dNTP mix,
100mM stock solution of dATP 100ul (final conc. = 10mM)
100mM stock solution of dTTP 100ul (final conc. = 10mM)
100mM stock solution of dGTP 100ul (final conc. = 10mM)
100mM stock solution of 5-Me-dCTP 50ul (final conc. = 5mM)
1M Tris-HCl pH8.0 10ul (final conc. = 10mM)
Deionized water 640ul

1x Binding Buffer (steps 1-4 and 1-5)
2M NaCl
50mM EDTA pH 8.0

1x Blocking Buffer (step 1-5)
0.4% SDS
50ug/ml yeast tRNA

1x Wash buffer (step 1-5)
10mM Tris-HCl pH 7.5
0.2mM EDTA,
10mM NaCl
20% glycerol
40ug/ml yeast tRNA

1x TEN buffer for size-fractionation column equilibration and DNA elution
10mM Tris-HCl pH 8.0
0.1mM EDTA pH 8.0
25mM NaCl

2x TEN buffer for adapter annealing
20mM Tris-HCl pH8.0
0.2mM EDTA pH8.0
100mM NaCl
10x ligation buffer with Spermidine
60mM Tris-HCl pH 7.5
60mM MgCl2
50mM NaCl
1mg/ml BSA
70mM Beta-mercaptoethanol
1mM ATP
20mM DTT
10mM spermidine
Critical parameters and Troubleshooting

(A) GIS-PET analysis method

flcDNA library procedure (Part I)

Cloning vector quality

For library construction purposes, the quality of the cloning vector is critical, in that it should produce only a low background of no-insert clones, and a library with a high titer. For this reason, always ensure that the vector preparation is pre-validated before use, following the procedure described in Appendix (B). It is very important to freeze the validated vector preparation in small aliquots, as we have seen that repeated freeze-thawing of the vector DNA results in degradation and consequently undesirable results including decreasing library titers, and the prevalence of clones in which the polyA tails were not removed.

Generation of flcDNA- general points

- Do not use glycogen at any stage where it is not specifically mentioned, as this will interfere with the CapTrapper selection process.
- For all steps involving single-stranded DNA or RNA, it is preferable to use a siliconized or “lo-binding” 1.5ml or 1.7ml microfuge tube to avoid loss of nucleic acid.
- Phase Lock Gel tubes (Eppendorf) are very convenient for the numerous phenol/ chloroform extraction steps, and help to reduce losses of aqueous sample. It is important that Phase Lock Gel should not be used after reverse-transcription (step 1-1), as the presence of trehalose and sorbitol increases the solution density to the point that both phases form below the gel layer, making recovery difficult.

Vector-insert ligations

The quantity of size-fractionated cDNA used for ligation to the pGIS vector is important; we have found that at least 100ng per 30-40ng vector is best. Additionally, one should avoid the temptation of using the smallest fractions- these contain excess adapter DNA, which can block the vector cloning sites. Finally, ensure that as far as possible, separate ligations are set up for different size fractions of cDNA, and avoid pooling all size fractions together for the ligation. This ensures a more complete collection of cDNA of different sizes.

Quality of flcDNA library

Before proceeding to the ditagging section (Part II), it is important to assay the quality of the flcDNA library. The quality of the library can be inferred from a combination of different Quality Control checks, including counting total titer, determining the range of insert sizes by colony PCR, and most importantly, by sequencing a selection of clones. Check the DNA sequences for the percentage of full-length transcripts, as well as for the presence of all expected flanking restriction enzyme recognition sites. To obtain an estimated 10-fold coverage of a human transcriptome, we aim for a flcDNA library with a total titer of 1x10⁶ cfu.

Large-scale library plating

Use of Q-trays: it is important to dry the plates(containing LB agar) sufficiently before use, otherwise the condensate produced during subsequent incubation will result in smearing/ merging of growing colonies. We usually leave them open in a sterile hood until a pattern of “waves” appear on the agar surface. During incubation, do not stack Q-trays more than 3-high, to avoid excessive condensation.

Plasmid DNA maxiprep from bacteria scraped from Q-trays

Extraction of plasmid DNA from these “bacterial scrapings” is less efficient than starting with the usual liquid cultures. You should expect a yield of approximately 1mg plasmid DNA starting from 1x10⁶ scraped colonies (about 10 Q-trays); if your yields are lower than 600ug of plasmid DNA, check that you are using enough solution volume at each stage: e.g. if using the Qiagen maxiprep system, ensure that 20-50ml of each buffer P1, P2 and P3 is used. There is a very large quantity of precipitated flocculant debris after the P3 neutralization step- this can result in poor recovery of the supernatant if one uses QiaFilters directly, so it would be better to preclear the precipitate by centrifugation (Oak Ridge 50ml tubes in SS34 rotor, 12,000 rpm for 30min), followed by the use of QiaFilters if necessary for a final removal of particulates before introducing the supernatant into the maxiprep tips.
**Ditagging Procedure (Part II)**

**PET concatenation**
This duration of the ligation step may need to be optimized, depending on the quality of the purified PETs you obtain. Aim for a majority of concatenated DNA around 1000–2000 bp. Ensure that you have added spermidine in your ligation buffer- we have found that this greatly increases the efficiency of the desired linear concatenation.

**Cloning of BamHI-derived PET concatemers**
The partial redigestion of the concatenated DNA by BamHI prior to gel fractionation appears to be critical to the success of the subsequent cloning step. Follow the recommended conditions carefully, to avoid either under- or over-digestion. We have successfully used BamHI at a specific activity of 1 – 3 U/ug substrate DNA.

**Recovery of gel-purified DNA**
We have not found much difference in yields, using either the “gel-crush” method of recovering DNA (step 3-6), or electroelution using ElutaTubes. Electroelution however can result in significant time savings.

**(B) ChIP-PET analysis method**

**Quality of starting material**
It is critical that the starting material, viz. ChIP-enriched gDNA fragments are of a sufficient quality to generate a high-quality ChIP-DNA library. Before starting, ensure that the fragments were completely de-crosslinked, are of sufficient quantity (preferably 200ng or more), and are of an appropriate size-range (centered around 50bp). Poor quality starting DNA manifests itself in a ChIP-DNA library with low titer and/or excessively short inserts.

**Other**
The general precautions listed for the GIS-PET analysis method are also applicable to ChIP-PET.

**Anticipated results**

**(A) GIS-PET analysis method**
Starting with approximately 10ug of high-quality polyA RNA, one should be able to obtain a flcDNA library with a total titer of 5x10^5 to >2x10^6 cfu. A pool of about 1x10^6 colonies scraped from 10 or more Q-trays should generate about 1mg of plasmid DNA. In the ditagging part of the procedure, 10ug of plasmid DNA from the flcDNA library should produce about 1ug of purified, linearized, single-PET plasmid DNA that will be self-circularized to generate the single-PET plasmid library. This single-PET plasmid library has a typical total titer of >5x10^6 cfu. The final GIS-PET library in pZErO-1, constructed by cloning various size fractions of approximately 1ug of concatenated 50bp PETs, should have a total titer of about 1x10^6 cfu, with 7-12 PETs present per clone.

**(B) ChIP-PET analysis method**
Starting with 200ng of ChIP-enriched DNA, one should be able to generate a ChIP-DNA library with a total titer of about 4x10^6 to >1x10^7 cfu. The other downstream results would be similar to what is expected for GIS-PET.

**Time Considerations**

**(A) GIS-PET analysis method**
It will take 6 days to construct the flcDNA library starting from purified polyA RNA, and another 9 days to prepare the GIS PET library for sequencing. Factor in another 4 days for preparation and validation of cloning vectors.

**(B) ChIP-PET analysis method**
Starting from ChIP-enriched DNA fragments, it should take 3 days to construct the ChIP-DNA library, and another 9 days to prepare the GIS PET library for sequencing.
APPENDICES

(A) Oligonucleotide sequences
(B) Adapter annealing
(C) Cloning vector information, preparation and validation (pGIS4 and pGIS3)
(D) List of reagents, consumables and suggested suppliers
(E) PicoGreen quantitation of DNA
Appendix (A): Oligonucleotide sequences

Although PAGE-purified oligonucleotides are preferred, for reasons of economy HPLC purified ones are usually used, and in our hands they work well. Oligonucleotide sequence details are shown in Figure 3. The “p-” refers to a phosphate modification.

<table>
<thead>
<tr>
<th>Oligonucleotide Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GsuI-oligo dT primer:</td>
<td>5’-GAGCTAGTTCTGGAGTTTTTTTTTTTTTTTVN-3’</td>
</tr>
<tr>
<td>GIS(-N)6 adapter:</td>
<td>5’-CTAAACTCGAGGGCGGCCGATCCGACNNNNNN-3’</td>
</tr>
<tr>
<td></td>
<td>3’-TTTGAGCTCCGCGGGCCGCTAGGCTG-p-5’</td>
</tr>
<tr>
<td>GIS(-N)5 adapter:</td>
<td>5’-CTAAACTCGAGGGCGGCCGATCCGACGNNNNN-3’</td>
</tr>
<tr>
<td></td>
<td>3’-TTTGAGCTCCGCGGGCCGCTAGGCTG-p-5’</td>
</tr>
<tr>
<td>GIS-Sal adapter:</td>
<td>5’-p-GTCGGATCCTGGG-3’</td>
</tr>
<tr>
<td></td>
<td>3’-TTCTAGGACCAGCTAG-5’</td>
</tr>
<tr>
<td>PMR011 (M13(+))</td>
<td>5’-GATGTGCTGCAAGCGATAGAGG-3’</td>
</tr>
<tr>
<td>PMR012 (M13(-))</td>
<td>5’-AGCGGATAACAATTTCACACAGG-3’</td>
</tr>
<tr>
<td>Structure of a generic ~50bp BamHI-cohesive PET</td>
<td>5’-GATCCGACXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNAAGTCG-3’</td>
</tr>
<tr>
<td></td>
<td>3’-GCTGXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNTTCAGCCTAG-5’</td>
</tr>
<tr>
<td></td>
<td>Wherein X and N may be any of A, C, G or T.</td>
</tr>
<tr>
<td>pGIS7/4VERIF-TOP</td>
<td>5’-GGCCGCGTGAGCAAGAAGAAGCAGCAGAAAGACCCGTAGGATCAACCTCAGGAA-3’</td>
</tr>
<tr>
<td>pGIS7/4VERIF-BOT</td>
<td>5’-CTCGAGTTGGATCATCTTAGCTGCTTCTTCTGCTTGCTCAGC-3’</td>
</tr>
<tr>
<td>pGIS7/4VERIF-TOP and pGIS7/4VERIF-BOT</td>
<td>anneal to form a useful adapter for verifying the integrity of (NotI+BseRI)-cut pGIS4.</td>
</tr>
<tr>
<td></td>
<td>5’-GGCCGCGTGAGCAAGAAGAAGCAGCAGAAAGACCCGTAGGATCAACCTCAGGAA-3’</td>
</tr>
<tr>
<td></td>
<td>CGCACCCTTTCTTCGCTGCTCTTCTGTGGAGATCTAGGTCGTC</td>
</tr>
</tbody>
</table>

Figure 3. Custom oligonucleotide sequences
Appendix (B): Adapter annealing

Our oligos are ordered already pre-adjusted to 100uM concentration (Proligo). These are in water, so prior to annealing they have to be adjusted to the correct concentration of Tris-NaCl-EDTA (TNE).

2x TNE is 20mM Tris-HCl pH8.0, 100mM NaCl, 0.2mM EDTA (these are final concentrations)

In a 0.2ml PCR tube, mix together:
- oligonucleotide A 20ul
- oligonucleotide B 20ul
- 2x TNE 40ul

Heat 95°C x 2';
- ramp from 95°C to 75°C at 0.1°C/second, hold at 75°C x 2';
- ramp from 75°C to 65°C at 0.1°C/second, hold at 65°C x 2';
- ramp from 65°C to 50°C at 0.1°C/second, hold at 50°C x 2';
- ramp from 50°C to 37°C at 0.1°C/second, hold at 37°C x 2';
- ramp from 37°C to 20°C at 0.1°C/second, hold at 20°C x 2';
- ramp from 20°C to 4°C at 0.1°C/second, hold at 4°C indefinitely until ready to collect.

Important! Keep all annealed adapters cold.

Quality check:
(1) Run ssDNA oligos vs annealed adapters on the same PAGE minigel (either 20%, or gradient 4-20%);
(2) In the case of the GIS-(N)5 and GIS-(N6) adapters, a simple NotI digestion should fragment the adapter in 2, showing that the NotI site is present and correct.
Appendix (C): GIS cloning vector information, preparation and validation (pGIS4, pGIS3)

Notes:
(i) The pGIS4 vector (Figure 4a; cloning site sequence details in Figure 4b) was created as an improvement to the published prototype pGIS1, and renders 3’ adapter ligation unnecessary. It contains an approximately 800bp stuffer which has to be excised before use by NotI + BseRI digestion, which simultaneously creates a NotI cohesive site to receive the 5’ adapter terminal region of the cDNA, and a TT overhang ready to receive the corresponding AA overhang at the 3’ end of the prepared cDNA insert.
(ii) The pGIS3 vector (Figure 5a; cloning site sequence details in Figure 5b) was created to facilitate direct ditagging of inserted blunt-ended DNA fragments. The restriction sites important for ditagging were built into the vector backbone itself. There is no stuffer fragment in pGIS3.

Figure 4a. Schematic diagram of the pGIS4 cloning vector for GIS-PET analysis
Figure 4b. Sequence details of the cloning site in pGIS4

| 1 | GGGCGAATTTC GATATCCGCG CCGCGCTTGG ATAAAGTCAG CAGCTTCCAC GCCAGCTTCA CACAAAAAGT GACTGACGGT AGCGGCGCGG CGGTGCAGGA |
| 101 | AGGTCAAGGGC GATCTGTGGG TGAAGAGTGTC AAACCTATTCC AACTGACATA TGAGCAAAAC TGATGAAAAC ATCTCTGTTTT CTGAGATTTG AAGACTGTTG |
| 201 | TTCTATAACC CGTTCGGTTCG GCAAGCTACG CAGCGATGCA CCACCAACCT AACCTTTTAC ATCTACTTTC AACGCTGTCG TGGGCTTTA TGGTAACTGTC ACTCTTCGCT |
| 301 | GGCACAGTGT AATATCGACG CGAAGCGCTG CCTGACCGAG GACACCGAGG AAGCTGACGT TCGAGATTTG TGGGCTTTA TGGTAACTGTC ACTCTTCGCT |
| 401 | TGGCACAATTC CATCGATTTA GCAACCTGGAG GCTGACCTGTG TCCACCTGAG TGGAGCTGAG GCTGACCTGTG TCCACCTGAG TGGAGCTGAG GCTGACCTGTG |
| 501 | CCTCTACCAT CGCGCAGAGG GCTCAGGCTG TATGTGAAAAT ATACGCGCCG GCAAGTCCGG TGGAGTACGT TGGGCTTTA TGGTAACTGTC ACTCTTCGCT |
| 601 | CCGCGGATGG CCAACGGGTTA AATATCGGCG GCAAGTCCGG TGGAGTACGT TGGGCTTTA TGGTAACTGTC ACTCTTCGCT |
| 701 | GGATATCGGG TACCTGTGCTG TGGGCTTGG TGGGCTTTA TGGTAACTGTC ACTCTTCGCT |
Figure 5a. Schematic diagram of the pGIS3 cloning vector for ChIP-PET analysis
Figure 5b. Sequence details of the cloning site in pGIS3

<table>
<thead>
<tr>
<th>NotI</th>
<th>MmeI</th>
<th>MmeI</th>
<th>SalI</th>
</tr>
</thead>
<tbody>
<tr>
<td>~~~~</td>
<td>~~~~</td>
<td>~~~~</td>
<td>~~~~</td>
</tr>
<tr>
<td>BseRI</td>
<td>BamHI</td>
<td>XhoI</td>
<td>BamHI</td>
</tr>
<tr>
<td>~~~~</td>
<td>~~~~</td>
<td>~~~~</td>
<td>~~~~</td>
</tr>
</tbody>
</table>

GGGCGAATTC GATATCGCGG CCGCGAGGAG TATGGATCCG ACTCGAGTCG GATCCTGGCT CCTCGTCGAC CTGCAGGCAT GCAAGCTTGA GTATTCTATACCCGCTTAAG CTATAGCGCC GGCGCTCCTC ATACCTAGGC TGAGCTCAGC CATAAGATAT
Appendix (C) (continued)

**C1. pGIS4**

**C1-1. Plasmid purification**
As per the Qiagen maxi protocol. However, to ensure that no contaminating bacterial genomic DNA is subsequently inserted into the prepared vector, we do an extra gel-purification of the maxiprep on agarose to isolate supercoiled DNA before digesting this with NotI/BseRI. Recovery (with the Qiagen gel extraction kit) is approximately 50%.

**Important**: the maximum capacity of each Qiaquick-spin tube is only 10ug, so be sure to determine the number of tubes needed beforehand.

**C1-2. Enzymatic digestion**
It is important to follow the conditions given here closely, as BseRI exhibits some non-specific activity.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>5ug purified supercoiled pGIS4</td>
<td>x ul</td>
</tr>
<tr>
<td>10x NEBuffer 3</td>
<td>20 ul</td>
</tr>
<tr>
<td>NotI (10U/ul)</td>
<td>5.0 ul (5-fold excess)</td>
</tr>
<tr>
<td>BseRI (4U/ul)</td>
<td>12.5 ul (5-fold excess; note BseRI is 75% active in Buffer 3)</td>
</tr>
<tr>
<td>BSA</td>
<td>2 ul</td>
</tr>
<tr>
<td>Water</td>
<td>to 200 ul</td>
</tr>
</tbody>
</table>

Incubate as 2x 100ul aliquots at 37°C, 3 hrs maximum; agarose gel purify.
Quantitate using Nanodrop, and resuspend in EB buffer at 40ng/ul.

**C1-3. Vector validation**
This section is generally applicable to any cloning vector. The purpose of this is to ensure that there is minimal contamination of the NotI+BseRI cut vector with either uncut or single-cut plasmid, which would otherwise contribute to unwanted background.

After quantitating the NotI+BseRI cut plasmid band, set up the following ligations (volumes for illustrative purposes only):

<table>
<thead>
<tr>
<th>(i) No ligase (uncut plasmid test)</th>
<th>(ii) Self-ligate (single-cut plasmid test)</th>
<th>(iii) Plus insert (positive control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40ng (NotI+BseRI) pGIS4</td>
<td>1 ul</td>
<td>1</td>
</tr>
<tr>
<td>80ng compatible stuffer*</td>
<td>- ul</td>
<td>4</td>
</tr>
<tr>
<td>5x ligation buffer (with PEG)</td>
<td>2 ul</td>
<td>2</td>
</tr>
<tr>
<td>T4 DNA ligase (5U/ul)</td>
<td>- ul</td>
<td>1</td>
</tr>
<tr>
<td>Deionized water</td>
<td>7 ul</td>
<td>2</td>
</tr>
</tbody>
</table>

Incubate 16°C overnight, transform into bacteria and count colonies: there should be none/very few for (i) and (ii), and many for (iii).

* Compatible stuffer: in this case, it would be a dsDNA fragment with a cohesive 5’ NotI site, and a 3’ AA overhang. It is possible to simply gel extract the smaller digestion product (after the NotI+BseRI vector digestion) and use this as an insert, however any non-specific cutting would be preserved in this product, giving rise to false positives in (iii). It is far better to construct your own pre-determined dsDNA adapter and use this as a control insert. We use the adapter obtained by annealing the oligonucleotides pGIS7/4VERIF-TOP and pGIS7/4VERIF-BOT (see also Appendices (A) and (B)).
C2. pGIS3
C2-1. Plasmid purification
As per the Qiagen maxi protocol. However, to ensure that no contaminating bacterial genomic DNA is subsequently inserted into the prepared vector, we do an extra gel-purification of the maxiprep on agarose to isolate supercoiled DNA before digesting this with XhoI. Recovery (with the Qiagen gel extraction kit) is approximately 50%.

Important: the maximum capacity of each Qiaquick-spin tube is only 10ug, so be sure to determine the number of tubes needed beforehand.

C2-2. Enzymatic digestion

Purified supercoiled pGIS3  5ug
10x NEBuffer 2   10ul
BSA     1ul
Xho I (20u/ul)   2ul
Water    to 100ul
Incubate at 37°C for 2 hrs; agarose gel purify.

Following the XhoI digest to linearize the vector (and expose the cloning “arms”), polish off the exposed 3’ overhangs using Mung Bean nuclease.

Purified, XhoI-digested pGIS3  approx. 2ug
10x reaction buffer   3ul
Mung Bean nuclease (Promega) use 5U/ug DNA
Deionized water    to 30ul
Incubate at 30 °C for 30min, then stop the reaction by adding 10% SDS to a final concentration of 0.01%, (hence, add 3.3ul of 10% SDS to the 30ul reaction).

Notes:
(i) T4 DNA polymerase or Klenow enzyme cannot be used for this end-polishing step, as they will not remove the 5’ overhangs to leave the Mmel sites directly adjacent to the inserted DNA fragment.  
(ii) Mung Bean nuclease is supplied at high concentration (70U/ul), and must be diluted before use. We typically dilute it to 5U/ul using 1x reaction buffer (supplied). It is critical to do the dilution just prior to use, as dilute MBN is unstable.  
(iii) This Mung Bean nuclease polishing step can be difficult, hence the importance of the vector validation step C2-3.

After the end-polishing reaction, use Qiagen PCR purification columns to remove buffer salts and enzymes, and do a Nanodrop quantitation

Next, the 5’-phosphate groups are removed to prevent self-circularization:
5 ug XhoI-cut, end-polished pGIS3   x ul
10x Antarctic phosphatase buffer   5 ul
Antarctic phosphatase (5U/ul) (NEB)   5 ul (use 5U/ug target)
Deionized water    to 50 ul
(note: volumes are only indicative, adjust as needed)

Incubate 37°C, 30 minutes (exactly!); then heat inactivate at 65°C for 20 minutes. There is no further need to do a phenol extraction, as Antarctic Phosphatase enzyme is completely inactivated by heating.

If necessary, dilute the prepared plasmid DNA in Qiagen EB buffer (or TE buffer) to a final concentration of 40 ng/ul (or less). If desired, one can run 2 ul (estimated to be 80ng) of the diluted DNA on an agarose gel against known standards, to confirm the quantity of DNA.
C2-3. Vector validation (See Appendix (C), step C1-3)
The prepared vector MUST now be validated and sequenced before use. This is to ensure that the MmeI sites were not damaged during the end-polishing step, as this would render the vector (and the entire library) useless. As a standard procedure, set up these ligation reactions (in 5x buffer with PEG):

(i) No ligase control reaction (uncut DNA check)
(ii) With ligase, no stuffer fragment (self-ligation check)
(iii) Add blunt, phosphorylated stuffer fragment (at least 80ng stuffer; positive control)
Appendix (D): List of reagents and suggested suppliers

<table>
<thead>
<tr>
<th>Item</th>
<th>Supplier</th>
<th>Catalog no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 M EDTA, pH 8.0, 100 ml</td>
<td>Ambion</td>
<td>9260G</td>
</tr>
<tr>
<td>1 M Tris, pH 8.0, 100 ml</td>
<td>Ambion</td>
<td>9855G</td>
</tr>
<tr>
<td>10% SDS, 500ml</td>
<td>Ambion</td>
<td>9822</td>
</tr>
<tr>
<td>10X TBE, 4 x 1 L</td>
<td>Ambion</td>
<td>9865</td>
</tr>
<tr>
<td>1M MgCl2, 100ml</td>
<td>Ambion</td>
<td>9530G</td>
</tr>
<tr>
<td>5 M NaCl, 100 ml</td>
<td>Ambion</td>
<td>9760G</td>
</tr>
<tr>
<td>5 M NH4OAc, 100ml</td>
<td>Ambion</td>
<td>9070G</td>
</tr>
<tr>
<td>Gel Loading Buffer II, 1.4ml</td>
<td>Ambion</td>
<td>8546G</td>
</tr>
<tr>
<td>GlycoBlue (15 mg/ml), 0.3 ml x 5</td>
<td>Ambion</td>
<td>9516</td>
</tr>
<tr>
<td>RNA Millennium Markers, 0.5-9kb</td>
<td>Ambion</td>
<td>7150</td>
</tr>
<tr>
<td>Millennium Markers™-Formamide, 0.5-9kb</td>
<td>Ambion</td>
<td>7151</td>
</tr>
<tr>
<td>Nuclease-free water, 10x50 ml</td>
<td>Ambion</td>
<td>9937</td>
</tr>
<tr>
<td>Phenol:Chloroform:IAA, 25:24:1, pH 6.6/7.9, 100 ml</td>
<td>Ambion</td>
<td>9730</td>
</tr>
<tr>
<td>Proteinase K Solution (20 mg/ml), 1.25 ml</td>
<td>Ambion</td>
<td>2546</td>
</tr>
<tr>
<td>RNaseZap, 250ml</td>
<td>Ambion</td>
<td>9780</td>
</tr>
<tr>
<td>Sodium Acetate, 3 M, pH 5.5, 100 ml</td>
<td>Ambion</td>
<td>9740</td>
</tr>
<tr>
<td>Yeast tRNA (10 mg/ml), 10 x 1 ml</td>
<td>Ambion</td>
<td>7118</td>
</tr>
<tr>
<td>Sodium Acetate, 3.0 M, pH 5.2</td>
<td>Amresco</td>
<td>E498-100ML</td>
</tr>
<tr>
<td>50x TAE buffer, 1L</td>
<td>BioRad</td>
<td>161-0743</td>
</tr>
<tr>
<td>Electro cuvettes, 50/pack</td>
<td>BioRad</td>
<td>165-2089</td>
</tr>
<tr>
<td>Sterilin Petri Dish 90x16 3 Vents S 500 pcs</td>
<td>BSN</td>
<td>101 VR20</td>
</tr>
<tr>
<td>Ampicillin, sodium salt, 25g</td>
<td>Calbiochem</td>
<td>171254</td>
</tr>
<tr>
<td>GelStar Nucleic Acid Gel Stain</td>
<td>Cambrex</td>
<td>50535</td>
</tr>
<tr>
<td>Lazy-L spreaders</td>
<td>Sigma</td>
<td>Z376779-1PAK</td>
</tr>
<tr>
<td>Spin-X® Plastic Centrifuge Tube Filters.sterile, 0.45um</td>
<td>Corning</td>
<td>8162</td>
</tr>
<tr>
<td>Dynabeads® M-280 Streptavidin, 10ml</td>
<td>Dynal</td>
<td>112.06</td>
</tr>
<tr>
<td>End-It™ DNA End-Repair Kit, 20 rxns</td>
<td>Epicentre</td>
<td>ER0720</td>
</tr>
<tr>
<td>DNA LoBind Tubes, 1.5 ml PCR clean, safe-lock, 250 pieces</td>
<td>Eppendorf</td>
<td>0030 108.051</td>
</tr>
<tr>
<td>Phase Lock Gel Light 2.0 (200/pac)</td>
<td>Eppendorf</td>
<td>E0032005101</td>
</tr>
<tr>
<td>6x loading dye solution, 1ml x 5</td>
<td>Fermentas</td>
<td>R0611</td>
</tr>
<tr>
<td>Gsu l, 5/ul, 100u</td>
<td>Fermentas</td>
<td>R0461</td>
</tr>
<tr>
<td>IPTG, dioxane-free 1g, 5g, 25g</td>
<td>Fermentas</td>
<td>R0391,2,3</td>
</tr>
<tr>
<td>T4 DNA Ligase (5 U/μl), 200 u (with PEG)</td>
<td>Fermentas</td>
<td>EL0334</td>
</tr>
<tr>
<td>GeBAflex Tubes Midi 3500 MWCO, 50-800ul, 30 tubes/kit</td>
<td>GeBA</td>
<td>D012</td>
</tr>
<tr>
<td>1 Kb Plus DNA Ladder, 250ug</td>
<td>Invitrogen</td>
<td>10787018</td>
</tr>
<tr>
<td>100 bp DNA Ladder, 50μg</td>
<td>Invitrogen</td>
<td>15628019</td>
</tr>
<tr>
<td>10mM dNTP mixed, 100ul</td>
<td>Invitrogen</td>
<td>18427-013</td>
</tr>
<tr>
<td>25 bp DNA Ladder</td>
<td>Invitrogen</td>
<td>10597011</td>
</tr>
<tr>
<td>cDNA Fractionation Column, 3/pack</td>
<td>Invitrogen</td>
<td>18092-015</td>
</tr>
<tr>
<td>lmMedia Zeo Agar</td>
<td>Invitrogen</td>
<td>Q621-20</td>
</tr>
<tr>
<td>lmMedia Zeo Liquid</td>
<td>Invitrogen</td>
<td>Q620-20</td>
</tr>
<tr>
<td>Low DNA Mass Ladder, 400ul</td>
<td>Invitrogen</td>
<td>10068013</td>
</tr>
<tr>
<td>Novex precast TBE Gels, 4-20%, 1.0 mm, 10 wells, ?/Box</td>
<td>Invitrogen</td>
<td>EC6225BOX</td>
</tr>
<tr>
<td>SUPERSCRIPT II, RNase H- RT, (200u/μl), 2,000u</td>
<td>Invitrogen</td>
<td>18064-014</td>
</tr>
<tr>
<td>Description</td>
<td>Supplier</td>
<td>Code</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>--------</td>
</tr>
<tr>
<td>SuperScript III, Reverse Transcriptase, (200u/ul), 2,000u</td>
<td>Invitrogen</td>
<td>18080-044</td>
</tr>
<tr>
<td>TOP10 electrocompetent cells, 20xns</td>
<td>Invitrogen</td>
<td>C4040-52</td>
</tr>
<tr>
<td>Trizol reagent, 200ml</td>
<td>Invitrogen</td>
<td>15596-018</td>
</tr>
<tr>
<td>cDNA Integrity Kit, 15 rxns</td>
<td>KPL</td>
<td>60-06-00</td>
</tr>
<tr>
<td>MACS mRNA Isolation Kit</td>
<td>Miltenyi Biotec</td>
<td>130-075-102</td>
</tr>
<tr>
<td>SYBR® Green I, 10,000x in DMSO, 50ul x 20</td>
<td>Molecular Probes</td>
<td>S7585</td>
</tr>
<tr>
<td>PPCO Oak Ridge Tube, 50ml, 50pcs/cs</td>
<td>Nalgene</td>
<td>3139-0050</td>
</tr>
<tr>
<td>BseRI (4,000 units/ml), 1,000 units</td>
<td>NEB</td>
<td>R0581L</td>
</tr>
<tr>
<td>Mmel, 2,000 units/ml, 500u</td>
<td>NEB</td>
<td>R0637L</td>
</tr>
<tr>
<td>Mung Bean Nuclease, 1500 u/150ul</td>
<td>NEB</td>
<td>M0250S</td>
</tr>
<tr>
<td>Not I (10,000 u/ml), 2,500 u</td>
<td>NEB</td>
<td>R0189L</td>
</tr>
<tr>
<td>SAM, 32 mM, 0.5 ml</td>
<td>NEB</td>
<td>B9003S</td>
</tr>
<tr>
<td>RNase ONE™ Ribonuclease (5–10u/ul), 5000u</td>
<td>Promega</td>
<td>M4265</td>
</tr>
<tr>
<td>RNasin® Plus RNase Inhibitor(40u/µl), 2,500u</td>
<td>Promega</td>
<td>N2611</td>
</tr>
<tr>
<td>T4 DNA polymerase, 5-10u/ul, 100u</td>
<td>Promega</td>
<td>M4211</td>
</tr>
<tr>
<td>Roll &amp; Grow® Plating Beads, 2 Pack</td>
<td>Q-biogene</td>
<td>5000-550</td>
</tr>
<tr>
<td>Buffer EB (250ml)</td>
<td>Qiagen</td>
<td>19086</td>
</tr>
<tr>
<td>HiSpeed Plasmid Maxi Kit (25)</td>
<td>Qiagen</td>
<td>12663</td>
</tr>
<tr>
<td>HotStarTaq DNA Polymerase, 4 x 250 units</td>
<td>Qiagen</td>
<td>203205</td>
</tr>
<tr>
<td>Qiaquick Gel Extraction Kit (250)</td>
<td>Qiagen</td>
<td>28706</td>
</tr>
<tr>
<td>Qiaquick PCR purification Kit (250)</td>
<td>Qiagen</td>
<td>28106</td>
</tr>
<tr>
<td>5-Me-dCTP, 5'-Methyl-2'-deoxycytidine-5'-triphosphate</td>
<td>Roche</td>
<td>10757047001</td>
</tr>
<tr>
<td>dATP, (dCTP), dGTP, dTTP, 25 uM each</td>
<td>Roche</td>
<td>1969064</td>
</tr>
<tr>
<td>(D+)Trehalose, dihydrate, 10g</td>
<td>Sigma</td>
<td>T5251</td>
</tr>
<tr>
<td>Diethyl Pyrocarbonate (DEPC), 100ml</td>
<td>Sigma</td>
<td>D5758</td>
</tr>
<tr>
<td>D-sorbitol</td>
<td>Sigma</td>
<td>S6021</td>
</tr>
<tr>
<td>Gel handler gel support, 10sh/pk</td>
<td>Sigma</td>
<td>Z37,695-7</td>
</tr>
<tr>
<td>Glycerol, 100ml</td>
<td>Sigma</td>
<td>G5516</td>
</tr>
<tr>
<td>Isopropanol, 500ml</td>
<td>Sigma</td>
<td>I-9516</td>
</tr>
<tr>
<td>NaIO₄, 25g</td>
<td>Sigma</td>
<td>S1878-25g</td>
</tr>
<tr>
<td>Nalgene centrifuge bottles, 500ml</td>
<td>Sigma</td>
<td>Z353744</td>
</tr>
<tr>
<td>spermidine</td>
<td>Sigma</td>
<td>S0266</td>
</tr>
<tr>
<td>DNA Ligation Kit Ver.2.1 (with Solutions I, II)</td>
<td>Takara</td>
<td>6022</td>
</tr>
<tr>
<td>TaKaRa ExTaq (Mg2+ Buffer), (5 units/µl), 250u</td>
<td>Takara</td>
<td>RR001A</td>
</tr>
<tr>
<td>TaKaRa LA Taq with GC Buffer (5 units/µl), 125 units</td>
<td>Takara</td>
<td>RR02AG</td>
</tr>
<tr>
<td>Wide-Range DNA Ladder (50-10,000 bp), 0.8ml (100 lanes)</td>
<td>Takara</td>
<td>3415A</td>
</tr>
<tr>
<td>Biotin (Long Arm) Hydrazide, 50 mg</td>
<td>Vector Labs</td>
<td>SP-1100</td>
</tr>
</tbody>
</table>

41
Appendix (E): PicoGreen quantitation of DNA

Adapted from http://www.molecularprobes.com/media/pis/mp07581.pdf.

Procedure

1. **Reagent Preparation**
Dilute concentrated PicoGreen reagent 200-fold in TE buffer just before the experiment. 50ul of 1× PicoGreen is used for each sample analysis. E.g., to make enough working solution to assay 20 samples (including blank controls), add 5ul concentrated PicoGreen to 995ul TE. Keep it from light.

2. **DNA Standard Curve Preparation**
Calf thymus DNA (Sigma-Aldrich, Cat. No.: D4522) is used to make the standard curve. 100ng/ul calf thymus DNA is prepared as a stock solution. First of all, decide the standard curve range according to the estimated concentration of DNA samples. Below is an example of standard curve preparation:

<table>
<thead>
<tr>
<th>Conc. after 50ul of 1× PicoGreen added (ng/ul)</th>
<th>1ng/ul calf thymus DNA(ul)</th>
<th>TE buffer (ul)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>0.05</td>
<td>5</td>
<td>45</td>
</tr>
<tr>
<td>0.025</td>
<td>2.5</td>
<td>47.5</td>
</tr>
<tr>
<td>0.01</td>
<td>1</td>
<td>49</td>
</tr>
<tr>
<td>0.005</td>
<td>0.5</td>
<td>49.5</td>
</tr>
<tr>
<td>0.0025</td>
<td>0.25</td>
<td>49.75</td>
</tr>
</tbody>
</table>

3. **Sample Preparation for analysis**
Dilute the DNA samples. Try to make the diluted DNA concentration fall in the range of the standard curve. Use 50ul of diluted DNA sample for each assay.

4. **Assay Plate Preparation**
Greiner 96-well clear flat-bottom plates (Cat. No.: 655101) are used. Add 50ul of TE (as blank control), 50ul of diluted DNA samples to wells of plate. Then, add 50ul of 1× PicoGreen into blanks and samples, protected from light.

5. **Measurement and Data analysis**
Measure the fluorescence of samples using Tecan Genios or other suitable fluorimeter. The working software is Magellan. Fluorescence value of the reagent blank is subtracted from that of each sample and DNA concentration of samples is determined from the standard curve by the software.
Literature Cited


