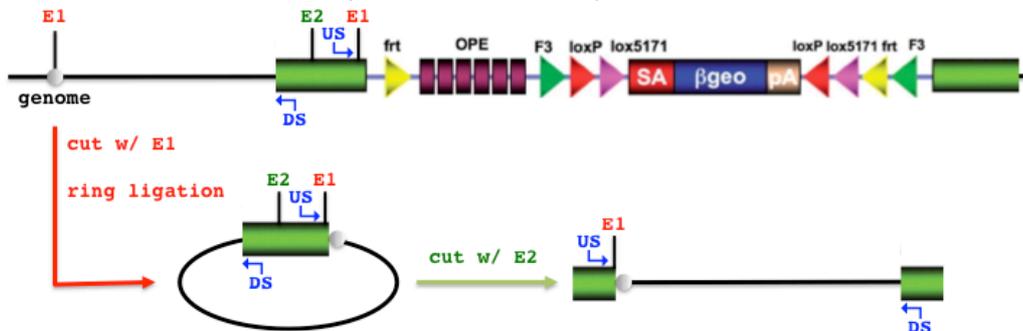


Inverse PCR: PCR Strategy to Map Genomic Integration Sites of Gene Trap Vectors by Next Generation Sequencing

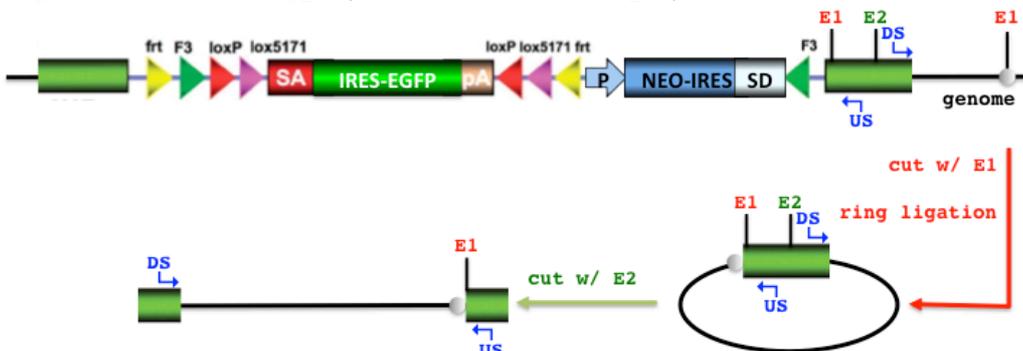
Enzyme1 (E1) is used to fragment the genome. As the recognition sequence for E1 is also present in the terminal repeat of the gene trap vector, it is possible to retrieve the exact integration site of the gene trap cassette within the genome by circularizing E1-digested gDNA (ring ligation) and subsequently amplifying the genomic region by inverse PCR (iPCR) using primers "US" and "DS". To improve iPCR efficiency, a linearization step using E2 was introduced, which re-opens the rings generated previously. Moreover, each integration site can be mapped by using two different enzymes E1. We would recommend you to split your samples and use both enzymes E1 on each sample in parallel.

Please keep in mind the "directionality" of the mapping strategies for different mutagens, as this affects the assignment of a particular mutation to the sense and anti-sense strand in the genome, respectively.

Mapping - 5' strategy (Lenti, Retro)



Mapping - 3' strategy (Tol2-GT, Tol2-pA)



- Mutagenesis systems

	Enzyme1 (E1)	Enzyme2 (E2)	Mapping Strategy
<input type="checkbox"/> Lenti	NlaIII, TaqI	PacI	5'
<input type="checkbox"/> Retro	NlaIII, MseI	SbfI	5'
<input type="checkbox"/> Tol2-GT	NlaIII, TaqI	PacI	3'
<input type="checkbox"/> Tol2-pA	NlaIII, TaqI	PacI	3'

- Solexa iPCR primer sequences

- DS_2
 AATGATACGGCGACCACCGAGATCTACACGAGCCAGAACCAGAAGGAACTTGAC
 - US_BCxx
 CAAGCAGAAGACGGCATAACGAGAT **INDEX** GTGACTGGAGTTCAGACGTGTGCTCTTC
- The DS primer is common for all PCR reactions, and binds in the genetrapp.
- The US primer contains a sequence that will bind to the oligo present on the NGS flowcell surface, an index of 4 to 8 bases, and then a sequence that binds in the genetrapp.
- The **INDEX** is a custom barcode (4-8bps) that allows to "label" all the PCR reactions from 1 complex sample. For each sample you will use two different US primers with two different indices: one for the PCR reaction of the "gDNA digested by E1-1" and one for the PCR reaction of the "gDNA digested by E1-2". For NGS, all samples can then be mixed and loaded into one NGS flowcell. With the index sequence you will be able to say from which original sample the PCR fragment comes from.

- Solexa flowcell primer sequences (i.e. custom 1st read)

- Lenti CAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA
- Retro 1:1 mix of
 - GAGTGATTGACTACCCGTCAGCGGGGGTCTTTCA
 - TGAGTGATTGACTACCCACGACGGGGGTCTTTCA
- Tol2-GT/Tol2-pA CACTTGAGTAAAATTTTTGAGTACTTTTTACACCTCTG
- standard Illumina 2nd read and index primers

- Reagents and Buffers

RNAse A (Qiagen, 19101, 100mg/ml)
 optional: a phase lock system
 Roti-Phenol/Chloroform/Isoamyl alcohol (Roth, A156.2)
 mixture of Chloroform:Isoamyl alcohol 24:1 (Sigma, Merck)
 Isopropanol, 70% Ethanol and TE Buffer
 Picogreen, Invitrogen
 Restriction enzymes and buffers, New England Biolabs
 T4 DNA ligase and buffer, Roche Applied Science

PCR kit (polymerase and buffer)
QIAquick PCR Purification Kit (Qiagen) or Sera Mag
Speedbeads (Thermo, 45152105080350) & Binding Buffer
QIAquick Gel Extraction Kit (Qiagen)

- **prepare Genomic DNA lysis buffer (GDLB) and autoclave it**
 - 10mM Tris-HCl pH 8.0
 - 5mM EDTA
 - 100mM NaCl
 - 1.0% SDS
 - 1 mg/ml Proteinase K

NGS-iPCR Protocol

1. Genomic DNA prep

- lyse cell pellet according to pellet size in 2-10mL GDLB
- inc. 55°C, o/n
- add 1:1000 RNase A (Qiagen, 100mg/ml) for ~1h, 37°C
- 2x Phenol/Chloroform/Isoamyl alcohol extractions (using a phase lock system)
- 1x Chloroform/Isoamyl alcohol extraction (using a phase lock system)
- Isopropanol precipitation of gDNA
- wash gDNA pellet w/ 70% EtOH
- resuspend in 0,1-2mL TE Buffer
- quantify DNA using Picogreen

2. Digestion of samples w/ enzymes E1 in parallel (per sample two separate reactions!)

gDNA 100ng/ μ l	40 μ l
10x CutSmart	8 μ l
Enzyme1	3 μ l
dH2O	<u>29 μl</u>
	80 μ l

inc. 37°C or 65°C (TaqI), o/n

3. Purify restriction digests (Qiagen PCR kit or Mag Speedbeads), elute in 100 μ l in total

4. Ring ligation

E1-gDNA	100 μ l
10x LB	120 μ l

T4 DNA Ligase	4 μ l
dH2O	<u>976μl</u>
	1200 μ l

inc. 16°C, o/n

5. Heat-inactivate T4 DNA Ligase at 65°C for 15min

6. Linearize samples w/ enzyme E2 at 37°C for 2h

RL-E1-gDNA	1200 μ l
Enzyme2	2 μ l

7. Purify restriction digest (Qiagen PCR kit or Mag Speedbeads), elute in 100 μ l in total

8. iPCR reaction (5-10 reactions per sample, process half or all of the digest according to the expected DNA amount)

E2-RL-E1-gDNA from 7.	10 μ l
primer US 100 uM	0.1 μ l
primer DS 100 uM	0.1 μ l
10 mM dNTP mix	1 μ l
10x pol. buffer	5 μ l
20x polymerase	3 μ l
dH2O	<u>30.8μl</u>
	50 μ l

Biorad C1000 cycle parameters

95deg, 3'	
95deg, 15"	37x
61deg, 25"	
72deg, 75"	
72deg, 5'	

You are now dealing with very big molar amounts of amplicons!!!! As NGS detects single molecules, there is a very (!!!) high danger of contaminating other experiments with post PCR samples. We recommend using different pipet sets, rooms and equipment for post PCR samples.

Pool all the PCR products from one sample (5-10 reactions) and analyze 20 μ l on an agarose gel. You should get a smear band, beginning at around 400bp. Estimate the DNA amount and mix all PCR samples (both separated digests for all

samples) in a way that you use from each one approximately the same DNA amount. Load that mixture on another agarose gel and cut the part of the smear between 400 and 800bp. Extract the DNA from the gel using Qiagen's Gel Extraction Kit and use it for NGS.

IMPORTANT NOTE:

Before executing the protocol, please consider the expected complexity. You will receive 100-200 million reads from 1 lane. As this is usually multiple times more than the complexity of your input, you might be oversampling the experiment. This can – by misalignment – cause artefacts. For that reason, it can be advantageous to combine several experiments in 1 lane by indexing. Subdividing single experiments may not be advisable, as the data are not simply additive. In order to optimally use NGS, consider the scale of input and experiment. If you wish e.g. to quantitatively use the NGS data, the 200 mio reads must stem from 200 mio independent genomes. If the complexity of input is significantly lower, the experiment can be scaled accordingly.