Haplobank

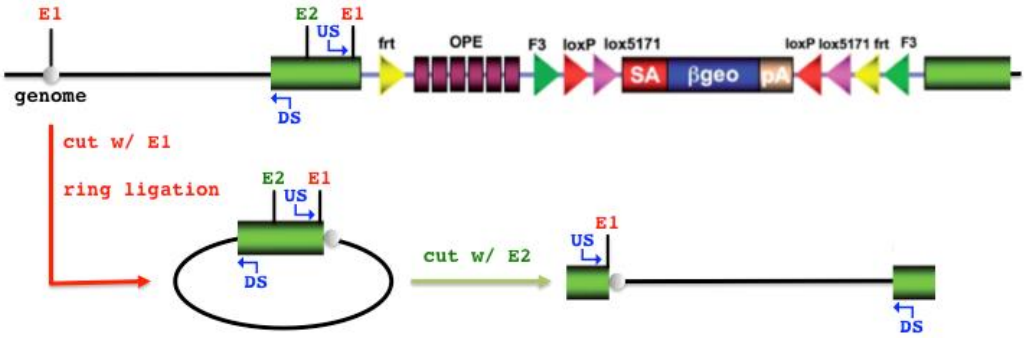
IMBA - Institute of Molecular Biotechnology GmbH
Dr. Bohr-Gasse 3
1030 Vienna, Austria

Telephone: + 43-1-79 044-4881
Telefax: + 43-1-79 044-110
E-Mail: office@haplobank.at

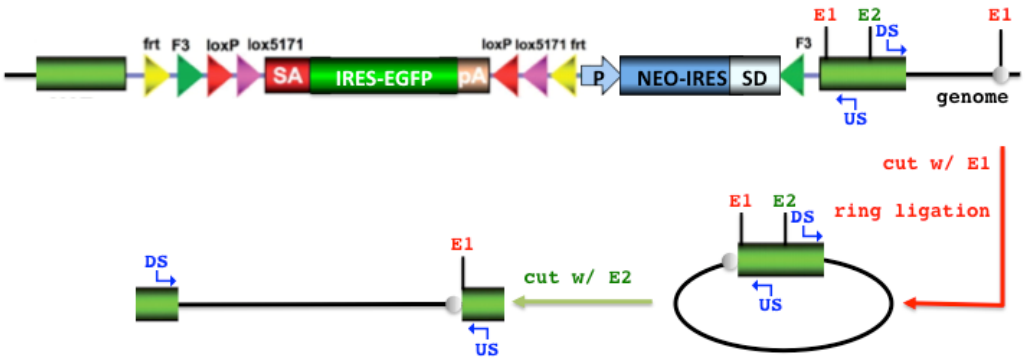
INVERSE PCR - PCR STRATEGY TO MAP GENOMIC INTEGRATION SITES OF GENE TRAP VECTORS BY NEXT GENERATION SEQUENCING

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 universal primers "US" and "DS". To improve iPCR efficiency, a linearization step using E2 was introduced, which re-opens the rings. Moreover, each integration site can be mapped by using two different Enzyme1s. We would recommend splitting the sample in two and using both enzymes E1 individually.
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-EGT, Retro-EGT)



Mapping - 3' strategy Tol2-EGT, Tol2-polyA)



Haplobank

IMBA - Institute of Molecular Biotechnology GmbH
Dr. Bohr-Gasse 3
1030 Vienna, Austria

Telephone: + 43-1-79 044-4881
Telefax: + 43-1-79 044-110
E-Mail: office@haplobank.at

MUTAGENISIS SYSTEMS	Enzyme 1 (E1)		Enzyme 2 (E2)	Mapping Strategy
Lenti-EGT	NlaIII	TaqI	PacI	5´
Retro-EGT	NlaIII	MseI	SbfI	5´
Tol2-EGT	NlaIII	TaqI	PacI	3´
Tol2-polyA	NlaIII	TaqI	PacI	3´

SOLEXA iPCR PRIMER SEQUENCES

DS_2: AATGATACGGCGACCACCGAGATCTACACGAGCCAGAACCAGAAGGAACTTGAC
US_BCxx: CAAGCAGAAGACGGCATACGAGATINDEXGTGACTGGAGTTCAGACGTGTGCTCTTC

The DS primer is common for all PCR reactions and binds to the gene trap.
The US primer contains a sequence that will bind with one part to the oligo present on the NGS flowcell surface, an index of 4 to 8 bases, and with the other to a sequence that binds to the gene trap. 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 analysis, all samples can then combined and loaded into one NGS flowcell. Based on the index sequence, it will be feasible to identify from which sample the PCR fragment originates.

SOLEXA FLOWCELL PRIMER SEQUENCES (i.e. custom 1st read)

Lenti-EGT	CAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA
Retro-EGT 1:1 mix of:	GAGTGATTGACTACCCGTCAGCGGGGGTCTTTCA
	TGAGTGATTGACTACCCACGACGGGGGTCTTTCA
Tol2-EGT/Tol2-polyA	CACTTGAGTAAAATTTTTGAGTACTTTTTACACCTCTG

Standard Illumina 2nd read and index primers

REAGENTS AND BUFFERS

RNAse A	Qiagen, 19101 (100mg/ml)
Roti-Phenol/Chloroform/Isoamyl alcohol	Roth, A156.2
mixture of Chloroform:Isoamyl alcohol 24:1	Sigma, Merck
Optional: Phase Lock Gel	VWR 733-2478
Isopropanol, 70% Ethanol, TE Buffer	
Picogreen	Invitrogen, P7589
Restriction enzymes and buffers	New England Biolabs
T4 DNA ligase and buffer	Roche Applied Science
PCR kit (polymerase and buffer)	
QIAquick PCR Purification Kit or	Qiagen, 28106
Sera-Mag SpeedBeads	GE 11829912
illustra ExoProStar 1-Step Kit	GE 11911411

Haplobank

IMBA - Institute of Molecular Biotechnology GmbH
Dr. Bohr-Gasse 3
1030 Vienna, Austria

Telephone: + 43-1-79 044-4881
Telefax: + 43-1-79 044-110
E-Mail: office@haplobank.at

Preparation of Genomic DNA lysis buffer (GDLB):

10mM Tris-HCl pH 8.0
5mM EDTA
100mM NaCl
1.0% SDS
1 mg/ml Proteinase K
→ autoclave

NGS-iPCR PROTOCOL

1. **Genomic DNA prep**
   - Lyse cell pellet according to pellet size in 0,5-5mL GDLB
   - Incubate at 55˚C, over night
   - Add 1:1000 RNAse A (100mg/ml) for ~1h at 37˚C
   - 2x Phenol/Chloroform/Isoamyl alcohol extractions (using Phase Lock Gel)
   - 1x Chloroform/Isoamy alcohol extraction (using a phase lock system)
   - Isopropanol precipitation of gDNA
   - Wash gDNA pellet with 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 seperate reactions!)**

gDNA 100ng/µl	40 µl
10x CutSmart	8 µl
Enzyme1	3 µl
dH2O	29 µl
	80 µl

→ incubate 37˚C or 65˚C (TaqI), over night

2. **Purify restriction digests (Qiagen PCR Purification Kit or Sera-Mag SpeedBeads), elute in 100µl in total**

3. **Ring ligation**

E1-gDNA	199 µl
10x LB	120 µl
T4 DNA Ligase	4 µl
dH2O	976 µl
	1200 µl

→ incubate 16˚C, over night

4. **Heat-inactivate T4 DNA Ligase at 65˚C for 15min**

Haplobank

IMBA - Institute of Molecular Biotechnology GmbH
Dr. Bohr-Gasse 3
1030 Vienna, Austria

Telephone: + 43-1-79 044-4881
Telefax: + 43-1-79 044-110
E-Mail: office@haplobank.at

5. **Linearize samples w/ enzyme E2 at 37˚C for 2h**

RL-E1-gDNA	1200 µl
Enzyme2	2 µl

6. **Purify restriction digest (Qiagen PCR kit or Sera-Mag SpeedBeads), elute in 100µl in total**

7. **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 µl
primer US 100 µM	0.1µl
primer DS 100 µM	0.1µl
10 mM dNTP mix	1 µl
10x Polymerase Buffer	5 µ
20x Polymerase	3 µl
dH2O	30.8µl
	50 µl

8. **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 PCR products from one sample (5-10 reactions) and analyze 20µl on an agarose gel. You should get a smear band, beginning at around 400bp.
- Estimate the DNA concentration and combine the individual PCR samples (both separated digests for all samples) using equal amounts.
- Load the mix on a fresh agarose gel and excise a band positioned between 400-800bp. Extract the DNA from the gel using Qiagen’s Gel Extraction Kit and submit the DNA to NGS analysis.

Haplobank

IMBA - Institute of Molecular Biotechnology GmbH
Dr. Bohr-Gasse 3
1030 Vienna, Austria

Telephone: + 43-1-79 044-4881
Telefax: + 43-1-79 044-110
E-Mail: office@haplobank.at

IMPORTANT NOTE:

Before performing this type of analysis, 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 is 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 should be be scaled accordingly.