

Mapping summary

454 Sequencing & read target enrichment analysis

MGP

January 28, 2013

On-target enrichment calculation

- ▶ **On-target reads:** Bam files were intersected against genes ($\pm 2\text{Kb}$) in gencode15.gtf:

```
intersectBed -abam <bamfile> -b gencode15.bed \  
> targetbam
```

- ▶ **Off-target and on-gencode reads:**

1. Reads not mapped on-target ($\pm 2\text{Kb}$):

```
intersectBed -v -abam <bamfile> -b gencode15.bed \  
> offtargetbam
```

2. Resultant set (from 1.) was intersected against genes ($\pm 2\text{Kb}$) in the full gencode file:

```
intersectBed -abam offtargetbam \  
-b gencode.v7.annotation.bed > offgencode
```

On-target enrichment on gencode (v7) genes ($\pm 2\text{Kb}$)

RACE	Tissue	Mapped reads	On % target	On-gencode off-target %	On-gencode off-target reads/Kb [A]	On-target reads/Kb [B]	Enrichment [B]/[A]
3'	brain	496841	6.2	86.5	0.2	1.4	5.9
	heart	452807	3.1	90.2	0.2	0.6	2.9
	kidney	269083	7.2	85.0	0.1	0.9	7.0
	liver	217409	4.1	87.0	0.1	0.4	3.9
	lung	402061	2.7	92.6	0.2	0.5	2.4
	spleen	466363	2.8	91.9	0.2	0.6	2.5
	testis	371106	8.5	80.9	0.2	1.4	8.7
5'	heart	434678	3.7	85.4	0.2	0.7	3.6
	kidney	482098	2.5	71.2	0.2	0.5	2.9
	liver	542400	1.0	42.1	0.1	0.3	2.0
	lung	467009	5.1	89.4	0.2	1.1	4.8
	spleen	497850	2.0	85.3	0.2	0.5	1.9
	testis	632212	14.4	70.6	0.2	4.2	17.0

- ▶ Enrichment found for all tissues (1.9x-17.0x), being significantly greater in testis results (8.7x and 17.0x).

Assembly mapping: on-target enrichment on gencode (v7) genes ($\pm 2\text{Kb}$)

RACE	Tissue	Mapped contigs	On % target	On-gencode off-target %	On-gencode off-target reads/Kb [A]	On-target reads/Kb [B]	Enrichment [B]/[A]
3'	brain	11173	2.9	91.0	0.006	0.015	2.6
	heart	9519	2.6	93.9	0.005	0.011	2.3
	kidney	5179	5.1	90.8	0.003	0.012	4.7
	liver	4638	3.0	92.4	0.002	0.006	2.7
	lung	9198	2.5	94.1	0.005	0.010	2.2
	spleen	11492	2.4	93.7	0.006	0.012	2.1
	testis	9406	5.7	86.3	0.004	0.024	5.5
3'	heart	7377	3.8	91.8	0.004	0.013	3.5
	kidney	10736	2.4	91.6	0.005	0.012	2.2
	liver	8015	2.3	89.8	0.004	0.009	2.2
	lung	10681	3.4	93.0	0.005	0.017	3.0
	spleen	14248	1.8	93.2	0.007	0.012	1.6
	testis	15498	4.7	83.8	0.007	0.033	4.6

Reads in genecode genes (v7) and intergenic regions

- ▶ Genes in `genecode.v7.annotation.gtf` were extended 2Kb, 15Kb and 50Kb upstream/downstream
- ▶ Intergenic reads were considered to be those mapping out of the extended `genecode.v7.annotation.gtf`

Reads in genecode (v7) genes and intergenic regions

RACE	Tissue	Mapped reads	% reads on genes $\pm 2\text{Kb}$	% Intergenic reads (genes $\pm 2\text{Kb}$)	% Intergenic reads (genes $\pm 15\text{Kb}$)	% Intergenic reads (genes $\pm 50\text{Kb}$)
3'	brain	496841	92.6	7.4	3.0	1.8
	heart	452807	93.3	6.7	4.1	3.3
	kidney	269083	92.2	7.8	5.0	4.1
	liver	217409	91.0	9.0	6.3	5.1
	lung	402061	95.3	4.7	1.8	0.9
	spleen	466363	94.6	5.4	2.2	1.0
	testis	371106	89.3	10.7	6.0	2.7
5'	heart	434678	89.1	10.9	9.0	8.5
	kidney	482098	73.6	26.4	24.3	23.5
	liver	542400	43.1	56.9	55.8	55.1
	lung	467009	94.4	5.6	2.7	2.2
	spleen	497850	87.3	12.7	10.3	9.3
	testis	632212	84.9	15.1	10.6	7.8