

Mapping summary

454 Sequencing & read target enrichment analysis

MGP

December 18, 2012

On target reads

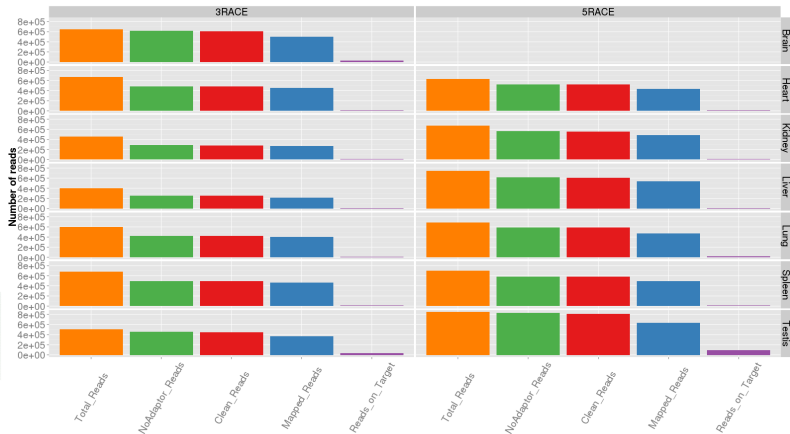
- ▶ ± 2000 bases were added to minimum and maximum target regions of each gene in gencode15.gtf.
- ▶ On target reads were extracted from bam files obtained by BLAT (95% identity):

```
intersectBed -abam <bamfile> \  
-b gencode15.nochr.extended.bed > <newbam>
```

On-target reads stats

RACE	Tissue	Total mapped %	Total split %	Mapped on target %	Split on target %
3'	Kidney	94.6	20.9	6.8	25.5
	Lung	94.7	14.5	2.5	17.5
	Liver	85.3	23.5	3.5	28.3
	Spleen	94.7	11.9	2.7	13.8
	Heart	93.8	15.5	2.9	26.9
	Testis	82.5	22.1	7.0	31.4
	Brain	81.1	12.7	5.0	15.2
5'	Kidney	86.2	23.5	2.1	28.5
	Lung	79.0	46.2	4.0	57.5
	Liver	88.9	12.0	0.9	30.3
	Spleen	85.2	16.7	1.7	27.3
	Heart	83.0	42.4	3.1	45.8
	Testis	77.7	31.9	11.2	63.1

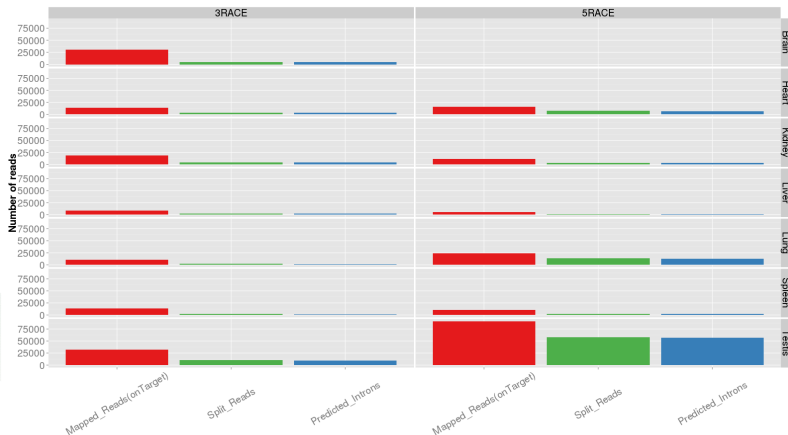
On-target reads stats



On-target reads stats

RACE	Tissue	Mapped on target %	Split on target %	% Introns on target
3'	Kidney	6.8	25.5	23.8
	Lung	2.5	17.5	14.1
	Liver	3.5	28.3	26.5
	Spleen	2.7	13.8	11.1
	Heart	2.9	26.9	23.9
	Testis	7.0	31.4	29.9
	Brain	5.0	15.2	13.6
5'	Kidney	2.1	28.5	26.4
	Lung	4.0	57.5	54.9
	Liver	0.9	30.3	28.1
	Spleen	1.7	27.3	25.3
	Heart	3.1	45.8	44.2
	Testis	11.2	63.1	61.8

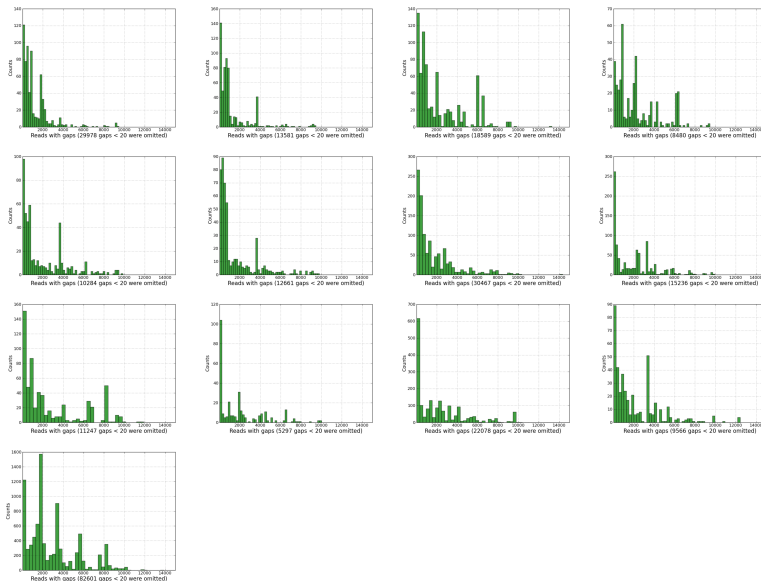
On-target reads stats



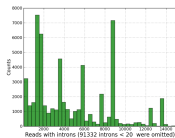
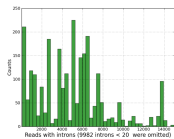
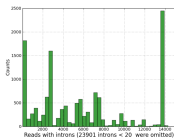
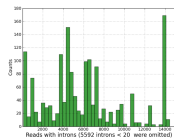
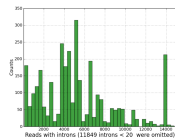
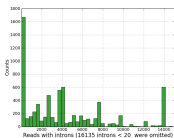
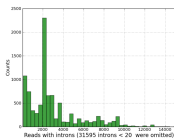
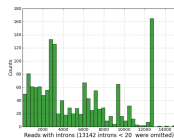
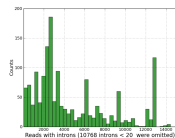
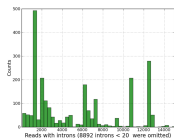
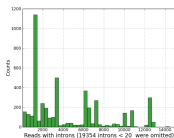
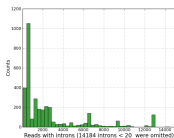
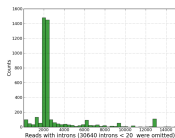
On-target reads stats

- ▶ Very low percentage of on-target reads (0.9 – 11.2).
- ▶ ~ 60% of split reads in 5' Lung and 5' Testis samples.

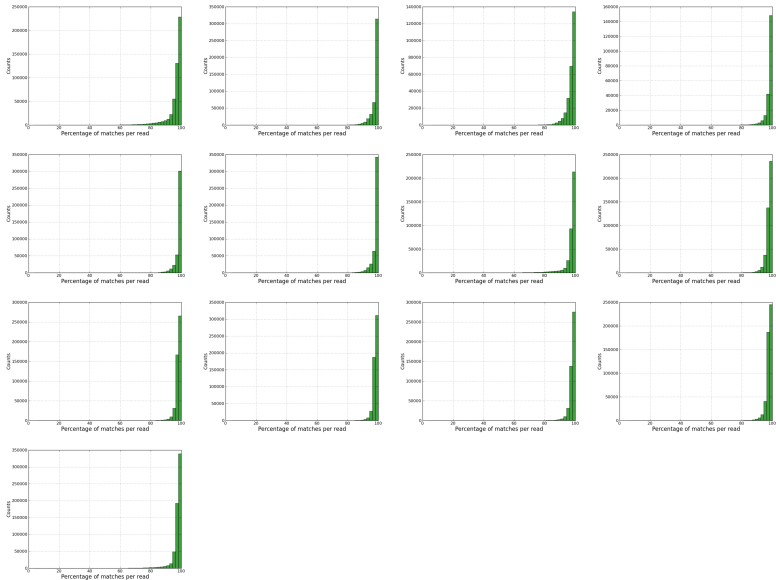
Distribution of the size of the longest gap in a read



Distribution of sizes of the longest intron in a read



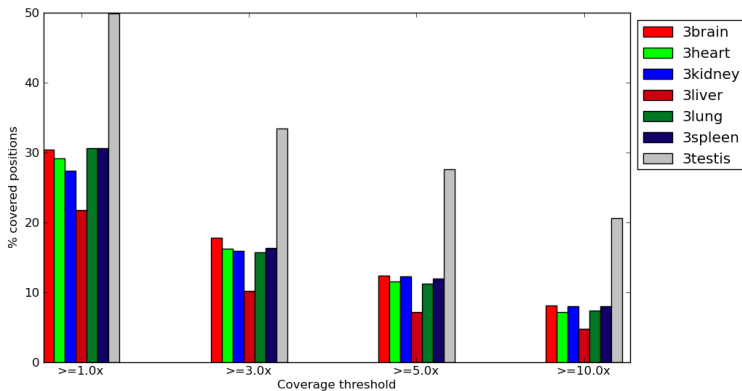
% Mapped bases per total reads



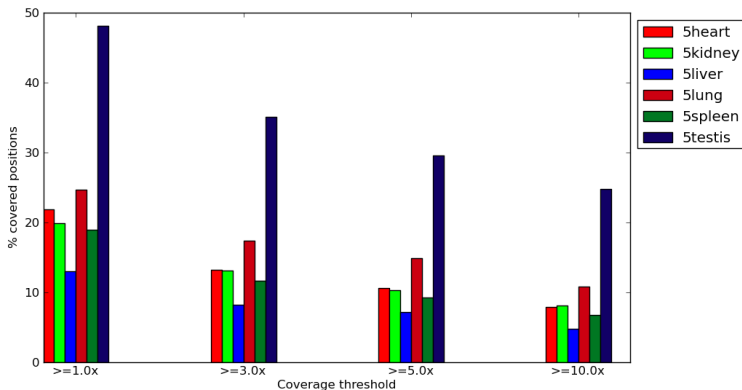
% Transcript positions covered

- ▶ A transcript region was built selecting exons and UTR regions for each gene from gencode15.gtf.
- ▶ We checked the percentage of positions on this transcript region that were covered in each bam file.
- ▶ We need the exact genomic coordinates of the primers used for 5' and 3' RACE in order to make a more accurate analysis for only the amplified RACE regions.

% Transcript positions covered (3' RACE)



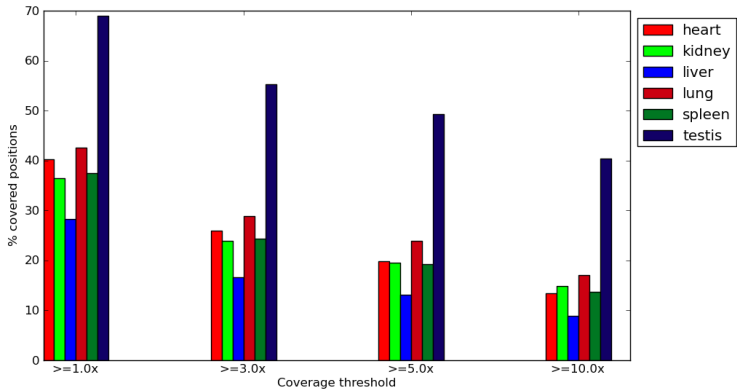
% Transcript positions covered (5' RACE)



% Transcript positions covered

- ▶ We merged each 3' bam with its corresponding 5' mate, obtaining 6 *merged* bams for heart, kidney, liver, lung, spleen and testis.
- ▶ We checked the percentage of positions of the transcript region that were covered in each of these 6 bams.

% Transcript positions covered (merged bam)



Conclusions

- ▶ Low number of on-target reads.
- ▶ No overall significant differences are found in the percentage of split reads between reads mapped on target and total mapped reads, except for 5' testis, 5' spleen, 5' liver, 3' testis and 3' heart
- ▶ Testis results outperforms the results obtained for the rest of samples.