In the Illumina run, SA12021, base calling was done by RTA. But because the lane chosen as control lane was found to have skewed base composition, off line base calling was done using Lane 4 as the control lane. The control lane is used to calculate cross talk matrices and prephasing/phasing corrections, that are applied on all lanes when base calling.

Two samples, Sample_SET2_RAP2 and Sample_R1_1_L002_R1 with and without OLB.

The original fastq files are at:

Without OLB: /raid/jcruz/corral/SA12021/Unaligned_beforeOLB
With OLB: /raid/jcruz/corral/SA12021/Unaligned

The fasta and quality files made by dhivya from the fastq files are at: /home/daras/tmp/olb_comp/data

Number of reads passing filter is almost the same.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Reads passing filter without OLB</th>
<th>Reads passing filter with OLB</th>
</tr>
</thead>
<tbody>
<tr>
<td>SET2_RAP</td>
<td>208104</td>
<td>208148</td>
</tr>
<tr>
<td>R1_1</td>
<td>457376</td>
<td>457502</td>
</tr>
</tbody>
</table>

The base distributions and quality distributions (by base position) look similar with and without OLB (graphs for without OLB, followed by with OLB)
Dominant sequences look very much the same (looking at the first 30 bases). Adaptor count (simply by grepping for GATCGGAAGA) is very similar. The diff output for the first 1000 and last 1000 sequences was also looked at. For the first 1000 reads, ~300 of them had differences, but it looked like only in one base. For the last 1000 reads, ~400 reads were different, but again by one or two bases.

Need to align both data files and check on alignment results.