Sequence variation
– step-by-step solutions to exercises

Session 3 – Dr Jason Wong

Introductory bioinformatics for human genomics workshop, UNSW
31st July 2014 – 1st August 2014
Exercises

1. Retrieve nonsense SNPs from the 1000 Genomes project for MLH1 using the Table Browser.

2. Also visualise this in the Genome Browser and find out what substitution has caused the nonsense SNPs.

3. How many of these SNPs are also COSMIC mutations? (Hint: requires intersection function in Table browser – or visualise directly in genome browser)
1. Retrieve missense SNPs from the 1000 Genomes project for MLH1

1. Browse to MLH1 in UCSC
2. Go to table browser

Select “Table browser” from “Tools” menu”

Select track

Select filter

Check “position”. The coordinates should be where the browser last was.
1. Check 1000genomes
2. Check nonsense
There are two SNPs

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Start Position</th>
<th>End Position</th>
<th>SNP ID</th>
<th>Value</th>
<th>Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr3</td>
<td>37089129</td>
<td>37089130</td>
<td>rs35001569</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>chr3</td>
<td>37090505</td>
<td>37090506</td>
<td>rs63750114</td>
<td>0</td>
<td>+</td>
</tr>
</tbody>
</table>
2. Also visualise this in the Genome Browser and find out what substitution has caused the nonsense SNPs.

Go back to Genome Browser

Right click anywhere on ‘All SNPs track’
Check ‘stop_gain’ – this has the same meaning as ‘nonsense’.

Change Splice site to ‘gray’ if saved from before.
Here are the two nonsense SNPs

Change track view to ‘pack’ to get the SNP ids
Note in both cases, both are tiallelic SNPs. However, the ‘T’ allele in both cases appears to be very rare and not actually a 1000 Genomes SNP. The SNP only passed our filter because it is being called a 1000 Genome SNP based on the other allele while the rare ‘T’ allele caused a nonsense SNP.
3. How many MLH1 COSMIC somatic mutations are also 1000 Genome SNPs?

1. Browse to MLH1 in UCSC
2. Go to table browser

Select “Table browser” from “Tools” menu”

Select track

Check “position”. The coordinates should be where the browser last was.

Select filter

Select “Table browser” from “Tools” menu”

Select track

Check “position”. The coordinates should be where the browser last was.

Select filter

Select “Table browser” from “Tools” menu”

Select track

Check “position”. The coordinates should be where the browser last was.

Select filter
1. Check 1000genomes

2. Check nonsense
Create intersection
Select group “Phenotype and Literature”
Then “COSMIC” track

<table>
<thead>
<tr>
<th>Intersect with All SNPs(138)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select a group, track and table to intersect with:</td>
</tr>
<tr>
<td><strong>group:</strong> Phenotype and Literature  <strong>track:</strong> COSMIC</td>
</tr>
<tr>
<td><strong>table:</strong> COSMIC (cosmic)</td>
</tr>
</tbody>
</table>

Intersect All SNPs(138) items with bases covered by COSMIC:

- [ ] All All SNPs(138) records that have any overlap with COSMIC
- [ ] All All SNPs(138) records that have no overlap with COSMIC
- [ ] All All SNPs(138) records that have at least **80** % overlap with COSMIC
- [ ] All All SNPs(138) records that have at most **80** % overlap with COSMIC

Intersect bases covered by All SNPs(138) and/or COSMIC:

- [ ] Base-pair-wise intersection (AND) of All SNPs(138) and COSMIC
- [ ] Base-pair-wise union (OR) of All SNPs(138) and COSMIC

Check the following boxes to complement one or both tables. To complement a table means to include a base pair in the intersection/union if it is not included in the table.

- [ ] Complement All SNPs(138) before base-pair-wise intersection/union
- [ ] Complement COSMIC before base-pair-wise intersection/union

[submit] [cancel]
One of the SNP is also a COSMIC mutation
Display COSMIC mutations as dense
Zoom in region where the nonsense SNPs are
COSMIC mutation lines up with rs35001569 but not rs63750114
Although there are two mutations affecting the site of this SNP, neither are actually nonsense mutations. 1422601 is a missense substitution while 26083 is an in-frame deletion.