***EXPLANATION***

Because of Galaxy bug, Martin and I haven't been able to find a way to get a list of <u>overlapping pieces of intervals</u> using scripts on MB06. Probably there is a way, according to Martin's opinion, but it is really tricky.

This kind of analysis is easier with Galaxy (there is a specific tool: Operate on genomic intervals → Intersect → Overlapping pieces of intervals) but unfortunately the bug came out every time we tried to use it. This is the outcome you can get using this Galaxy tool:
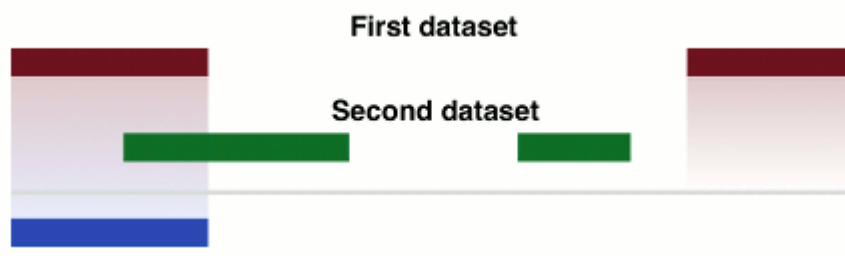


*Overlapping pieces of Intervals returns intervals that indicate the exact base pair overlap between the first dataset and the second dataset.*

And the blue regions in this example are exactly the type of result we want!

However, as I mentioned before, it is pretty difficult to get this kind of results using scripts (according to Martin), so we have found an easier way…

So instead of the intersection to get overlapping pieces of intervals, we can directly intersect the motifs with CNVs: this type of intersection tool gives you a list of those motifs which overlaps with big CNVs (see the figure down below).



*Overlapping Intervals returns entire intervals from the first dataset that overlap the second dataset. The returned intervals are completely unchanged, and this option only filters out intervals that do not overlap with the second dataset. [Galaxy]*

<u>Apparently this seems to be the wrong type of intersection</u>, because it gives you the regions of the first dataset which intersect one or more regions of the second dataset (it doesn't give you the start and end of each intersected region).

But, <u>if you think about the dimensions (length) of motif regions and of CNVs regions</u>, you can figure out how we can use this type of intersection to get exactly what we want!

I try to explain as clearer as I can… (otherwise for more explanation you can ask to Martin).
Reasoning: **because motif regions are <u>really small </u>regions (~ 19 bp) while the deletions/duplications** (found from patients) **are <u>really big</u>**, we can assume that **probably the motifs are** not partially but **completely included or excluded in these big deletions/duplications** (see the figure down below).

**1 dataset** { **Motifs**

**2 dataset** { **Deletions**

**Duplications**

**Intersecting regions**