

Count UMIs

```
cat $BATCH | while read PCR_WELL; do
  awk '{
    split($4, arr, "|");
    if (!seen[arr[1]]) {
      seen[arr[1]] = 1;
      count[arr[2]]++;
    }
  } END {
    for (sample in count)
      print sample "\t" count[sample];
  }' $RMDUP_SPLIT_BED_DIR/$PCR_WELL.bed \
  | sort -k1,1 \
  >$OUTPUT_DIR/$PCR_WELL.UMI.count
```



```
cat $BATCH | while read PCR_WELL; do
    awk '{
        split($4, arr, "|");
        if (!seen[arr[1]]) {
            seen[arr[1]] = 1;
            count[arr[2]]++;
        }
    } END {
        for (sample in count)
            print sample "\t" count[sample];
    }' $RMDUP_SPLIT_BED_DIR/$PCR_WELL.bed \
    | sort -k1,1 \
    >$OUTPUT_DIR/$PCR_WELL.UMI.count
```

← Do this stuff on every line

← Do this stuff once all the lines have been processed

Awk has a rich library of functions

Function	Behavior
substr	Extract a substring from a string
split	Split up a string according to separator characters
index	Find a string inside of another string
length	Return the length of a string
strtonum	Convert a string to a number
sprintf	Create a string from data according to a format
sub	Replace a part of a string with another string