HLA typing: whole exomes

Clemens Messerschmidt

21.09.2015

For 11 exomes, which were used in the ATHLATES paper and have known HLA types, we performed HLA typing with the tool optitype.

The following table shows the reference types and the optitype predictions with 4-digit resolution:

```
print(all_alleles)
```

```
##           V2         V3         V4         V5         V6         V7
## SRR359301 "A*02:03"  "A*31:01"  "B*13:01"  "B*48:01" "nottyped" "C*03:04"
## SRR360148 "A*02:01"  "A*36:01"  "nottyped" "B*40:01" "C*03:04"  "nottyped"
## SRR359098 "nottyped" "A*68:02"  "nottyped" "B*53:01" "nottyped" "nottyped"
## SRR359102 "A*30:02"  "A*66:01"  "nottyped" "B*41:02" "C*05:01"  "nottyped"
## SRR359108 "nottyped" "A*68:02"  "nottyped" "B*53:01" "nottyped" "nottyped"
## SRR359298 "nottyped" "A*24:07"  "B*27:04"  "B*39:05" "C*08:01"  "C*12:02"
## SRR360655 "A*30:02"  "nottyped" "nottyped" "B*57:03" "C*02:10"  "nottyped"
## SRR360288 "A*02:01"  "A*02:11"  "B*15:04"  "B*35:05" "C*01:02"  "nottyped"
## SRR359103 "nottyped" "A*02:01"  "B*18:01"  "B*57:01" "nottyped" "nottyped"
## SRR359295 "A*02:03"  "nottyped" "B*35:03"  "B*55:02" "C*04:01"  "C*12:03"
## SRR360391 "A*02:01"  "nottyped" "nottyped" "B*40:02" "C*03:04"  "nottyped"
##           V2         V3         V4         V5         V6         V7
## SRR359301 "A*02:03" "A*31:01" "B*13:01" "B*48:01" "C*03:03" "C*03:04"
## SRR360148 "A*02:01" "A*36:01" "B*35:01" "B*40:01" "C*03:04" "C*04:01"
## SRR359098 "A*03:01" "A*68:02" "B*35:01" "B*53:01" "C*04:01" "C*04:01"
## SRR359102 "A*30:02" "A*66:01" "B*18:01" "B*41:02" "C*05:01" "C*17:01"
## SRR359108 "A*03:01" "A*68:02" "B*35:01" "B*53:01" "C*04:01" "C*04:01"
## SRR359298 "A*11:02" "A*24:07" "B*27:04" "B*39:05" "C*08:01" "C*12:02"
## SRR360655 "A*30:02" "A*74:01" "B*15:03" "B*57:03" "C*02:10" "C*07:01"
## SRR360288 "A*02:01" "A*02:11" "B*15:04" "B*35:05" "C*01:02" "C*04:01"
## SRR359103 "A*01:01" "A*02:01" "B*18:01" "B*57:01" "C*07:01" "C*07:01"
## SRR359295 "A*02:03" "A*03:01" "B*35:03" "B*55:02" "C*04:01" "C*12:03"
## SRR360391 "A*02:01" "A*68:01" "B*07:02" "B*40:02" "C*03:04" "C*07:02"
```

For all alleles that possess a precise enough reference, prediction and reference were compared. The number of possible hits is the number of alleles per sample that is typed with a precision >= 4-digit, e.g. HLA-A*02:01

```
print(accordance)
```

```
##           optitype possible hits
## SRR359301        5             5
## SRR360148        4             4
## SRR359098        2             2
## SRR359102        4             4
## SRR359108        2             2
## SRR359298        5             5
## SRR360655        3             3
## SRR360288        5             5
```

```
## SRR359103        3             3
## SRR359295        5             5
## SRR360391        3             3
```

```
colSums(accordance)
```

```
##      optitype possible hits
##            41            41
```

```
colSums(accordance)/ colSums(accordance)[2]
```

```
##      optitype possible hits
##             1             1
```

On our small sample set of 11 wes, optitype correctly predicts 66/66 for 2-digit and 41/41 for 4-digit resolution, i.e. 100 % accordance with the reference.