HLA typing: whole exomes

Clemens Messerschmidt

21.09.2015

For 12 panels, which were used in the HLAssign paper and have known HLA types, we performed HLA typing with the tools optitype, bwakit, hlassign.

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

```
print(all_alleles)
```

```
##             V2         V3         V4         V5         V6
## CELL_ID_108 "nottyped" "nottyped" "nottyped" "nottyped" "nottyped"
## CELL_ID_109 "A*24:03"  "A*33:03"  "B*15:12"  "B*46:01"  "C*01:02"
## CELL_ID_122 "A*02:01"  "A*24:23"  "B*15:01"  "nottyped" "C*03:03"
## CELL_ID_13  "A*02:05"  "A*32:01"  "B*40:01"  "B*49:01"  "C*03:04"
## CELL_ID_16  "A*11:01"  "A*24:03"  "B*15:02"  "B*55:02"  "nottyped"
## CELL_ID_163 "nottyped" "nottyped" "B*07:02"  "B*44:17"  "nottyped"
## CELL_ID_165 "nottyped" "nottyped" "B*35:31"  "nottyped" "nottyped"
## CELL_ID_18  "A*02:01"  "A*32:01"  "B*15:01"  "B*27:08"  "C*03:04"
## CELL_ID_21  "nottyped" "nottyped" "nottyped" "nottyped" "nottyped"
## CELL_ID_235 "A*02:01"  "A*31:01"  "B*44:02"  "B*51:01"  "C*14:02"
## CELL_ID_36  "A*24:10"  "A*29:01"  "B*07:05"  "B*51:01"  "C*07:04"
## CELL_ID_38  "A*30:02"  "A*66:02"  "B*18:01"  "B*58:01"  "C*07:01"
## CELL_ID_39  "A*01:02"  "A*66:01"  "B*58:01"  "B*58:02"  "C*03:02"
## CELL_ID_41  "A*11:01"  "A*26:01"  "B*07:05"  "B*55:02"  "C*01:02"
## CELL_ID_45  "A*02:01"  "A*33:03"  "B*15:16"  "B*44:03"  "C*04:01"
## CELL_ID_56  "A*02:11"  "A*68:01"  "B*35:05"  "B*40:04"  "C*03:04"
## CELL_ID_6   "A*03:01"  "A*30:01"  "B*15:10"  "B*58:01"  "C*07:18"
## CELL_ID_79  "A*24:02"  "A*24:02"  "nottyped" "nottyped" "nottyped"
## CELL_ID_94  "A*02:01"  "A*24:02"  "B*51:01"  "B*54:01"  "C*01:02"
## CELL_ID_99  "A*03:01"  "A*03:01"  "B*47:01"  "B*47:01"  "C*06:02"
##             V7         V2         V3         V4         V5         V6
## CELL_ID_108 "nottyped" "A*03:01" "A*31:01" "B*15:18" "B*35:03" "C*04:01"
## CELL_ID_109 "nottyped" "A*24:03" "A*33:03" "B*15:12" "B*46:01" "C*01:02"
## CELL_ID_122 "C*03:04"  "A*02:01" "A*24:23" "B*15:01" "B*51:01" "C*03:03"
## CELL_ID_13  "C*07:01"  "A*02:05" "A*32:01" "B*40:01" "B*49:01" "C*03:04"
## CELL_ID_16  "C*12:03"  "A*24:02" "A*25:01" "B*18:01" "B*51:09" "C*01:02"
## CELL_ID_163 "nottyped" "A*03:01" "A*32:01" "B*07:02" "B*44:02" "C*05:01"
## CELL_ID_165 "C*12:04"  "A*02:01" "A*02:01" "B*35:31" "B*57:01" "C*02:02"
## CELL_ID_18  "C*06:02"  "A*02:01" "A*32:01" "B*15:01" "B*27:08" "C*03:04"
## CELL_ID_21  "nottyped" "A*02:01" "A*02:01" "B*44:02" "B*51:01" "C*07:04"
## CELL_ID_235 "C*16:04"  "A*02:01" "A*31:01" "B*44:02" "B*51:01" "C*14:02"
## CELL_ID_36  "C*15:05"  "A*24:10" "A*29:01" "B*07:05" "B*51:01" "C*07:04"
## CELL_ID_38  "C*07:01"  "A*03:01" "A*11:01" "B*07:02" "B*51:01" "C*07:02"
## CELL_ID_39  "C*06:02"  "A*01:02" "A*66:01" "B*58:01" "B*58:01" "C*03:02"
## CELL_ID_41  "C*07:02"  "A*11:01" "A*26:01" "B*07:05" "B*55:02" "C*01:02"
## CELL_ID_45  "C*14:02"  "A*02:01" "A*33:03" "B*15:16" "B*44:03" "C*04:01"
## CELL_ID_56  "C*04:01"  "A*02:11" "A*68:01" "B*35:05" "B*40:04" "C*03:04"
## CELL_ID_6   "C*08:04"  "A*03:01" "A*30:01" "B*15:10" "B*58:01" "C*07:01"
## CELL_ID_79  "nottyped" "A*24:02" "A*24:02" "B*52:01" "B*52:01" "C*12:02"
```

```
## CELL_ID_94  "C*14:02"  "A*01:01" "A*24:02" "B*08:01" "B*44:06" "C*05:01"
## CELL_ID_99  "C*06:02"  "A*03:01" "A*03:01" "B*47:01" "B*47:01" "C*06:02"
##              V7        V2         V3        V4         V5        V6
## CELL_ID_108 "C*07:04" "A*03:205" "A*31:21" "B*15:18"  "B*35:03" "C*04:01"
## CELL_ID_109 "C*03:03" "A*24:03"  "A*33:03" "B*15:12"  "B*46:01" "C*01:51"
## CELL_ID_122 "C*03:04" "A*02:01"  "A*24:23" "B*15:01"  "B*51:01" "C*03:03"
## CELL_ID_13  "C*07:01" "A*02:05"  "A*32:53" "B*40:221" "B*49:01" "C*03:04"
## CELL_ID_16  "C*12:03" "A*24:02"  "A*25:01" "B*18:01"  "B*51:09" "C*01:02"
## CELL_ID_163 "C*07:02" "A*03:01"  "A*32:01" "B*07:02"  "B*44:17" "C*05:01"
## CELL_ID_165 "C*06:02" "A*02:01"  "A*02:01" "B*35:31"  "B*57:01" "C*02:02"
## CELL_ID_18  "C*06:02" "A*02:01"  "A*32:01" "B*15:01"  "B*27:08" "C*03:04"
## CELL_ID_21  "C*14:02" "A*02:01"  "A*02:01" "B*44:02"  "B*51:01" "C*07:04"
## CELL_ID_235 "C*16:04" "A*02:01"  "A*31:01" "B*44:02"  "B*51:01" "C*14:02"
## CELL_ID_36  "C*15:05" "A*24:10"  "A*29:01" "B*07:05"  "B*51:01" "C*07:04"
## CELL_ID_38  "C*15:06" "A*11:117" "A*68:71" "B*07:26"  "B*51:01" "C*03:04"
## CELL_ID_39  "C*06:02" "A*01:02"  "A*66:01" "B*58:01"  "B*58:01" "C*03:02"
## CELL_ID_41  "C*07:02" "A*11:01"  "A*26:01" "B*07:06"  "B*59:01" "C*01:02"
## CELL_ID_45  "C*14:02" "A*02:01"  "A*33:03" "B*15:16"  "B*44:03" "C*04:01"
## CELL_ID_56  "C*04:01" "A*02:11"  "A*68:01" "B*35:05"  "B*40:04" "C*03:04"
## CELL_ID_6   "C*08:04" "A*03:62"  "A*30:01" "B*15:10"  "B*58:01" "C*07:18"
## CELL_ID_79  "C*12:02" "A*24:02"  "A*24:02" "B*52:01"  "B*52:01" "C*12:02"
## CELL_ID_94  "C*07:01" "A*01:01"  "A*24:02" "B*08:01"  "B*44:06" "C*05:01"
## CELL_ID_99  "C*06:02" "A*03:01"  "A*03:01" "B*47:01"  "B*47:01" "C*06:02"
##              V7        V2        V3        V4        V5        V6
## CELL_ID_108 "C*07:04" "A*03:01" "A*31:01" "B*15:18" "B*35:03" "C*04:01"
## CELL_ID_109 "C*03:04" "A*24:03" "A*33:03" "B*15:01" "B*46:01" "C*01:02"
## CELL_ID_122 "C*03:03" "A*02:01" "A*24:23" "B*15:01" "B*51:01" "C*03:03"
## CELL_ID_13  "C*07:01" "A*02:05" "A*32:01" "B*40:01" "B*49:01" "C*03:04"
## CELL_ID_16  "C*12:03" "A*24:02" "A*25:01" "B*18:01" "B*51:09" "C*01:02"
## CELL_ID_163 "C*07:02" "A*03:01" "A*32:01" "B*07:02" "B*44:17" "C*05:01"
## CELL_ID_165 "C*06:02" "A*02:01" "A*02:01" "B*35:31" "B*57:01" "C*02:02"
## CELL_ID_18  "C*06:02" "A*02:01" "A*32:01" "B*15:01" "B*27:08" "C*03:04"
## CELL_ID_21  "C*14:02" "A*02:01" "A*02:01" "B*44:02" "B*51:01" "C*07:04"
## CELL_ID_235 "C*16:04" "A*02:01" "A*31:01" "B*44:02" "B*51:01" "C*14:02"
## CELL_ID_36  "C*15:29" "A*24:10" "A*29:01" "B*07:05" "B*51:01" "C*07:04"
## CELL_ID_38  "C*07:02" "A*03:01" "A*11:01" "B*07:02" "B*51:01" "C*07:02"
## CELL_ID_39  "C*06:02" "A*01:02" "A*66:01" "B*58:01" "B*58:02" "C*03:02"
## CELL_ID_41  "C*07:02" "A*11:01" "A*26:01" "B*07:06" "B*55:02" "C*01:02"
## CELL_ID_45  "C*14:02" "A*02:01" "A*33:03" "B*15:16" "B*44:03" "C*04:01"
## CELL_ID_56  "C*04:01" "A*02:11" "A*68:01" "B*35:05" "B*40:04" "C*03:04"
## CELL_ID_6   "C*08:04" "A*03:01" "A*30:01" "B*15:10" "B*58:01" "C*07:01"
## CELL_ID_79  "C*12:02" "A*24:02" "A*24:02" "B*52:01" "B*52:01" "C*12:02"
## CELL_ID_94  "C*07:01" "A*01:01" "A*24:02" "B*08:01" "B*44:06" "C*05:01"
## CELL_ID_99  "C*06:02" "A*03:01" "A*03:01" "B*47:01" "B*47:01" "C*06:02"
##              V7
## CELL_ID_108 "C*07:04"
## CELL_ID_109 "C*03:04"
## CELL_ID_122 "C*03:04"
## CELL_ID_13  "C*07:01"
## CELL_ID_16  "C*12:03"
## CELL_ID_163 "C*07:02"
## CELL_ID_165 "C*06:02"
## CELL_ID_18  "C*06:02"
## CELL_ID_21  "C*14:02"
```

```
## CELL_ID_235 "C*16:04"
## CELL_ID_36  "C*15:05"
## CELL_ID_38  "C*15:06"
## CELL_ID_39  "C*06:02"
## CELL_ID_41  "C*07:02"
## CELL_ID_45  "C*14:02"
## CELL_ID_56  "C*04:01"
## CELL_ID_6   "C*08:04"
## CELL_ID_79  "C*12:02"
## CELL_ID_94  "C*07:01"
## CELL_ID_99  "C*06:02"
```

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 bwakit hlassign possible hits
## CELL_ID_108        0      0        0             0
## CELL_ID_109        5      4        4             5
## CELL_ID_122        5      4        5             5
## CELL_ID_13         6      4        6             6
## CELL_ID_16         1      1        1             5
## CELL_ID_163        1      2        2             2
## CELL_ID_165        1      1        1             2
## CELL_ID_18         6      6        6             6
## CELL_ID_21         0      0        0             0
## CELL_ID_235        6      6        6             6
## CELL_ID_36         6      5        6             6
## CELL_ID_38         0      0        0             6
## CELL_ID_39         5      5        6             6
## CELL_ID_41         6      4        5             6
## CELL_ID_45         6      6        6             6
## CELL_ID_56         6      6        6             6
## CELL_ID_6          5      5        5             6
## CELL_ID_79         2      2        2             2
## CELL_ID_94         1      1        1             6
## CELL_ID_99         6      6        6             6
```

```
colSums(accordance)
```

```
##      optitype        bwakit      hlassign possible hits
##            74            68            74            93
```

```
(colSums(accordance)/ colSums(accordance)[4])*100
```

```
##      optitype        bwakit      hlassign possible hits
##      79.56989      73.11828      79.56989     100.00000
```

3 samples can be identified with having wrong reference types assigned to them. To compute performance, we remove them and redo the calculations.

```
accordance = accordance[-c(5,19,12),]
colSums(accordance)
```

```
##      optitype        bwakit      hlassign possible hits
##            72            66            72            76
```

```
(colSums(accordance)/ colSums(accordance)[4])*100
```

```
##      optitype        bwakit      hlassign possible hits
##      94.73684      86.84211      94.73684     100.00000
```

Clearly, optitype and hlassign show better performance than bwakit for our sample set of 20 panels (HLA enriched).