

Compare Immuannot and *Skirt* in KIR allele typing in HPRC samples

Running time

We ran skirt and Immuannot on 95 HPRC assemblies (one reference genome CHM13.0). Each run goes with three threads. On average, Immuannot took 231.948 wall clock seconds, while skirt took 1413.9 wall clock seconds.

HPRC calls

All Immuannot's KIR call (n=901) can be found in Skirt's output, and Skirt has five extra calls missed by Immuannot:

```
HG002#2#JAHKSD010000133.1 701392..707153 KIR3DS1# --> low confidence
HG002#2#JAHKSD010000425.1 19868..25629 KIR3DS1# --> low confidence
HG00741#1#JAHALY010000117.1 374240..376067 KIR3DL2# --> low confidence
HG02630#2#JAHADP010000210.1 866052..878183 KIR2DS4*00101$ -> high confidence
NA19240#1#JAEOM010000197.1 440542..452673 KIR2DS4*00101$ -> high confidence
```

Example mappings are selected based on the positions (3rd column in the data above) and the mismatch number coded as 'NM:i:X' in the minimap2's output.

HG002#2#JAHKSD010000133.1 701392..707153 KIR3DS1#

Two example CDS split-alignments:

```
KIR:KIR00749      1165      0      950 +
HG002#2#JAHKSD010000133.1 701392 707153 950 950 NM:i:0
cg:Z:34M999N36M744N285M1488N300M1580N295M
cs:Z::34~gt999ag:36~gt744ag:285~gt1488ag:300~gt1580ag:295 KIR3DS1*082 1165 bp
```

```
KIR:KIR00871      1165      0      950 +
HG002#2#JAHKSD010000133.1 701392 707153 950 950 NM:i:0
cg:Z:34M999N36M744N285M1488N300M1580N295M
cs:Z::34~gt999ag:36~gt744ag:285~gt1488ag:300~gt1580ag:295 KIR3DS1*107 1165 bp
```

CDS mapping rates $950/1165 = 81\%$

HG002#2#JAHKSD010000425.1 19868..25629 KIR3DS1#

Two example CDS split-alignments:

```
KIR:KIR01654      1165      0      950  +
HG002#2#JAHKSD010000425.1      26976      19868      25629      950 950 0      NM:i:0
cg:Z:34M999N36M744N285M1488N300M1580N295M
cs:Z::34~gt999ag:36~gt744ag:285~gt1488ag:300~gt1580ag:295      KIR3DS1*0130126 1165 bp
```

```
KIR:KIR01676      1165      0      950  +
HG002#2#JAHKSD010000425.1      26976      19868      25629      950 950 0      NM:i:0
cg:Z:34M999N36M744N285M1488N300M1580N295M
cs:Z::34~gt999ag:36~gt744ag:285~gt1488ag:300~gt1580ag:295      KIR3DS1*1230101 1165 bp
```

CDS mapping rates $950/1165 = 81\%$

HG00741#1#JAHALY010000117.1 374240..376067 KIR3DL2#

Two example CDS split-alignments:

```
KIR:KIR00926      1368      0      356      +
HG00741#1#JAHALY010000117.1 374240 376067 353      356      60      NM:i:3
cg:Z:34M729N36M742N286M
cs:Z::25~ct:8~gt729ag:36~gt742ag:251*ag:14*cg:19      KIR3DL2*105 1368 bp
```

```
KIR:KIR00938      1368      0      356      +
HG00741#1#JAHALY010000117.1 374240 376067 354      356      60      NM:i:2
cg:Z:34M729N36M742N286M
cs:Z::25~ct:8~gt729ag:36~gt742ag:251*ag:34      KIR3DL2*10002 1368 bp
```

CDS mapping rates $356/1368 = 26\%$

HG02630#2#JAH AOP010000210.1 866052..878183 KIR2DS4*00101\$

KIR2DS4*0010101 gene length is 16092, but the call allele length is 12131, indicating large intron deletion(s)

```
KIR:KIR00045      915      0      915      -
HG02630#2#JAH AOP010000210.1 866052 878183 915      915      60      NM:i:0
cg:Z:42M98N53M462N105M1195N51M3168N294M1552N300M2461N36M2280N34M
cs:Z::42~ct98ac:53~ct462ac:105~ct1195ac:51~ct3168ac:294~ct1552
ac:300~ct2461ac:36~ct2280ac:34
KIR2DS4*0010101 915 bp
```

```
KIR:KIR00092      915      0      915      -
HG02630#2#JAH AOP010000210.1 866052 878183 915      915      60      NM:i:0
```

cg:Z:42M98N53M462N105M1195N51M3168N294M1552N300M2461N36M2280N34M
 cs:Z::42~ct98ac:53~ct462ac:105~ct1195ac:51~ct3168ac:294~ct1552
 ac:300~ct2461ac:36~ct2280ac:34
 KIR2DS4*0010102 915 bp

CDS mapping rates $915/915 = 100\%$

NA19240#1#JAHEOM010000197.1 440542..452673 KIR2DS4*00101\$

KIR2DS4*0010101 gene length is 16092, but the call allele length is 12131, indicating large intron deletion(s)

KIR:KIR00045 915 0 915 +
 NA19240#1#JAHEOM010000197.1 440542 452673 915 915 56 NM:i:0
 cg:Z:34M2280N36M2461N300M1552N294M3168N51M1195N105M462N53M98N42M
 cs:Z::34~gt2280ag:36~gt2461ag:300~gt1552ag:294~gt3168ag:51~gt1195
 ag:105~gt462ag:53~gt98ag:42
 KIR2DS4*0010101 915 bp

KIR:KIR00092 915 0 915 +
 NA19240#1#JAHEOM010000197.1 440542 452673 915 915 56 NM:i:0
 cg:Z:34M2280N36M2461N300M1552N294M3168N51M1195N105M462N53M98N42M
 cs:Z::34~gt2280ag:36~gt2461ag:300~gt1552ag:294~gt3168ag:51~gt1195
 ag:105~gt462ag:53~gt98ag:42
 KIR2DS4*0010102 915 bp

CDS mapping rates $915/915 = 100\%$

Three mismatches caused by gene fusion:

[#HG02723#1#JAHEOU010000160.1](#)

skirt: 7567681..7579926 KIR2DL3*00101_e1245-2DP1*00201_e678
 immuannot: 7568452..7579926 KIR2DP1*new

[#HG02630#2#JAHADP010000210.1](#)

skirt: 880672..896028 KIR2DS4*00101_e124567-3DL1*03501_e8
 immuannot: 880650..896295 KIR2DS4*new

[#NA19240#1#JAHEOM010000197.1](#)

skirt: 422697..438053 KIR2DS4*00101_e124567-3DL1*03501_e8
 immuannot: 422431..438394 KIR2DS4*new

Conclusion

1. Immuannot is four times faster than skirt on full genome assembly annotation
2. Match rates between Immuannot and skirt on HPRC samples $898/901 = 99.7\%$, three mismatches are caused by gene fusion that Immuannot can be distinguished in current version.
3. Skirt has five calls that missed by Immuannot, three are in low confidence because the CDS are partially mapped, two are in high confidence with fully CDS match. Immuannot missed them because of large size intron deletion(s).