HLA variant identification techniques in African Populations

Martin O. Pollard, S. Peacock, N. Park, C. Pomilla, M. Quail, D. Gurdasani, Manjinder S. Sandhu

Wellcome Trust Sanger Institute, University of Cambridge

Introduction

The Human Leukocyte Antigen (HLA) genes located on chromosome 6 are responsible for regulating immune function via antigen presentation and are one of the determining factors for stem cell and organ transplantation compatibility. Implicated by numerous association studies in autoimmune disorders, cancer, vaccine response and both non-infectious and infectious disease risk; there is a strong motivation to study this region. However this region is hypervariable; has a complex LD structure and lacks good population references for the African region which we are studying. This pilot study will aid us selecting methods to type the HLA for the Genome Diversity in Africa Project allowing us to build a population reference and mitigate these issues.

Samples and Nomenclature

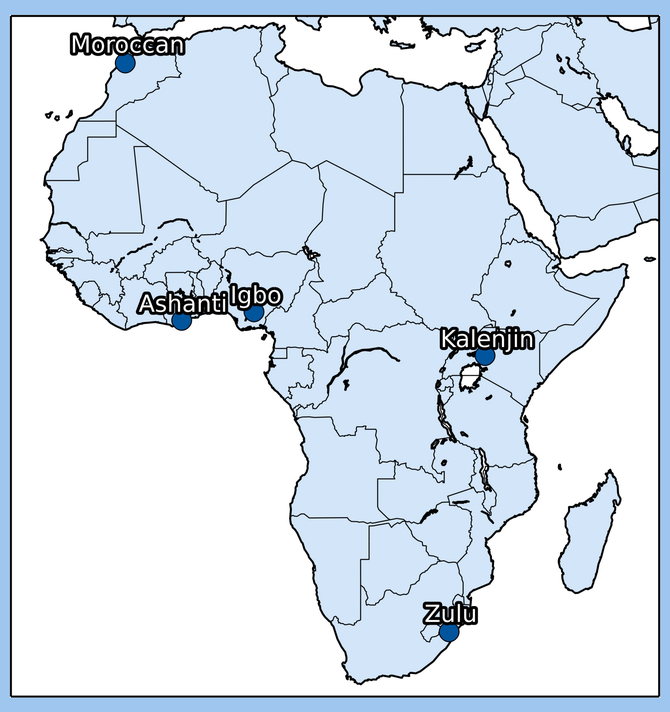
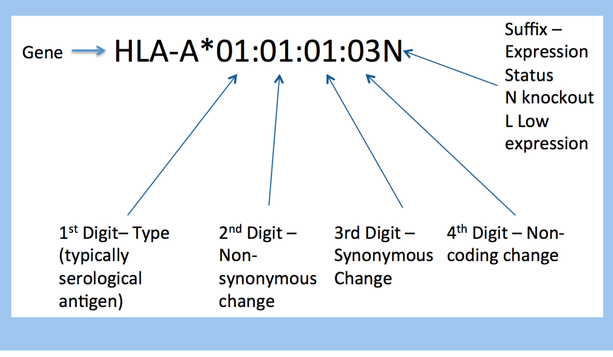


Figure 1: 125 samples were Sanger Sequenced, 25 from each population. A subset of 4 from Igbo, Kalenjin, Ashanti and Zulu were sequenced on PacBio as a pilot.



Sanger Method

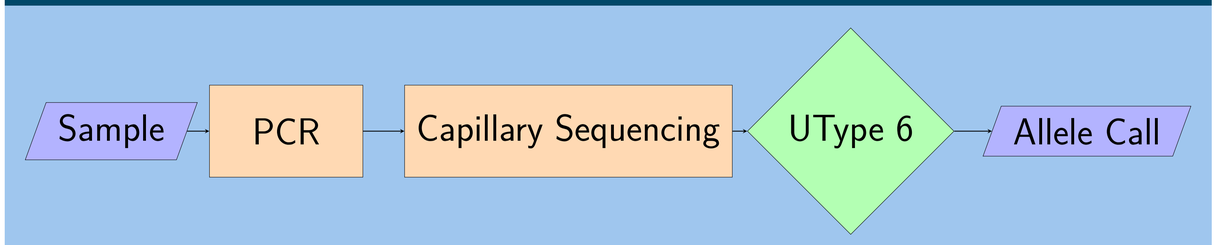


Figure 2: All samples were Sanger typed for HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1, -DRB3/4/5.

PacBio Method

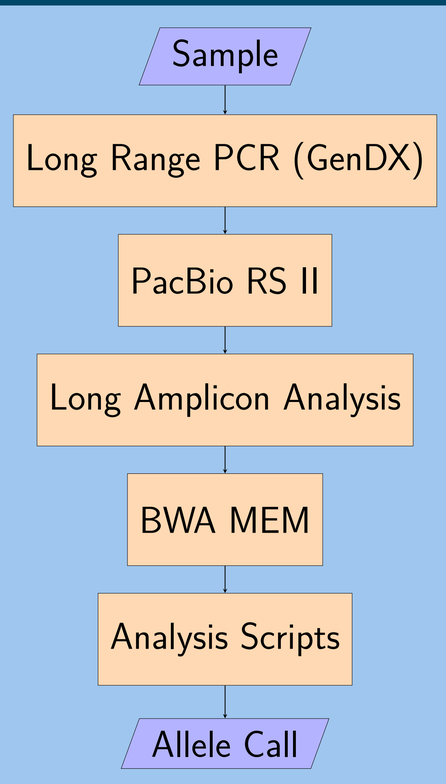


Figure 3: Four samples were typed on the PacBio machine, using GenDX primers.

Results

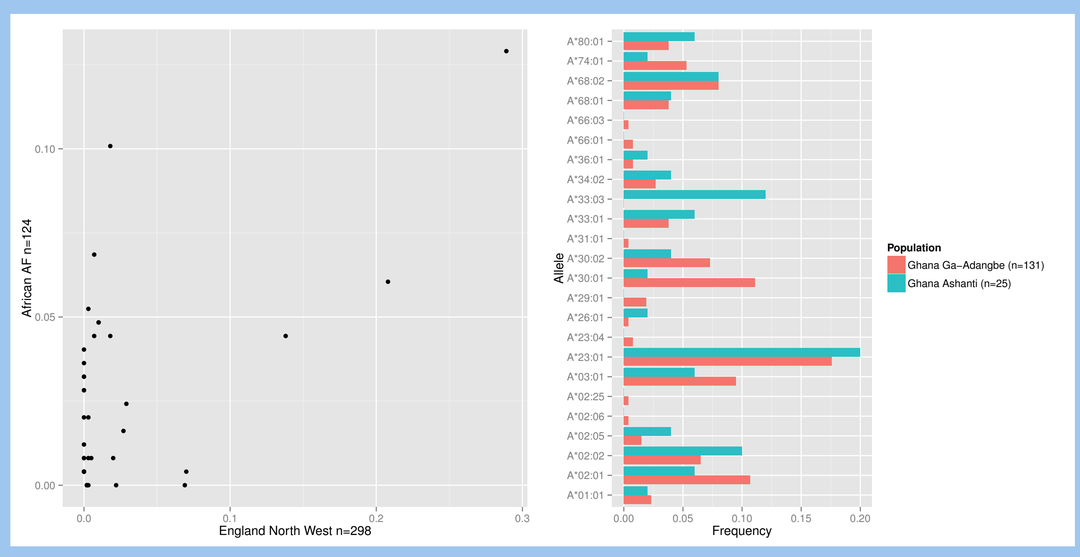


Figure 4: Left is concordance of our African AF to [1] for HLA-A. Right is comparison of our Sanger result to [2] for HLA-A.

The results from our Sanger sequencing pilot were promising with our small sample detecting alleles not found in nearby populations as shown in figure 4, we also found evidence for at least 2 novel alleles. The samples sequenced were not clonally separated so we could not ascertain phase without reference to the IMGT database, introducing some reference bias. Additionally as is standard for clinical typing we only sequenced exons 2, 3 & 4 in the Class I genes and exons 2 & 3 for most of the Class II genes apart from DPB1 & DRB3/4/5 where we sequenced 2, 3 & 4.

Fields	A	B	C	DPA1	DPB1	DQA1	DQB1	DRB1	DRB2	DRB3	DRB4	DRB5
2	**2,271**	**2,856**	**1,919**	**16**	**442**	**22**	**517**	**1,202**	**1**	**45**	**9**	**19**
3	**872**	**1,068**	**777**	**19**	**104**	**17**	**277**	**544**	**0**	**10**	**3**	**2**
4	**39**	**34**	**39**	**5**	**4**	**15**	**12**	**10**	**0**	**4**	**4**	**0**

Table 1: *Number of fields for each allele*

To aid us in devising mapping strategies we surveyed the IMGT/HLA database (table 1). This showed that the number of alleles where there is enough depth of sequencing to reveal intronic variation is quite small. For example HLA-B has 2526 alleles sequenced for exons 2 and 3, 603 for 2-4 and only 182 full sequences.

Sample	A		B		C	
Igbo	Known A*02:01:01:01	Known A*23:01:01	Known B*35:01:01:02	Known B*53:01:01	Known C*04:01:01:01	Known C*16:01:01
Ashanti	Known A*02:05:01	Novel A*80:01:01:01	Novel B*07:02:01	Known B*53:01:01	Known C*04:01:01:01	Known C*07:02:01:01
Kalenjin	Known A*02:01:01:01	Novel A*30:02:01:02	Novel B*50:01:01	Known B*51:01:01	Known C*02:02:02:01	Known C*06:02:01:02
Zulu	Novel A*03:01:01:01	Known A*24:02:01:01	Known B*07:02:01	Novel B*58:01:01:01	Known C*06:02:01:01	Known C*07:02:01:01

Table 2: Results of the Pacbio Sequencing with nearest allele detected.

The PacBio sequencing covered all exons and introns in the class I genes and detected 6 novel alleles as shown in table 2. Three of these contained intronic variation producing new 4 figure types but at least three affected Exon 3 producing a new 2 figure type.

Conclusion

This pilot has demonstrated that PacBio Sequencing allows you to discover hidden HLA diversity in a relatively small number of African samples. Our next step will be to explore a larger population of samples from the Genome Diversity in Africa project.

References

[1] Ana Alfirevic, Faviel Gonzalez-Galarza, Catherine Bell, Klara Martinsson, Vivien Platt, Giovanna Bretland, Jane Evely, Maike Lichtenfels, Karin Cederbrant, Neil French, Dean Naisbitt, B Kevin Park, Andrew Jones, and Munir Pirmohamed.
In silico analysis of hla associations with drug-induced liver injury: use of a hla-genotyped dna archive from healthy volunteers.
Genome Medicine, 4(6):51, 2012.

[2] Paul J. Norman, Jill A. Hollenbach, Neda Nemat-Gorgani, Lisbeth A. Guethlein, Hugo G. Hilton, Marcelo J. Pando, Kwadwo A. Koram, Eleanor M. Riley, Laurent Abi-Rached, and Peter Parham.
Co-evolution of human leukocyte antigen (hla) class i ligands with killer-cell immunoglobulin-like receptors (kir) in a genetically diverse population of sub-saharan africans.
PLoS Genet, 9(10):e1003938, 10 2013.

Acknowledgements

Zulu samples - Fraser Pirie and Ayesha Motala Dep Diabetes and Endocrinology, School of Clinical Medicine, University of Kwazulu-Natal, Durban, South Africa
Igbo & Kalenjin - Charles Rotimi, National Human Genome Research Institute, National Institute of Health (NIH), Bethesda, USA
Moroccans - Charles Agyemang, Dep of Public Health, Academisch Medisch Centrum (AMC), Amsterdam, The Netherlands
Ashanti - Richard Cooper, Dep of Public Health Sciences, Medical School, Loyola University, Chicago, USA