

Focus meeting: Best practice using GRCh38

Tuesday, March 31 2015, 13.00-15.30, SURF, Utrecht

Around 30 bioinformaticians working with Next Generation Sequencing data gathered on March 31, 2015 to discuss their work related to using the new human reference genome build GRCh38.

Ies Nijman from UMC Utrecht opened the meeting with an introduction on GRCh38 and the current solutions and improvement on mapping and variant calling using this new assembly. Pieter Neerincx from UMC Groningen explained to us the development and the methodology of lift-over tools and highlighted challenges remained for lift over from GRCh38. Pjotr Prins from UMC Utrecht further dived into the development of alignment program to support the new assembly, e.g., bwa-mem. One major change in GRCh38 is the inclusion of large number of alternative reference loci. As pointed out by Pjotr, un-careful dealing with these contigs will lead to low map-ability and wrong variant-calling.

In the second half of this meeting, Zuotian Tatum and Wibowo Arindrarto both from LUMC presented two biological studies using GRCh38 as reference. For a RNA-seq study, Zuotian reported a high correlation on gene expression level between using hg19 and GRCh38. However, there is still a lot need to be done for gene annotation on GRCh38. Wibowo reported a study on a difficult gene region using both hg19 and GRCh38. Clear difference can be seen but as Wibowo pointed out that the new assembly does not necessarily produce better results. Hence, it is crucial to discuss with biologist who has extensive knowledge in the gene and how it functions to make the correct selection and conclusion.

Concluding, this was a very productive meeting and all participants extended their understanding on the new GRCh38. Major take home messages are that (1) we observed relatively slow pickup of this new assembly in community (e.g., BROAD GATK), (2) there are quite some positive attitude towards GRCh38 as we see many improvements. However, tough regions still though, complex gene families might still not be accurately assembled. (3) there is a lacking of proper support for the ALT contigs in mapping and variant calling. Intermediate conclusion is to drop these in using GRCh38 until this support is there, or a completely new reference approach.

Participants:

Ies Nijman (UMCU), Wibowo Arindrarto (LUMC), Pieter Neerincx (UMCG), Wai Yi Leung (LUMC), Bart de Koning (MUMC), Saskia Hiltemann (ErasmusMC), Maarten Kooyman (SURFsara), Yuri Hoogstrate (ErasmusMC), Jochem, Zuotian Tatum (LUMC), Sander Bervoets (ServiceXS), Daniel Borràs (ServiceXS), Stef van Lieshout (VUMC), Martijn Vermaat (LUMC), Peter van 't Hof (LUMC), Robert Ernst (UMCU), Rianne Jansen (The Hyve), Peter Kok (The Hyve), Stephan Nouwens (ErasmusMC), Alexander Schoenhuth (CWI) Robert Kraaij (ErasmusMC), Johan den Dunnen (LUMC), Job van Riet (ErasmusMC), Bert Eussen (ErasmusMC), Tom Brands (ErasmusMC), Pjotr Prins, Walter de Valk (ErasmusMC), Thomas Chin-A-Woeng (ServiceXS), Jeroen Laros (LUMC), Rutger Brouwer (ErasmusMC), Harmen van de Werken (ErasmusMC), Theo de Vet (ErasmusMC), Sander Bollen (LUMC), Lars Eijssen (Maastricht University), Tim Dorscheidt (The Hyve)