Using NextFlow in healthcare: Non-invasive prenatal diagnosis

Authors: Helen Curley, Joe Larkman, Natalie Barrows, Chipo Mashayamombe-Wolfgarten.

Background:

Non-Invasive Prenatal Testing (NIPD) is a safer alternative to traditional invasive prenatal screening. NIPD analyses cell-free foetal DNA present in the maternal bloodstream to detect certain genetic conditions during pregnancy. NIPD boasts reduced risk of miscarriage, minimal discomfort to the pregnant mother, and early detection of genetic abnormalities.

We couple NIPD with Relative haplotype dosage (RHDO), which works by analysing the haplotypes inherited from both parents and comparing them with the foetal DNA. Haplotypes are specific combinations of genetic markers that are inherited together on the same chromosome from one parent.

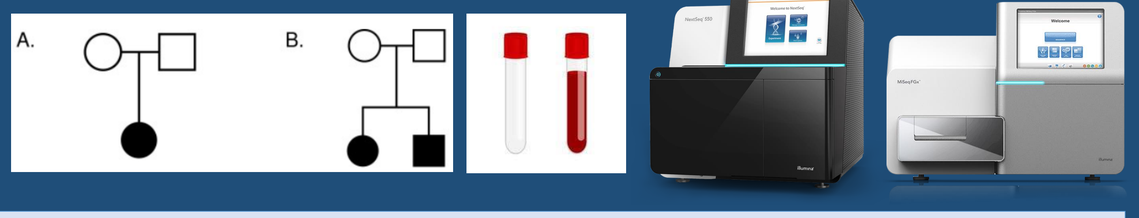
We use the NIPD-RHDO pipeline to investigate diseases such as Duchene muscular dystrophy (DMD), spinal muscular atrophy (SMA), congenital adrenal hyperplasia (CAH), cystic fibrosis (CF) and retinoblastoma (RB1)[1,2].

Samples and sequencing:

Two vials of 10ml blood are taken from the expectant mother along with samples from controls which must include paternal blood and can include siblings.

50µl of cfDNA is extracted from the maternal plasma (20-100ng DNA), for all future purposes this is considered the proband. Whilst 35µl of genomic DNA is taken from leukocytes (100ng DNA).

Using the Roche's KAPA HyperCapture kit, a targeted enrichment library is created and sequenced on either MiSeq or NextSeq sequencers.



Overview:

We used Nextflow v22.10.4 and Python v3.11.4. Different config files for genome build and sequencer type are provided. BCL Convert (v4.0.3) is utilised to create FASTQs. Probeset CF_CAH_RB1 must use a no decoy reference genome in Bowtie 2 (v2.4.4), which is not available via iGenomes. Samblaster (v0.1.26), samtools (v1.16.1) and Picard Tools (v3.0.0) are used to remove duplications and collect quality metrics. Platypus (v0.8.1) is used as a genotyper rather than variant caller by providing coordinates to "call". Bedtools intersect (v2.30.0) uses conditional branching for each probeset and takes in bed files for each disease separately. All custom made modules call custom python scripts. Sequential probability ratio testing (SPRT) analysis is used to assign likelihood of disease inheritance and is currently in Python v3.6.8.

Discussion:

Our genomics laboratory hub is responsible for the diagnosis, prognosis and treatment stratification for a population of 18 million. We have observed a continuous rise in demand for computing power and storage resources. To address this growing need, we are now exploring cloud computing solutions. In this regard, Nextflow emerges as a logical and promising choice for optimising our processes and enhancing the efficiency of our genomic analyses. By leveraging cloud-based resources and the flexibility of Nextflow, we aim to streamline our workflows and meet the escalating demands of our genomic testing services. The pipeline presented here exemplifies one of our successful transitions to Nextflow, wherein we amalgamated standard nf-core modules, and nf-core modules with customised components tailored to our specific requirements.

References:

1. Parks, M., Court, S., Bowns, B. *et al.* Non-invasive prenatal diagnosis of spinal muscular atrophy by relative haplotype dosage. *Eur J Hum Genet* **25**, 416–422 (2017). 2. Scotchman, E., Shaw, H., Patemoster B. *et al.* Non-invasive prenatal diagnosis and screening for monogenic disorders. *Eur J Obs & Gyne* **253**, 320-327 (2020). 3. Ewels, P., Peltzer, A., Fillinger, S. *et al.* The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnology **38**, pages276–278 (2020).