Results

Story: (1) first to phase *de novo* variants with PacBio sequencing, (2) replicated previous *de novo* SNV results, (3) demonstrated that the class of *de novo* indels most associated with DNA replication (CCCs) are mostly explained by paternal age, consistent with spermatogenesis

Fraction of *de novo* variants that were phased using Illumina reads: 0.4

Standard deviation (across probands): 0.06190185592012657

Fraction of *de novo* variants that were phased using Illumina reads: 0.8322324966974901

Standard deviation (across probands): 0.03070946196107359

When classifying SNPs, results were consistent with Iceland paper

* correlation between paternal age and number of de novos
* C > T more maternal than paternal
* C > A more paternal than maternal
* Paternal age effect more consistent across mutational classes than maternal age effect

CCC indels were positively correlated with paternal age and not correlated with maternal age – confirms the idea that CCC mutations are related to polymerase slippage during replication

|  |  |  |  |
| --- | --- | --- | --- |
| *Position* | *Indel* | *Context* | *Type* |
| chr8:117967436 | +T | TTAAATATTTTTTT | HR |
| chr5:52931910 | -A | CCAATTAAAAAAA | HR |
| chr12:71038252 | -GTG | taacttgtggtgttt | CCC |
| chr2:158890470 | +GAG | CAGAACTGAGGAGCAT | CCC |
| chr20:39588834 | -AGG | GAGAGAAGGAGATGT | non-CCC |
| chr4:14567444 | +AACCC | ACATAATATATAACCCAACCTACCTT | non-CCC |

Note: only one of the non-CCC indels was an insertion, rest were deletions. See if this is a pattern with illumina data as well as Netherlands/Oscar’s data