Problems:

* Predicting local indel mutation rates (at the genic or region level) would serve as a powerful control for disease association studies as well as an invaluable tool for understanding evolution, especially in repetitive regions.
  + Here we extend previous work to predicting indels
* Biology: Three quarters of de novo indels arise through polymerase slippage. Polymerase slippage implies a replication-associated event, yet previous studies failed to identify an association between de novo indels and parental age. This is attributable to either a lack of power (cite 1000G) or not applying previously defined indel classification techniques (cite Iceland)
* Technology: The hypothesized benefits of phasing DNVs with long-read sequencing have not been tested.

*De novo* variants

* New variants in the offspring that are not in the parents
* Fundamental to evolution and important for disease pathogenesis
* Understanding how DNVs arise has impacts on evolution, reproductive decision-making, knowledge of disease pathogenesis, and predictions mutagenic somatic events (i.e., not germline). Models of DNVs provide
* One method for understanding DNVs is assigning them to a parent of origin. This improves our understanding because then the influence of the biology of sperm and oocytes on DNV rates becomes a testable hypothesis.
* Both spermatogonia and oocytes undergo approximately ~20 cell divisions as primordial germ cells before puberty
* After puberty sperm continuously undergoes mitosis, while oocytes are stable cells. Furthermore, both occupy different environments. Teasing apart the mutagenic properties of stable vs labile cells has important implications for non-germline cell types.
* Review introductions/conclusions of recent *de novo* variant papers!

*De novo* variant phasing

* In order to determine if a de novo variant arose in sperm or oocyte, they have to be phased (i.e., assigned) to the parent of origin
* Two methods for phasing DNVs
* Three-generation haplotype phasing
  + Phases x% of DNVs
  + Requires sequencing of offspring of the person to whom the DNV is assigned, which is not possible for anyone who has not had children and difficult to collect larger families.
* Read-based phasing
  + identifying heterozygous variants on the same read that are uniquely inherited from mom or dad
  + Phases x% of DNVs with Illumina
  + Limitation: read length
  + Here we use long-read technology to phase DNVs that were original identified with short-read sequencing

Patterns of *de novo* SNVs and indels

* What is known about germline *de novo* SNVs?
  + Review results
* What is known about germline *de novo* indels?

Here, we perform the first long-read phasing of *de novo* variants, replicate previous findings with *de novo* SNVs, and define the mutagenic properties of *de novo* insertions and deletions (indels).