Ultra-Sensitive Targeted Haplotype Phasing Using TELL-Seq

Genomics

Veronika Mikhaylova (vmikhaylova@universalsequencing.com), Universal Sequencing Technology, **Long Pham**, Universal Sequencing Technology, **Peter Chang**, Universal Sequencing Technology, **Yu Xia**, Universal Sequencing Technology, **Christian Boles**, Sage Sciences, **Andrew Anfora**, Universal Sequencing Technology, **Ivan Garcia-Bassets**, Universal Sequencing Technology, **Yong Wang**, Universal Sequencing Technology, **Tom Chen**, Universal Sequencing Technology Corp

Humans have maternal and paternal sets of matched chromosomal copies. Combined, these two sets typically differ from the reference human genome at 4 to 5 million sites in every individual, although many of these differences are present only in one of the two copies (heterozygous). Resolving whether neighboring differences belong to the same or different copies (a process known as 'phasing') is critical for many research and clinical applications. Highly accurate and low-cost conventional short-read next-generation sequencing (NGS) methods fail to phase heterozygous sites located more than a few hundred bases apart. Phasing must then rely on population allelic frequencies, parental genotypes, or long-read NGS methods with limited sequencing sensitivity or fidelity. Alternative to conventional short-read NGS, linked-read approaches capture long-range information during NGS library preparation. Using linked reads, transposase enzyme-linked long-read sequencing (TELL-Seq) can phase an entire genome. Here, we sought to determine whether TELL-Seq can also phase discrete DNA fragments in targeted experiments. First, we show the phasing of two double somatic mutations in a 3.4kb fragment of amplified PIK3CA cDNA, relevant for cancer prognosis and treatment. Second, we show the phasing of seven polymorphic sites in a 13kb fragment of amplified genomic DNA containing part of the SCN10A locus, relevant for cardiac arrest risk assessment. Finally, we show the phasing of the BRCA1 and BRCA2 loci in the Ashkenazi Trio, two genes associated with breast cancer. We used CRISPR-Cas9 and pulse-field electrophoresis to excise and isolate these two almost 200kb fragments. We developed a computational pipeline that outputs phased heterozygous sites aligned to the reference genome to aid these three phasing analyses. Together, our results demonstrate high phasing accuracy combining conventional NGS data collection with a linked-read library method using DNA fragments of varying sizes.