Submitter Name: Alexander E Urban
Submitter email: aeurban@stanford.edu

Technically challenging types of genome sequence variants with relevance to brain function, and new approaches to their analysis

Alexander E Urban[1,2]

[1]Department of Psychiatry and Behavioral Sciences, Stanford University; [2]Department of Genetics, Stanford University

State-of-the-art whole-genome sequencing (WGS) cannot fully resolve several potentially important types of genome sequence variants. Advanced computational approaches, alone or in combination with advanced sample preparation and the emerging long-read sequencing technologies, can be employed to address this challenge, and two examples for such approaches will be given here.

Retrotransposons such as LINE-1 or *Alu* elements are present in large numbers in the human genome, and their activity can cause somatic genome variation in the human nervous system, which is hypothesized to have relevance to brain development and neuropsychiatric disease. However, the detection of individual somatic mobile element insertions presents a difficult signal-to-noise problem. We developed a machine-learning method (RetroSom) to detect somatic retrotransposition events in very deep WGS data. We found that there was anatomical distribution of the LINE-1 insertions in neurons and glia across both hemispheres, indicating retrotransposition occurred during early embryogenesis, and with the potential to affect gene expression.

Large copy number variants (CNVs) have been found to be strongly associated with disorders such as schizophrenia or autism. Their presence in the germline genome is relatively easily detected with standard approaches. However, about 5% of the human genome is constituted by segmental duplications (SegDups), long stretches of SegDups very often mask the boundaries of large CNVs, and current methods cannot determine the exact breakpoints of such large CNVs. We have developed CRISPR-targeted ultra-long read sequencing (CTLR-Seq), which allowed us to assemble and resolve the SegDup-rich breakpoint regions of the psychiatrically relevant large CNVs in 22q11 and 16p11.