Московский семинар по биоинформатике

Заседание 324

5 сентября 2018, среда, 19.00

VLADIMIR B. SEPLYARSKIY

Harvard Medical School

DECIPHERING SIGNATURES OF MUTATIONAL PROCESSES IN HUMAN GERMLINE

Stereotypic mutational processes operating in human germline are the source of genetic diversity and the cause of hereditary diseases. Patterns of germline mutations vary on different scales including existence of single nucleotide hotspots, clusters of multiple mutations at scales of ten kilobases and megabase-scale variation. Observed variation is a consequence of exposures to a combination of unknown mutational processes. However, etiology, intensities and spectrum of mutational processes in germline are almost unexplored.

Current understanding of mutational processes in human cells is primarily based on the cancer data. Inference of mutational signatures extracted from cancers relies on the fact that individual tumors have dramatic variation of exposure to a mutational process

Here, we employed variability of mutational patterns along the genome to infer underlying mutational processes. We assume that heterogeneity of mutational spectra between loci is driven by different relative contribution of stereotypic mutational processes.

To formally extract mutational signatures, the genome was binned in non-overlapping windows of fixed size (e.g. 20 or 100 kilobases) and spectra were compared between windows. Very rare polymorphisms (allele frequency below 10-4) from GNomad and TOPMEDs projects served as proxy of mutations. Inference of signatures was formulated as a matrix decomposition problem, where a matrix of mutation type rates (rows) in windows (columns) was approximated by product of exposure matrix (per-window intensities of unknown processes) and spectra matrix (per-process spectra of mutation rates). Using independent component analysis to perform matrix decomposition we discovered seven major mutational processes including signatures created by transcription coupled nucleotide excision repair, error-prone asymmetric bypass of DNA damages during replication, signature associated with replication timing, signature associated with repeat expansion and oocyte-specific signature. These signatures were consistent with patterns of de novo germline mutations excluding any non-mutational sources of their origin. Oocyte-specific signature is localized in regions with disproportionally high fractions of mutations of maternal origin, which were recently discovered in studies of de novo mutations in human trios. This signature is active in several genomic regions comprising about 5% of the genome; has a representative scale of about 20 megabases, but shows the highest intensity on non-transcribed strand of long genes (WWOX1, CSMD1, RBFOX1). Also, we show that replication-associated signatures predict fork polarity and inter-origin distance in germline opening up an avenue of mutation-based inference of molecular features in human cells.

Рабочий язык семинара – русский

Заседание семинара пройдет в ФББ МГУ, Воробьевы Горы 1, к. 73, комната 221.

Сайт семинара (проезд, архив и т.д.): http://www.rtcb.iitp.ru/msb/Index.htm