Computational studies of mutational sequence signatures in cancer genomes

Akademisk avhandling

Som för avläggande av medicine doktorsexamen vid Sahlgrenska akademin, Göteborgs universitet kommer att offentligen försvaras i **Arvid Carlsson**, Akademikum, Medicinaregatan 3, Göteborg

Torsdagen den **4:e november 2021**, klockan **09:00**

av **Markus Lindberg**
Fakultetsopponent:
Dr. Inigo Martincorena
Wellcome Sanger Institute, Cambridge, UK

Avhandlingen baseras på följande delarbeten

I.   Intragenomic variability and extended sequence patterns in the mutational signature of ultraviolet light.
     Lindberg M., Boström M., Elliott K., Larsson E.
     Proc Natl Acad Sci U S A. 2019 Oct 8;116(41):20411-20417

II. An extended mutational signature model reveals contextual motifs informative of local mutation rate in cancer.
     Lindberg M., Larsson E.
     Manuscript

III. Independent somatic evolution underlies clustered neuroendocrine tumors in the human small intestine.
     Elias E., Ardalan A., **Lindberg M.**, Reinsbach S., Muth A., Nilsson O., Arvidsson A., Larsson E.
     Accepted for publication in Nature Communications

SAHLGRENSKA AKADEMIN
INSTITUTIONEN FÖR BIOMEDICIN

GÖTEBORGS
UNIVERSITET

Computational studies of mutational sequence signatures in cancer genomes

Markus Lindberg

Avd för medicinsk kemi och cellbiologi, institutionen för biomedicin, Sahlgrenska akademin, Göteborgs universitet, Sverige, 2021.

Abstract

Cancer typically forms when mutational processes modify key cancer driver genes, resulting in positive selection and tumor growth. As such, mutational processes are at the core of the disease. Trinucleotide mutational signatures have emerged in the last decade as essential tools for analysis of mutational processes. These models describe the relative probability of mutagenesis at different trinucleotide contexts for a variety of mutational processes.

In UV exposed cancers, the DNA sequence "TTCCG" constitutes UV mutation hotspots in active promoters, which cannot be represented by trinucleotide-based mutational signatures. In the first study, we expand the mutational signature of UV and demonstrate that its trinucleotide profile depends on cytosine methylation, and that this stems from increased CPD formation at methylated sites. We also show that incorporation of longer sequence patterns into the signature model better describes the UV mutational process. Furthermore, we show that such extended signature models increase accuracy when separating driver mutations from passengers. Strongest effect on mutation probability was from TTCCG in expressed promoters, but other sequence patterns also significantly modulate mutation probability in UV exposed melanoma.

In the second study, we build on this concept further and develop a bioinformatics tools capable of estimating longer sequence patterns' effect on mutation rates in conjunction with trinucleotide contexts. We then applied our tool on 27 cancer types to explore sequence patterns with modulating effects on mutation frequencies. Homopolymer patterns were found to be the strongest effectors, but pentamers of higher complexities were also found to increase mutation rates among multiple cancers.

Finally, in the last study we analyze the mutational properties of small intestine neuroendocrine tumors (SI-NETs). Mutational signature analysis reveals a lack of mutagenic processes active in these tumors, which goes together with its notably low mutation burden and low frequency of driver mutations. The most striking result in this study is that despite the multifocal nature (multiple tumors in close proximity) of this cancer type, each tumor has evolved independently.

In summary, this thesis demonstrates the utility of mutational signatures, and highlight novel approaches to signature analysis that incorporate longer sequence patterns.

Keywords: Cancer, genomics, mutational signatures, ultraviolet light, multifocal tumors