ITCR Monthly Meeting

October 4, 2019
Eliezer Van Allen
Felix Dietlein
Dana-Farber Cancer Institute, Boston
Passenger and driver mutations in cancer

Stratton et al. Nature 2009
Searching for a needle in the haystack
The background mutation rate varies across the human genome

Lawrence et al. Nature 2013
synonymous mutations provide a way to account for the variation of the regional background mutation rate in the coding space
Nucleotide contexts provide an additional signal in favor of driver mutations.

- **Passenger mutations**: Random mutations that do not contribute to cancer.
- **Driver mutations**: Mutations that confer a selective advantage and are associated with cancer.

Detection of cancer genes based on mutation counts and nucleotide context:
- **Strong relative increase of mutation count**
  - Non-cancer gene
  - Cancer gene
- **Low relative increase**
  - Non-cancer gene
  - Cancer gene

**Functional consequence**:
- Low background mutation rate
  - Search for genes with increased number of mutations above the local background mutation rate.
- High background mutation rate
  - Search for genes with increased number of mutations in an unusual context.

**Accumulation in functional positions**
- Driver mutations accumulate in specific functional positions.

**Distribution dependent on nucleotide context**
- Nucleotide contexts provide additional signal.

Dietlein et al. submitted
### Mutational signatures in cancer

**Signature 1A**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 1B**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 2**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 3**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 4**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 5**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 6**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 7**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 8**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 9**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 10**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 11**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 12**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 13**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 14**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 15**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 16**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 17**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 18**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 19**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 20**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

**Signature 21**

- C>A
- C>G
- C>T
- T>A
- T>C
- T>G

---

*Figure adapted from Alexandrov et al. Nature 2013*
nucleotide contexts may provide a way to identify driver mutations in the non-coding space
Nucleotide contexts provide an additional signal in favor of driver mutations.
Nucleotide contexts provide an additional signal in favor of driver mutations

Dietlein et al. submitted
A catalog of driver genes in human cancer
**Aim 1** To define the landscape of noncoding cancer driver mutations using nucleotide context

**Rationale:** The role of noncoding mutations in tumor formation largely remains unknown. The emerging availability of large whole-genome sequencing datasets creates an ample opportunity to develop new mutational significance algorithms.

**Hypothesis:** Current mutational significance algorithms have been developed for coding mutations. Applying them to noncoding regions does not exploit the full potential of whole-genome sequencing datasets.

**Approach:** Develop a new statistical model that accounts for:
- Context-dependent distribution of passenger mutations
- Modeling of the background mutation rate
- Partitioning of the background mutation rate:
- Modeling the sequence composition of the reference genome
- Coverage fluctuation
Extended nucleotide contexts around passengers contain a signal.
Extended nucleotide contexts around passengers contain a signal.

- **Fig. S2**
  - Passenger mutations are enriched in broad and characteristic nucleotide contexts.
  - We observed that passenger mutations were surrounded by characteristic nucleotide contexts. We visualized these characteristic contexts by sequence logo plots.
  - **Left:** For each mutational process, mutation frequencies of the reference (ref) nucleotides (C vs. T) and mutation types (mut: transitions (C/T), type-I transitions (A), type-II transitions (G)) are visualized as logo plots.
  - **Right:** Sequence logo plots represent the relative over-representation (deviation) of the nucleotides around mutations relative to their expected frequencies in the human exome. In other words, the height of each flanking nucleotide indicates its impact on the local mutation probability. Most characteristic nucleotide contexts clearly exceed the trinucleotide context (±1) around mutations. For instance, the nucleotide context associated with UV exposure (motif 7) contains almost exclusively C>T mutations; T's are overrepresented in 5' adjacency and C/G followed by T's are overrepresented in 3' adjacency. Hence, sequence logo plots provide a convenient way to visualize the dependence of local mutation probabilities on the broad nucleotide context.

Dietlein et al. submitted
A composite likelihood model to account for extended nucleotide contexts

Dietlein et al. submitted
Aim 2  To evaluate the clinical significance of passenger mutation distribution patterns for immunotherapies

**Rationale:** Despite the clinical relevance of immunotherapies, robust markers for patient stratification are currently lacking. The most common parameter to predict clinical response is the mutational burden.

**Hypothesis:** Recent studies have suggested that not only the total number of passenger mutations but also their underlying biology mediates clinical response to immunotherapy. We and others have shown that the underlying biology of passenger mutations is not only reflected by their surrounding trinucleotide context but that a relevant signal is in the extended nucleotide context.
Aim 2  To evaluate the clinical significance of passenger mutation distribution patterns for immunotherapies

Approach: Use extended nucleotide contexts for “improved” mutational signatures

• Apply our composite likelihood model to a sequencing dataset of 249 tumors with clinically annotated outcomes to immune checkpoint therapy
• Evaluate the immunogenicity of each individual passenger mutation by using established algorithms, such as POLYSOLVER algorithm and NetMHCpan-3.0
• Determine which passenger mutations from the whole-exome sequencing data can be rediscovered in the corresponding RNA sequencing data of the same tumor.
• Develop an integrative model to combine these aspects and predict the immunotherapy response of a tumor, based on the immunogenicity of its passenger mutations