Oriol Gómez Lores
1528867
Temes Actual Bioinformàtica

BIOINFORMATIC METHODS TO IDENTIFY CANCER DRIVER GENES

1 INTRODUCTION

Cancer is a group of illnesses which are characterized by high cell proliferation and a chance to invade other tissues. It is a genetic illness, and most cancers are due to somatic mutations. These mutations affect different pathways which activate convenient genes or inactivate those which would hinder cancer development.

2

Cancer is due to mutations which are slowly selected according to the 'fitness' they 'give' to the cancer within the growing environment. Driver genes are genes which, when mutated, give a selective adaptation and start the cancer. On average, cancers have a lot of mutations, but just 4.6 are driver mutations.

3 HOW CAN WE IDENTIFY DRIVER MUTATIONS? 4

Countless patients with the illness are analysed, and the mutations which meet the following requirements are possible candidates:

1. **Recurrence.**
2. **Functional impact.**
3. **Clustering.**

These requirements are signals of positive selection and are complementary. Mutations can be found comparing observed data vs expected data. These mutations are localized on genome regions, depending on chromatin state. Furthermore, some mutations are more frequent than others (C>T due to UV light), and it is those infrequent ones which are looked into, as they may be candidates for drive mutations, because we do not expect to naturally see them.

1. ONCODRIVECLUSTL

1. We see mutations.
2. We distribute these mutations randomly and observe a statistical distribution of the mutations.
3. We get the chance of having a cluster within such context, and compare with the patient, to see whether the cluster is likely or not.
4. Can be combined with 3D-protein structure software.

2. INTOGEN

1. Combines various software in order to identify all driver mutations found under combined criteria.
2. Combines 7 different methods, weighting them differently according to how well they perform.
3. Only identifies genes encoding proteins, what about non-coding genes?

3. boostDM

1. Predicts driver mutations.
2. Builds a model with mutations we know that are drivers and those that are passenger.
3. Compares this information with mutations seen by IntoGen and tries to explain why hypothetical mutations could potentially be driver mutations.

4. CANCER GENOME INTERPRETER

1. Compiles information from many databases and many algorithms.
2. We can introduce a gene and a mutation, and a cancer type.
3. Yields the information available per mutation, and which treatments are available (Prescription).