Visualization Challenges of Variant Interpretation in Multiscale NGS Data

E. Ståhlbom, J. Molin, C. Lundström and A. Ynnerman

Design study

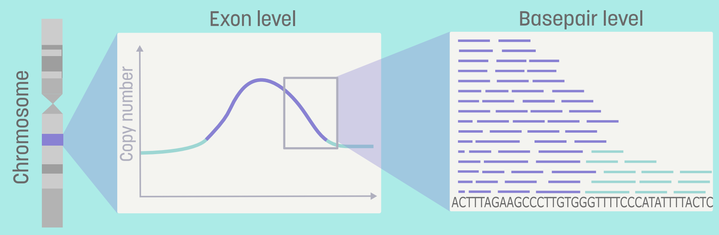
We are using a design study methodology and focus on copy number variant visualization in the clinic.

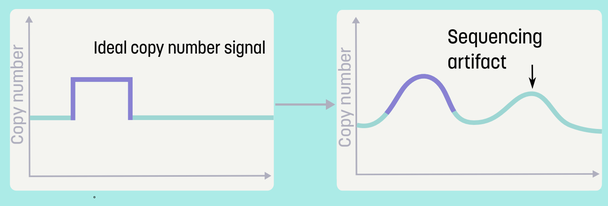
Design studies emphasize collaboration with experts and end users, as well as immersion in their world and view of the data. The primary goal is to understand the data and tasks that users need to carry out with the data.

These results are from:

- Two semi-structured deep interviews with analysts
- Three site visits to hospital genomics labs
- One shorter interview with two geneticists

Results: Five challenges

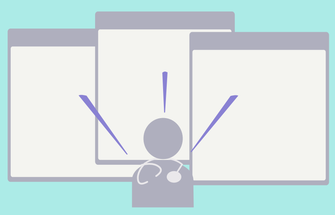


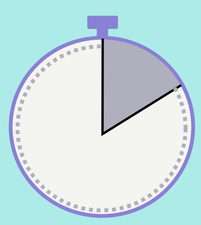


Many times there appears to be a duplication which can not be confirmed with other methods.

The data must be closely reviewed, is often run through a number of programs, and is compared to other samples.

When a variant is found it is looked up in multiple databases to understand its clinical relevance.



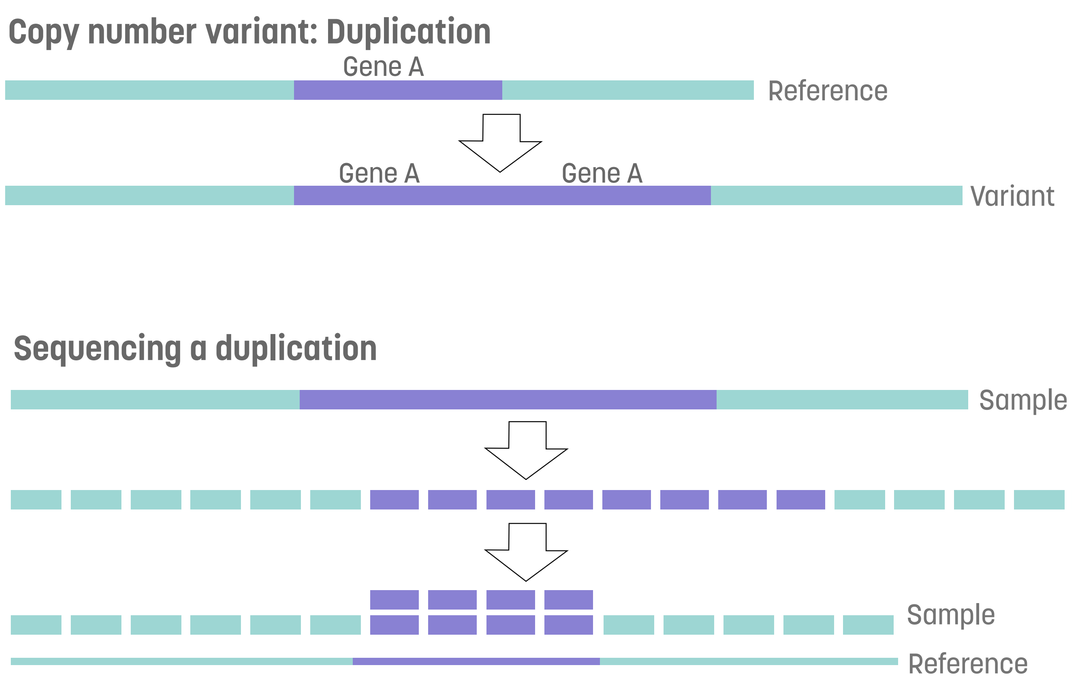


- **Multiscale data** with important information on all levels
- **Artifacts** introduced by the sequencing can cause confusion
- **Unknown significance** of a finding means that it can not be used for diagnosis or chosing a treatment
- **Multiple sources** of information means many windows and tabs to keep track of
- **Efficiency** is needed since there are many cases to review

Background

Next generation sequencing (NGS) data is **large** and **noisy**. Clinicians look for genetic variants, and assess if they are real and their effect on the patient. Harmful genetic variants can cause **cancer** and **inherited diseases**.

Copy number variants are large sections of the genome that have been duplicated or deleted. They are **longer than the DNA fragments** used in DNA sequencing, so they are **difficult to find.**



When a gene is duplicated the patient has an extra copy of that gene, compared to the reference genome.

The sample DNA is split into fragments which are sequenced. If a patient has a duplication, there will be more fragments from the duplicated area. The fragments are mapped to a reference genome, and then the ones from the duplication pile on top of each other.

The sequencing process introduces multiple biases and erors, which study subjects had to deal with in their day to day work.

LINKÖPING UNIVERSITY

SECTRA