New concept proposal

Gene Panel

Author	Eelke van der Horst, Femke Kopmels	**Date last updated**	31/10/2023
Project	General interest	**Contact**	DCC
Dataset release	2024.1	**Consulted expert**	-

1 Rationale

Gene panels are used for targeted screening in both clinical and research applications. When a gene panel is used for target enrichment as part of Library Preparation, information on the gene panel and its focus genes are required to interpret downstream results. Therefore, we propose to introduce a *Gene Panel* concept that can be used to add metadata on the focus genes of the panel to the *Library Preparation* concept, which is part of a *Sequencing Assay*.

2 Comparison to other standards/data models

2.1 FAIR Genomes

FAIR Genomes uses 'target enrichment kit' (fg:Sample_preparation_Target_enrichment_kit), with definition "Indicates which target enrichment kit was used to prepare the sample. Target enrichment is a pre-sequencing DNA preparation step where DNA sequences are either directly amplified (amplicon or multiplex PCR-based) or captured (hybrid capture-based) in order to only focus on specific regions of a genome or DNA sample". It refer to NCIT:C154307 | enrichment |, which is defined as "any of various techniques designed to select or increase a target item in a mixed sample". FAIR Genomes' target enrichment kits are of type GENEPIO:0000081 | sequencing kit |.

2.2 ENA

ENA has 'library selection' as an experiment information attribute, which is part of a sequence read submission. Values can be selected from a list of permitted values. However, this list is quite generic with values such as "PCR: target enrichment via PCR", without providing the name of the specific gene panel kit or information on the genes in the gene panel.

3 Concept information

Concept or concept compositions or inherited	General concept name	General description	Contextualized concept name	Contextualized description	Type	Standard	Value set or subset	Meaning binding	Cardinality for composedOf
concept	Gene Panel	collection of genes that have been grouped for testing	Gene Panel	collection of genes that have been grouped for testing					
composedOf	focus gene	gene of focus associated to the concept	focus gene	gene that is targeted by the gene panel	Gene				1:n
composedOf	data file	data file associated to the concept	data file	BED file associated to gene panel	Data File				0:n

General concept name	Cardinality for concept to Administrative Case	Cardinality for concept to Data Provider	Cardinality for concept to Subject Pseudo Identifier	Cardinality for concept to Source System
Gene Panel	-	1:1	-	-

4 Impact on the SPHN Dataset

Optional (if existing concepts need to be adapted because of this new concept, state here the currently released version of the existing concept and the proposed adapted version)

5 Discussion

We propose to separate the physical gene panel kit from the information on the list of focus genes of the panel with an optional BED file. This allows for the option of providing custom gene panels that don't have a brand name, for instance in the case of custom gene panels, and also allows to only use a gene panel kit's name without providing information about the focus genes. The gene panel kit is indicated by Library Preparation's target enrichment kit property. Note that targeted sequencing sequencing can also target segments of chromosomes, *i.e.* much larger than specific genes.

6 Example

Example of AmpliSeq for Illumina BRCA Panel. The name of the kit is indicated in Library Preparation > target enrichment kit.

focus gene:
- gene identifier
  - code: HGNC:1100 |BRCA1 DNA repair associated|
- organism
  - code:
    - name: Homo sapiens
    - coding system and version: NCBI Taxonomy - 2023-07-28
    - identifier: 9606

focus gene:
- gene identifier
  - code: HGNC:1101 |BRCA2 DNA repair associated|
- organism
  - code:
    - name: Homo sapiens
    - coding system and version: NCBI Taxonomy - 2023-07-28
    - identifier: 9606