New concept proposal

Protein

Author	**Jan Armida**	**Date last updated**	**01/10/2022**
Project	Genomic Concepts	**Contact person**	Jan Armida
Status	Accepted	**Consulted expert**	WG

1 Rationale

Alterations at nucleic acid level can influence protein structure and function which may lead to phenotypic variability and development of disease. The possibility to add a layer of information beyond genomics and refer to a specific protein is therefore an important and potentially actionable piece of information that needs to be communicated to the clinician.

2 Comparison to other standards/data models

2.1 HL7 FHIR

In HL7 FHIR a protein can be described using the element “Sequence” or “Molecular Sequence” and by specifying the Sequence Type as Amino Acid Sequence.

2.2 SNOMED CT

SNOMED CT provides various codes related to “Protein”. Notably it is possible to refer directly to specific proteins with a well-known function or classes of proteins important in clinical settings. In general, it is possible to refer to a “Protein” as a substance using the code 88878007 | Protein (substance) |

2.3 GA4GH Phenopackets

GA4GH Phenopackets do not contemplate the building block for protein. Instead, it is possible to refer to a specific protein using the GA4GH VRS (Variant Representation Specification) standard. The link is therefore made within a *Variant Descriptor* block that de facto extends the existing VRS.

3 Concept information

Contextualized concept name	Contextualized description	Type	Standard	Value set	Meaning binding
Protein	molecule composed by one or more chains of amino acids				SNOMED CT: 88878007 | Protein (substance) |; SO:0000104 polypeptide
protein identifier	unique protein id according to a specific nomenclature, e.g. UniProtKB	Code	UniProtKB or other		
organism	organism associated to the protein	Organism			

4 Impact on the SPHN Dataset

The addition of *Protein* does not require any further change in the current SPHN Dataset release.

5 Discussion

The possibility to refer to a protein as a separate instantiation and not only when associated to a specific variant increase the flexibility and reusability of the concept in different contexts.