

Afrapportering på data management pilot projekt

Afrapporteringsdato	2018, January 15th	Afrapportering #	2
Projekt navn	Actionable Biomarkers DK		
Projekt start	2016, August 1st	Projekt slut	2018, June 30th

Realiserede resultater/leverancer ultimo 2017

- **WP1: Data capture, data harmonization, conversion of unstructured data into structured biomarker formats**
- **WP2: Data management effort addressing primary data types: genome and proteome sequences**
- **WP3: DTU text mining effort addressing full length papers for novel biomarker detection**

Westergaard D, Stærfeldt H-H, Tønsberg C, Jensen LJ, Brunak S. Text mining of 15 million full-text scientific articles. bioRxiv 2017; : 162099.

- **WP4: Data management workflows implementing the condensation of genomic and proteomic data into actionable biomarkers**

Besenbacher, S. et al. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. Nat Commun 6, 5969 (2015).

Marett, L. et al. Sequencing and de novo assembly of 150 genomes from Denmark as a population reference. Nature Skov, L., Consortium, T. D. P.-G. & Schierup, M. H. Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion. PLoS Genet. 13, e1006834 (2017).

Jensen, J. M. et al. Assembly and analysis of 100 full MHC haplotypes from the Danish population. Genome Res. gr.218891.116 (2017). doi:10.1101/gr.218891.116

Vazquez, M., Pons, T., Brunak, S., Valencia, A. & Izarzugaza, J. M. G. wKinMut-2: Identification and Interpretation of Pathogenic Variants in Human Protein Kinases. Hum. Mutat. 37, 36–42 (2016).

Forventede resultater/leverancer medio 2018

- **WP5: Secure private cloud effort for biomarker workflows on Computerome (DTU/KU) and ABACUS 2.0**
 - a) Establish a "joint cloud" integrating Abacus and Computerome for the life sciences; Negotiations to further promote this idea ongoing.
 - b) In progress for genomics - proteomics is not in the loop from the start, but is expected to join shortly. VM containerization down to network level (=> HW agnostic) and APIs for automating the ordering of resources have already been completed. Establishing tunnel and call back between Computerome and Abacus for storing and updating the biomarkers is in progress. Likewise, with resource pooling.

Konstaterede eller forventede afvigelser i forhold til de planlagte leverancer

Her ønskes ingen lange forklaringer.

Milestone (T2.1): A cloud compatible, implemented workflow for genomics and proteomics data preparing for biomarker extraction (M12).

Challenged by lack of assets able to understand the particular life science data analysis workflows. Furthermore, the Abacus technology lacks sufficient support. Current status is that, albeit still subtle to further consideration, an updated and enhanced workflow for biomarker extraction is readily available.

Forbrugt bevilling

Bruges hele bevillingen?

The budget is geared through other project funding and the project is on track to reach all milestones **in time**.