D2.2

Ultra deep sequencing of prognostic biomarkers

Project number:	668858
Project acronym:	PrECISE
Project title:	PrECISE: Personalized Engine for Cancer Integrative Study and Evaluation
Start date of the project:	1st January, 2016
Duration:	36 months
Programme:	H2020-PHC-02-2015

Deliverable type:	OTHER
Deliverable reference number:	PHC-668858 / D2.2 / 2.0
Work package contributing to the deliverable:	WP 2
Due date:	October 2017 – M22
Actual submission date:	1st December, 2017

Responsible organisation:	BCM
Editor:	Pavel Sumazin
Dissemination level:	PU
Revision:	V2.0

Abstract:	This report describes an effort to sequence prognostic biomarkers for prostate cancer
Keywords:	Graphical user interface, secure databank

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 668858.

This work was supported (in part) by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0324-2. The opinions expressed and arguments employed therein do not necessarily reflect the official views of the Swiss Government.

Editor

Pavel Sumazin (BCM)

Contributors

Pavel Sumazin (BCM)
Hyunjae Ryan Kim (BCM)
Maria Rodriguez Martinez (IBM)
Peter Wild (UZH)
Laura De Vargas (UZH)

Disclaimer

Executive Summary

We designed a prognostically predictive biomarker panel that targets thirty-six commonly mutated predictive genes and used this panel to profile forty-two Castraction Resistant Prostate Cancer (CRPC) patients. We report on predictive variants identified in each of the patients. To improve clonal inference and estimates of clonal composition we selected fifteen patients, based on these profiles and from D2.1. Some of these patients will be profiled for additional areas, and all will be profiled for copy number variations. These will be reanalysed by Chimaera (Manica et al., 2017) to reconstruct partial phylogenies based on observed frequencies of these mutations and copy numbers changes of their corresponding alleles. Copy number arrays will be used to test Chimaera's ability to correctly deconvolve mutation frequencies and copy numbers.

Revision 2.0: In this revision upon request all the appendices are made available for download through a public link. The appendices mentioned in the deliverable can be downloaded from the following link:
https://bcm.box.com/s/qxybcfjqr17t4oet321ykcxy61zqohyy

Contents

List of Tables

Chapter 1 Prognostically predictive panel

We selected a total of 36 genes whose mutations are enriched in prostate cancers for ultra-deep sequencing. ImmQuant was used to capture their coding regions and the completeness of the capture was verified against the human reference genome GRCh38. Target genes are given in the table below.

AKT1	CDH1	MED12	PMS2
AR	CDKN1B	MRE11A	PTEN
ATM	CHEK2	MSH2	RAD51C
ATR	EP300	MSH6	RAD52D
AURKA	ERG	MYC	RB1
BARD1	EZH2	MYCN	SPOP
BRACA1	FOXA1	NBN	TMPRSS
BRACA2	GEN1	PALB2	TP53
BRIP1	HOBX13	PIK3CA	ZNF595

Table 1: Target Genes

Targeted regions, human reference hg19, are provided in Appendix 1.

Chapter 2 Profiling

A total of forty-two The Prostate Cancer Outcomes Cohort Study from UZH (ProCOC) and Metastatic Prostate Cancer Biobank from UZH (metaProC) patients were selected for ultra-deep sequencing using our predictive panel. Their profiles—implemented in three batches and sequenced using Illumina HiSeq using 150bp pair-end reads by Sophia Genetics and with QC given in Appendix 2—produced 1,799 recurring candidate mutations with average coverage of 1,590 reads per sample. In total, across all samples, these mutations were observed 7,790 times. The mutations are given in Appendix 3. We are preparing a publication based on this data and will make BAM files freely available on ENA project PRJEB19193.

Chapter 3 Selection for additional profiling

Of the 1,799 recurring candidate mutations listed in Appendix 3, we selected sixty-eight recurring somatic mutations that alter protein expression, with MAF < 0.01 according to ExAC (Lek et al., 2016), favourable scores according to SIFT (Kumar et al., 2009) and Polyphen (Adzhubei et al., 2010), observed mutation frequencies above 2%, observation in COSMIC (Forbes et al., 2016), and coverage above six-hundred reads. These mutations targeted seventeen genes, but most importantly, many of these mutations co-occurred in the same patients. These mutations are given in Appendix 4.

We are particularly interested in studying mutation pairings to elucidate initiating mutations from mutations that impart growth and resistance advantages. Consequently, we selected fifteen patients for follow-up, including copy-number profiling and the sequencing of additional areas. In particular, we were interested in studying the relationship between P53 and PTEN, PIK3CA and BRCA2, and FOXA1 and P53 mutations pairs. The selected fifteen patients include twelve that were selected from the forty-two profiled as a part of the D2.2 cohort and three patients profiled as a part of the revised D2.1 cohort. These patients include P1, P2 and P9 from D2.1, and S24, S30, S36, S38, S47, S56, S57, S60, S62, S66, S9, and S41. The D2.1 cohort included ten ProCOC patients that were profiled using whole-exome sequencing and our mutation panel. Samples include patients with multiple resections across received therapy.

Chapter 4 Conclusion

For this deliverable we profiled forty-two patients using ultra-deep sequencing and targeting key prostate-cancer genes. Our results suggested that most patients have known cancer mutations in more than one locus, leading to a fundamental question: for recurring mutation pairs, is one mutation always initiating? If so, which mutation should be targeted? To begin resolving this open problem, we chose 15 patients with recurring mutation pairs for further molecular profiling and phylogeny reconstruction by Chimaera. Copy number assays, together with mutation profiles will be used to test Chimaera's ability to reconstruct mutation frequencies, copy number, and the order of mutation acquisition.

Chapter 5 References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nature methods *7*, 248-249.

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., and Ponting, L. (2016). COSMIC: somatic cancer genetics at high-resolution. Nucleic acids research *45*, D777-D783.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols *4*, 1073-1081.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B.*, et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285-291.

Manica, M., Chouvarine, P., Mathis, R., Wagner, U., Oehl, K., Saba, K., Roditi, L.D.V., Pati, A.N., Martínez, M.R., Wild, P.J.*, et al.* (2017). Inferring clonal composition from multiple biopsies of the same tumor. arXiv, 1701.07940.