The genomic DNAs were exacted and sonicated to an average size of 200bp (range 100-500bp). The targeted DNA fragments were captured pulldown and exon-wide libraries were created using the Roche SeqCap EZ Exome V3 and TruePrep DNA Library Prep Kit V2 for Illumina (#TD501, Vazyme, Nanjing, China), and paired-end sequence data was generated using Illumina HiSeq machines with the average sequencing depth of 30x and totally xx unique reads for xx sample in xx individuals. The sequence data, aligned to the human reference genome (NCBI build 37) using BWA [1], and sorted and removed PCR duplication using GATK 4.0 [2]. Somatic mutation calling was performed using Strelka2 [3] and VarDict [4]. Somatic mutations existing in at least two of the results of the three software were selected as high confident mutations and be involved in the further bioinformatics and bio-statistical analysis. Transition (Ti) and transversion (Tv) ratio was applied to measure the selection in cancer genomes and to show the characteristics of mutation profiles. Mutation and polymorphisms were annotated with the Ensembl Variant Effect Predictor (VEP) [5]. Most frequent mutations in the prostate cancer samples were shown with ComplexHeatmap (version 2.0.0)[6] and oncoprint with custom R script (R version:3.6.0). Genetic distances between samples were calculated with the hamming distance using the gene-based presence and absence of somatic mutations across samples. The neighbor joining algorithm from the phonon[7] were applied to estimate phylogenetic trees and MEGA7 [8] was used to show phylogenetic relationships between tumor sectors for each patient. Trunk, branch and private mutations were defined as the mutation was occurred in more than three, two and one samples in our project. Gene functional annotation were based on DAVID[9] and ReactomePA [10]. Genome-wide mutation-clinical analysis were conducted with linear regression, logistic regression or Wilcoxon test (rank sum or signed rank test) determined by distribution of response variable and explanatory variables. We didn’t conduct multiple test correction such since we are exploratory research and we want to identify certain interesting signals for further validation. We also conducted association study between mutation burden and clinical characteristics with same method above in which mutation burdens were averaged for each individual for multiple samples.

1. Slater, P.M., R. Grivell, and A.M. Cyna, *Labour management of a woman with carnitine palmitoyl transferase type 2 deficiency.* Anaesth Intensive Care, 2009. **37**(2): p. 305-8.

2. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.

3. Kim, S., et al., *Strelka2: fast and accurate calling of germline and somatic variants.* Nat Methods, 2018. **15**(8): p. 591-594.

4. Lai, Z., et al., *VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research.* Nucleic Acids Res, 2016. **44**(11): p. e108.

5. McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.

6. Gu, Z., R. Eils, and M. Schlesner, *Complex heatmaps reveal patterns and correlations in multidimensional genomic data.* Bioinformatics, 2016. **32**(18): p. 2847-9.

7. Schliep, K.P., *phangorn: phylogenetic analysis in R.* Bioinformatics, 2011. **27**(4): p. 592-3.

8. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.* Mol Biol Evol, 2016. **33**(7): p. 1870-4.

9. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

10. Yu, G. and Q.Y. He, *ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization.* Mol Biosyst, 2016. **12**(2): p. 477-9.