



The Proteus Genome Project

Interim project progress report

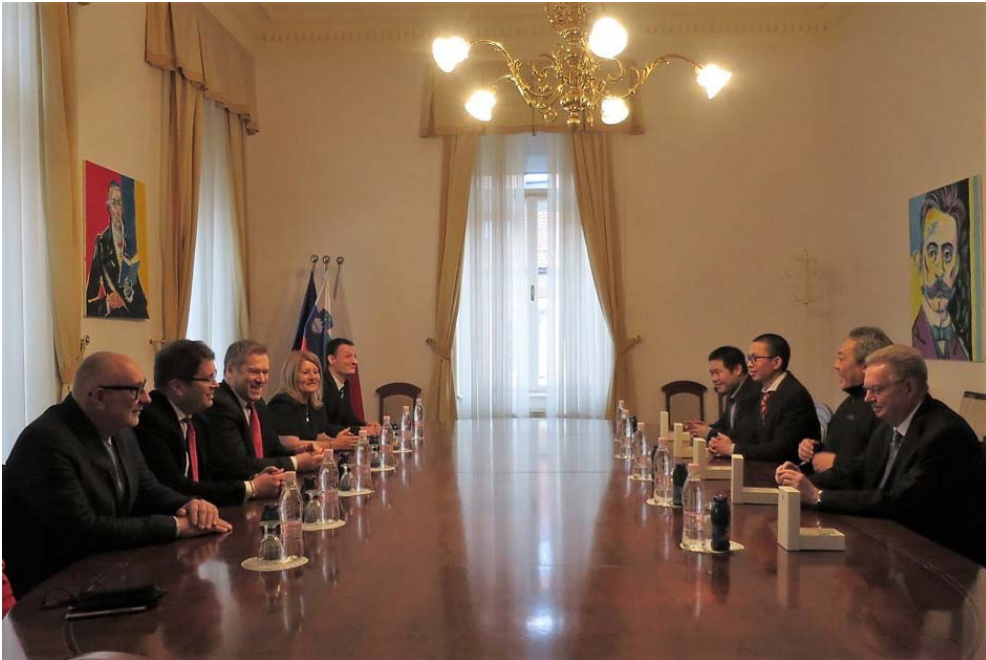
Yonglun Luo

Executive Director of the Lars Bolund Institute of Regenerative Medicine

25/11/2019

1. Project background and technology

Initiation of the PGP project



29 March. 2018

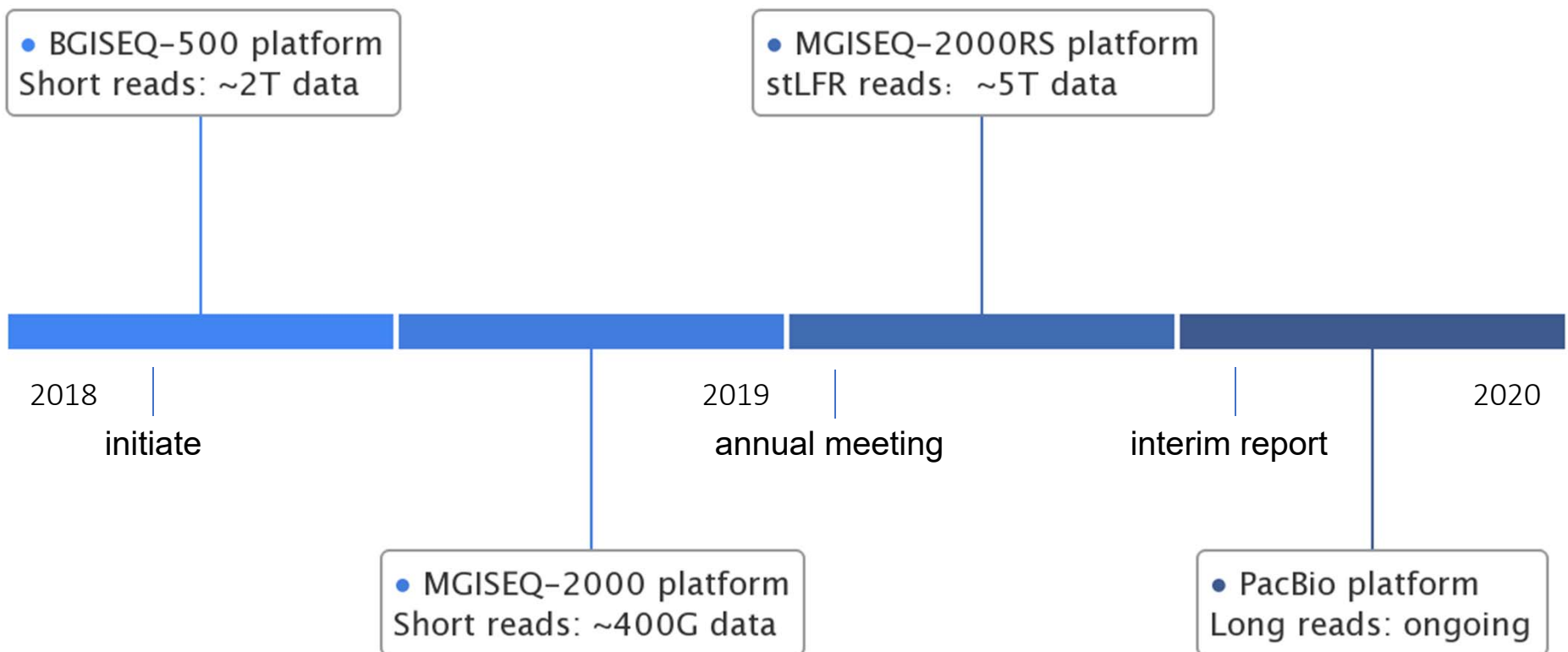
MGItech - DNB^{seq}

+



2. An overview of PGP

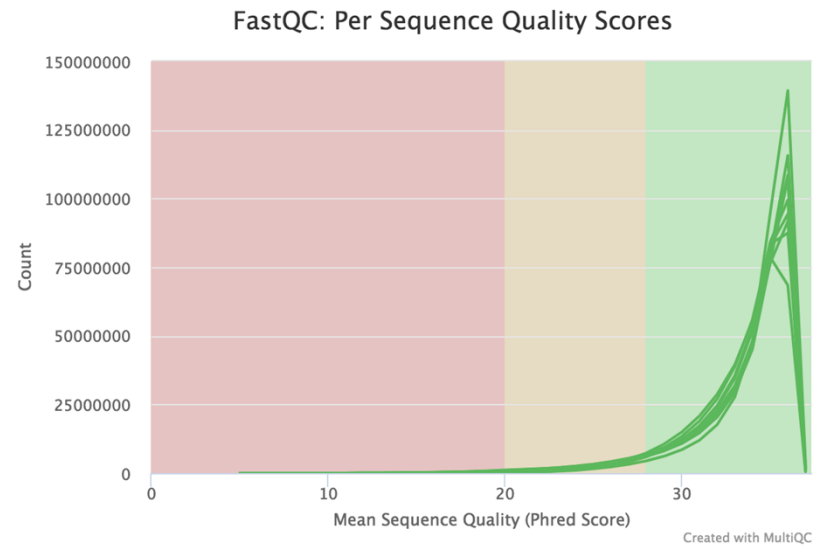
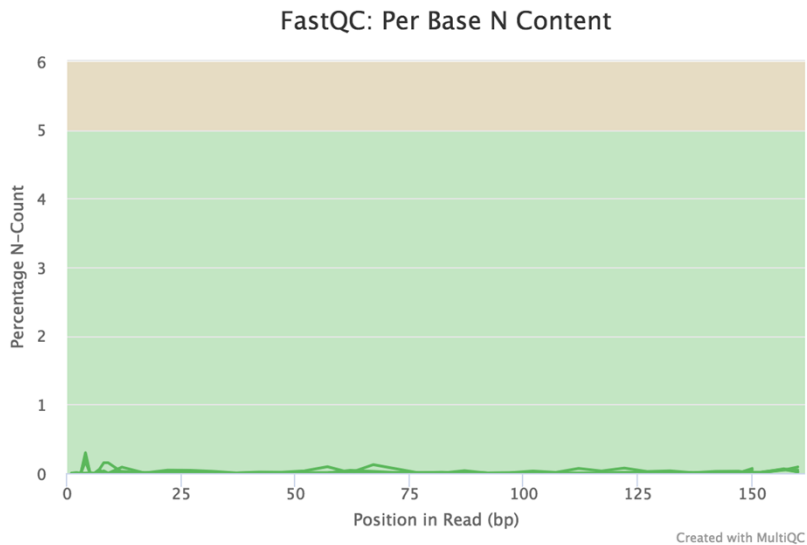
Two NGS platforms, Two technologies



3. Data output and QC

Total PGP output
Genome data = 7 Tb
Transcriptome data = 0.5 Tb

= **2500 human genome**

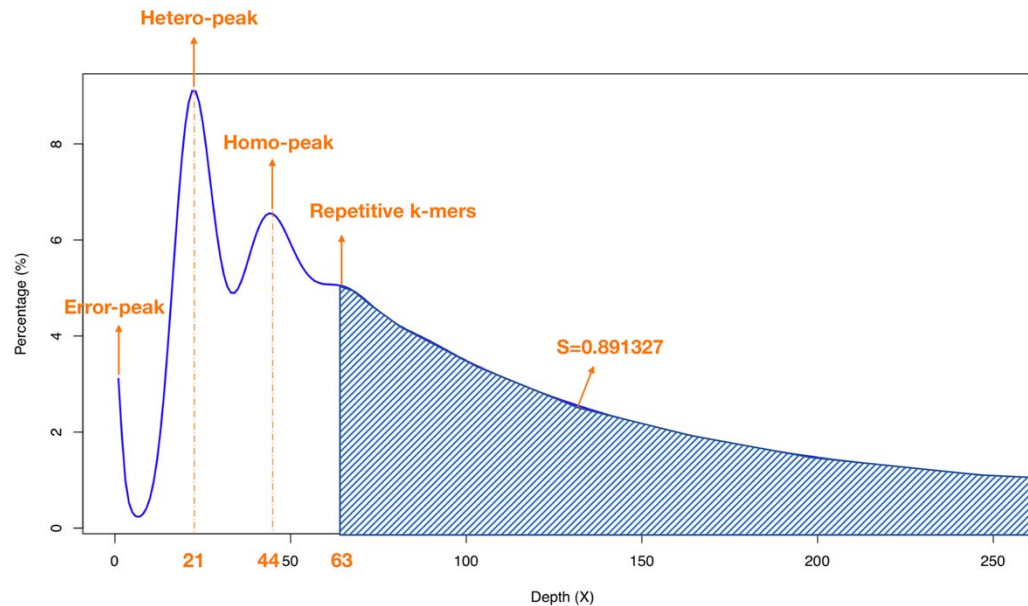


4. Interim Genome Assembly

**Proteus Genome Size
= ~ 42 Gb**

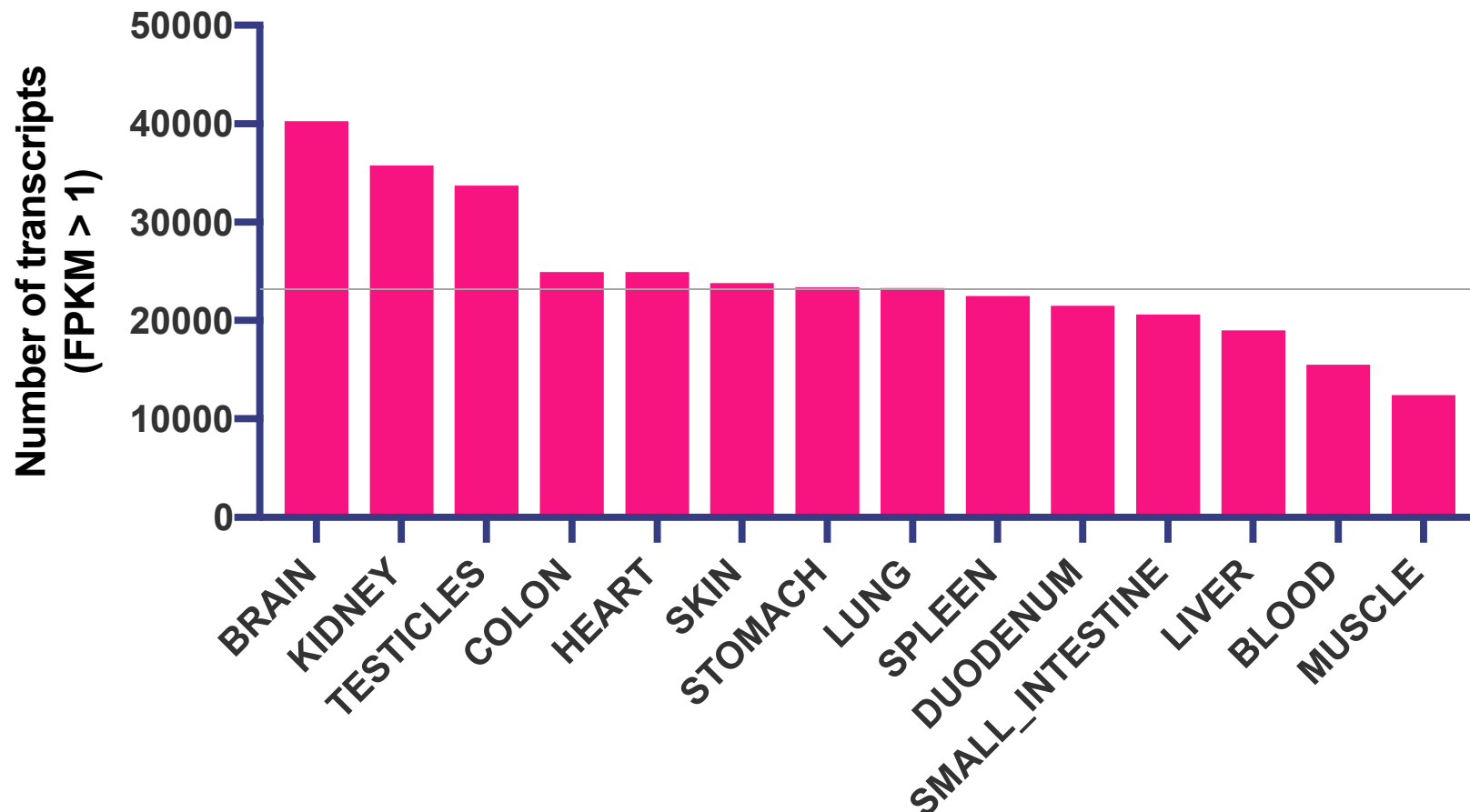
**Repetitive sequences
= 89%**

14X human genome



5. Interim Transcriptome Analysis

14 organs/tissues, average > 20,000 expressed gene



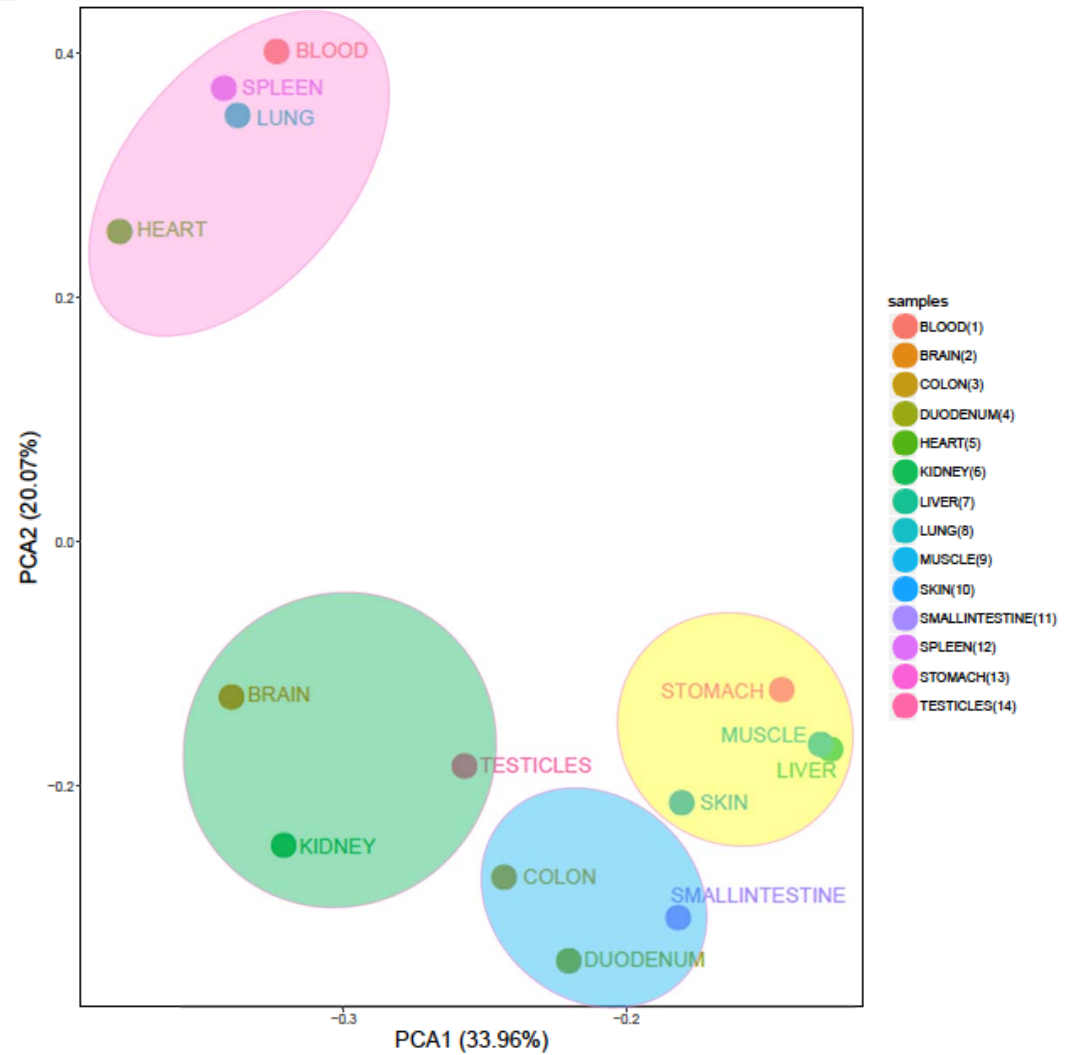
5. Interim Transcriptome Analysis

Immune/blood systems

Neural/barrier system

GI tract

Metabolic system



6. Future actions

1. Complete the single molecule long-read sequencing
2. Assemble the proteus genome as complete as possible
3. Analyze the proteus genome to provide novel insights for the evolution of this iconic species
4. Complete the analysis of transcriptome data
5. Report the PGP project in a top scientific journal

thanks

