Genome Institute of Singapore

www.a-star.edu.sg/attract

ATTRaCT

Thursday October 20, 2016
2:00-3:00pm
1613T

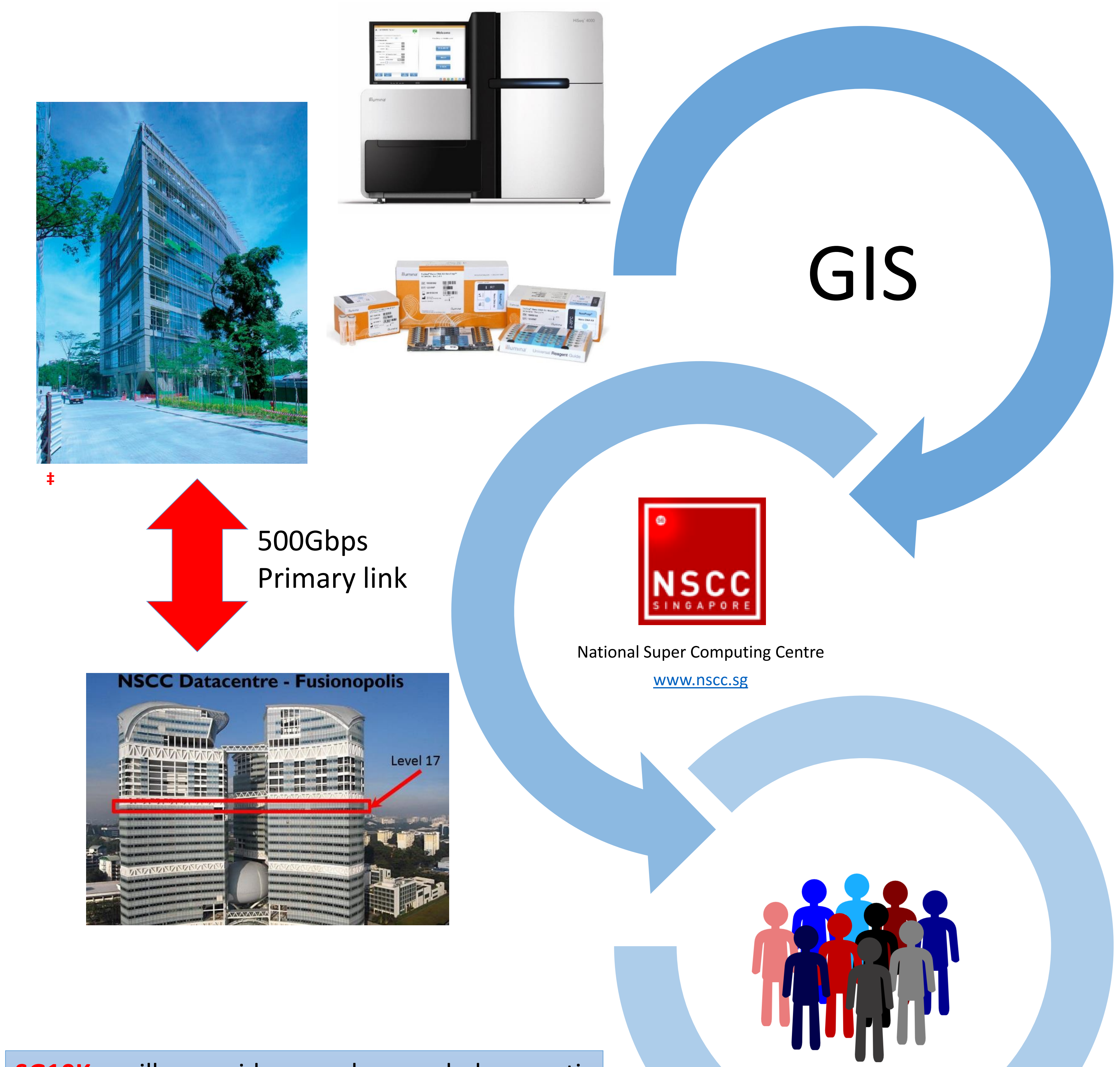
Agency for Science, Technology and Research
SINGAPORE
CREATING GROWTH, ENHANCING LIVES

Introducing “SG10K”: Cataloging genetic diversity and population structures in 10,000 South Asians

<u>Bellis C</u>[1], Irwan ID[1], Wang C[2], Soon WW[3], Wilm A[4], Shih CC[4], Ng HH[5], Liu J[1] & SG10K consortium
[1]Human Genetics 2, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore (A*STAR), Singapore
[2]Computational and Systems Biology, Genome Institute of Singapore, A*STAR, Singapore
[3]Next Generation Sequencing Platform, Genome Institute of Singapore, A*STAR, Singapore
[4]Scientific and Research Computing, Genome Institute of Singapore, A*STAR, Singapore
[5]Gene Regulation Laboratory, Genome Institute of Singapore, A*STAR, Singapore

Genetic variation plays an important role in a variety of human diseases and quantitative traits. Due to different underlying genetic architecture and contrasting environmental exposures, many genetic findings have shown population-specific characteristics, highlighting the importance of population diversity in human genetic studies. The Singapore population consists of three major ethnic groups, Chinese, Malay, and Indian, which together represent **~80%** of the Asian population. To empower biomedical and human genetic studies in Asian populations, the SG10K project will perform 12-13× WGS of 10,000 Singaporeans. Coupled with powerful bioinformatics tools, our study design will enable high-quality genotype calling for a full frequency spectrum of genetic variants segregating in the population. Our main objectives are to **(1)** comprehensively characterize genetic variation in Singapore population; **(2)** create a WGS reference panel for accurate genotype imputation in Asian population; and **(3)** generate a large control dataset for WGS-based genetic association study of diseases. Our extensive calculations have provided substantial preliminary evidence allowing us to confidently proceed with our plan to survey our 10,000 samples at a depth of coverage between 12-15× in our WGS strategy. Because the phased variant calling will become more powerful with the increased sample size (via more accurate haplotype information), >10× WGS will allow us to characterize the full spectrum of germ-line genetic variants (except the private ones) in 10K individuals with the similar accuracy and sensitivity as 30× WGS. To achieve our goals, we will be employing the Illumina TruSeq Nano DNA Library Preparation Kit and sequencing study participants on the Illumina HiSeq 4000 instrument, in-house at GIS. However, this project is a fully national effort across multiple institutions; our collaborative partners include **SingHealth Duke-NUS Institute of Precision Medicine**, **Singapore Eye Research Institute**, **Centre for Personalised and Precision Health**, **National University Health System** and several **Translational and Clinical Research Flagship** Programmes (Heart failure, Parkinson disease). To date we have taken possession of the complete SG10K cohort and generated >900 WGS.

Upon completion, this study will provide valuable genetic information to facilitate clinical and pharmaceutical research in Singapore populations and will empower genetic studies of Singapore and Asian-centric diseases.


‡
GIS
500Gbps
Primary link
NSCC Datacentre - Fusionopolis
Level 17
NSCC SINGAPORE
National Super Computing Centre
www.nscc.sg

SG10K will provide much needed genetic information of Asian populations while maintaining GIS and Singapore's ***leading position*** in human genetics in Asia

Objectives

- **Characterize genetic variation in Singapore**
- Create a reference panel for genome imputation in Singapore population
- Generate a control dataset for GWS-based genetic association study of disease

Strategy

- 10,000 Singaporean individuals, covering three local ethnic groups: Chinese, Malay, and Indian (representing ~80% of Asian population)
- WGS at **12-15×** coverage
- IT infrastructure for data share with local research community and international research efforts

Impact

- Precision medicine
- Industrial collaboration
- Biomedical research

Precision medicine

1. Provide a ***catalog*** of genetic variants and their frequency information for local populations.
2. ***Genomic diagnosis*** in congenital or rare diseases
3. Frequency for the relevance and economic analysis of ***PGx*** genetic biomarkers in local and Asian populations
4. <u>National genetic database</u>
   - Design optimized SNP arrays
   - Genotype large national cohort(s), up to 500K
   - Genome-wide coverage of genetic variants by ***imputation***

Industrial collaboration

1. Demonstrate Singapore’s capability in ***large-scale*** genome sequence analysis
2. Extensive catalog of functional variants, particularly ***loss-of-function*** variants in Asian population
3. Coupled with additional genotyping confirmation, empower genotype-based recall study with industrial partners

Biomedical research

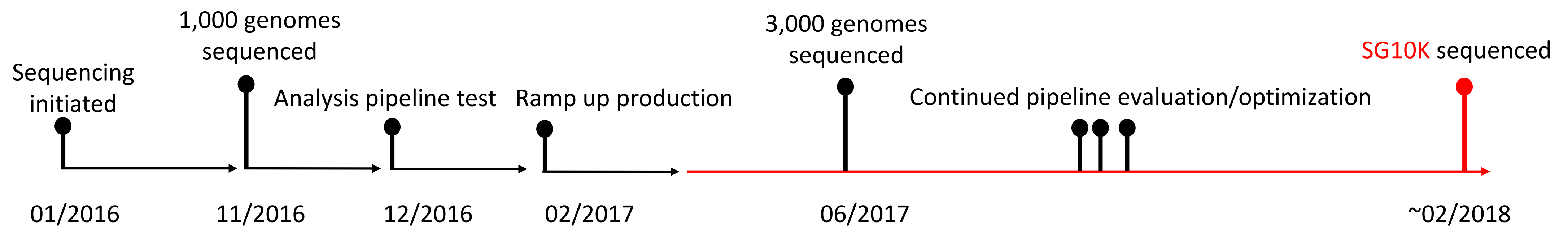
1. ***Efficient*** genomic imputation using large reference panel
2. Enhance coverage of genetic variants and thus the ***scientific value*** of existing GWAS datasets
3. Create common control database that will ***empower*** future WGS-based genetic studies of disease
4. Provide datasets for methodological development
5. ***Genetic structure*** of local populations

<u>Study status</u>; SG10K production sequencing to date has generated almost 1,000 WGS. We aim to run analytical pipeline assessment and optimization reviews at several milestones, namely n=1,000/3,000/5,000 WGS.

SG10K represents a significant undertaking for GIS, A*STAR and the local research community that will position Singapore at the forefront of the precision medicine thrusts, globally.

Without the support and commitment from our funding bodies and study collaborators, our initial objectives would not be possible.

Projected timeline



ASHG 2016
VANCOUVER • OCTOBER 18-22, 2016

‡ https://commons.wikimedia.org/w/index.php?curid=1370607