Developing the Korean human genomic DNA reference material for genomic sequencing

Young-Kyung Bae[1], Jina Kim[2,3], Joohon Sung[2,3], Inchul Yang[1]

[1]Center for Bioanalysis, Division of Chemical and Medical Metrology, KRISS, Daejeon, Republic of Korea; [2]Interdisciplinary program in Bioinformatics, [3]Genome & Health Big Data Branch, Department of Health Science, School of Public Health, Seoul National University, Seoul, Republic of Korea

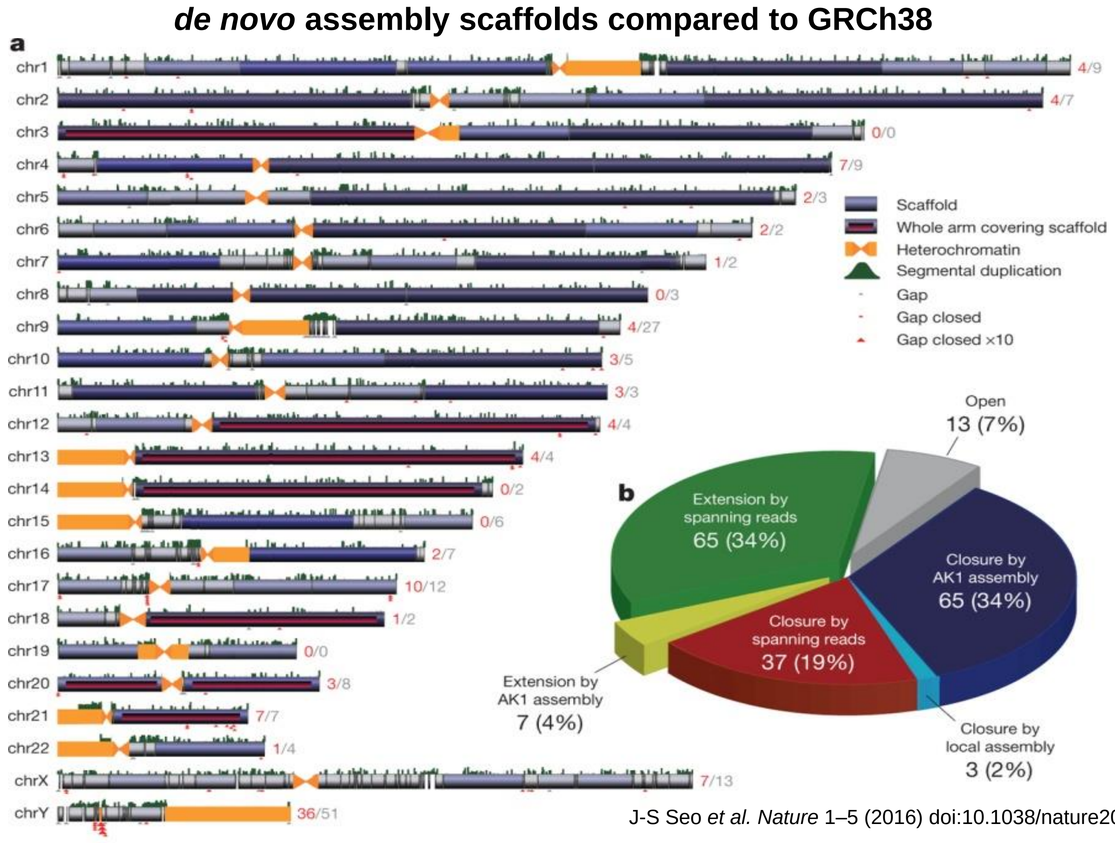
Presenter: Ji-Youn Lee

Abstract

Genome sequencing has become a key component in precision medicine. The majority of sequencing practices, short reads are compared and mapped using the GRCh37/38 human genome assembly as the reference. In order to cover putative nucleotide variants and structural alterations in Koreans, we have developed the Korean human genomic DNA reference material (KRISS RM 111-10-014) and the corresponding database with a variant call file containing SNPs (single nucleotide polymorphisms) and small insertions and deletions. Also a tab-delimited bed file is available in the following link (http://147.47.68.110:3000/). These together can be applied to whole genome sequencing, exome sequencing, and targeted sequencing. Specifically, the genomic DNA reference material can be used for evaluating reliability of library construction, sequencing chemistry and the downstream bioinformatics algorithms for mapping, alignment, and variant calling. We expect that this pair of the matched reference material and the genome database will be valuable in improving the public healthcare in the era of precision medicine.

Utilizing AK1 information for the Korean reference genome

***de novo* assembly scaffolds compared to GRCh38**



J-S Seo *et al. Nature* 1–5 (2016) doi:10.1038/nature20098

a, Scaffold coverage over GRCh38 per chromosome. The blue shading represents scaffold size, with darker segments for longer scaffolds. Eight chromosomal arms are spanned by single scaffolds. Closed euchromatic gaps are labelled in red on each chromosome, with the total number of gaps in grey. b, Number of gaps closed using the AK1 assembly (blue), local assembly of long reads (light blue), and long reads alone (red). The number of extended gaps with AK1 assembly is represented in yellow, with long reads in green and open gaps in grey. The 65 dot plots of gaps closed with the AK1 assembly can be found in the AK1 genome browser (http://211.110.34.36/gbrowse2).

Production of KRISS human genomic DNA reference material

Analytical methods

For the determination of sequence information of the Korean genomic DNA, next generation sequencing also known as high throughput sequencing was used. The method involves genomic DNA extraction from cells, DNA quantification by UV-spectrophotometer and dNMP-based LC-MS/MS. The LC-MS/MS results were used to assign the concentration of DNA solution. For the sequencing raw data acquisition, the DNA library was constructed and applied to Illumina Hiseq X sequencer. The read length was set at 150 base pairs, and the mappable mean depth (post-alignment) was 56.30. For bioinformatics analysis, the fastq raw data from the sequencer is compared to the most current NCBI reference genome GRCh38. First, the fastq reads are aligned using Isaac aligner (developed by Illumina) using the parameters below and identified variants using Isaac variant caller version 2.0.13 with its default settings (base-quality-cutoff: 15, keep-duplicate: 1, default-adapters: AGATCGGAAGAGC*, *GCTCTTCCGATCT). Sequencing and bioinformatics analysis was performed by a subcontractor listed below. Results generated by the subcontractor were validated by KRISS.

Validation of the subcontractor's method for sequencing the reference genome

(1) The validity of the subcontractor's method was evaluated by comparing variant calling results (vcf) generated by the subcontractor after sequencing the NIST RM 8398 and those provided by NIST for this RM.

	Specificity (%)	Accuracy (%)
NIST high confidence vcf vs. subcontractor's vcf	99.97	99.97
NIST total vcf vs. subcontractor's vcf (alignment tool was matched: novoalign)	99.99	99.99

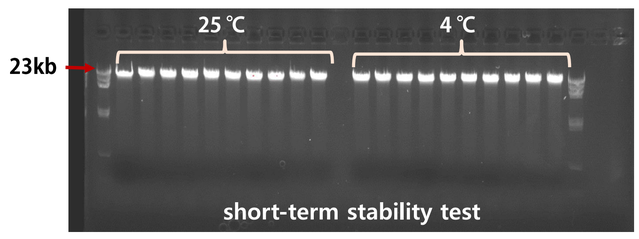
(2) the KRISS RM 111-10-014 was also analyzed by an independent laboratory (Theragen, Korea). The vcf files generated by these two laboratories were compared for their specificity and accuracy, which showed a good agreement. Vcf files against GRCh37 reference assembly for each data set was used for this analysis.

	Specificity (%)	Accuracy (%)
Theragen's vcf vs. subcontractor's vcf	99.99	99.98

(3) Additional validation tests using targeted sequencing for 107 SNV regions identified in KRISS RM genomic DNA when compared to NCBI reference genome GRCh38. Multiplex PCR products were pooled and applied to NGS, which generated 10,875,484 reads per vial on average. All the targeted SNVs called for within each amplicon was accurately matched with results from vcf data obtained from whole genome sequencing.

Stability & Homogeneity test

(1) By gel electrophoresis and UV spectrophotometer, the physical properties and DNA concentration were measured. The size of DNA remained unchanged over time.

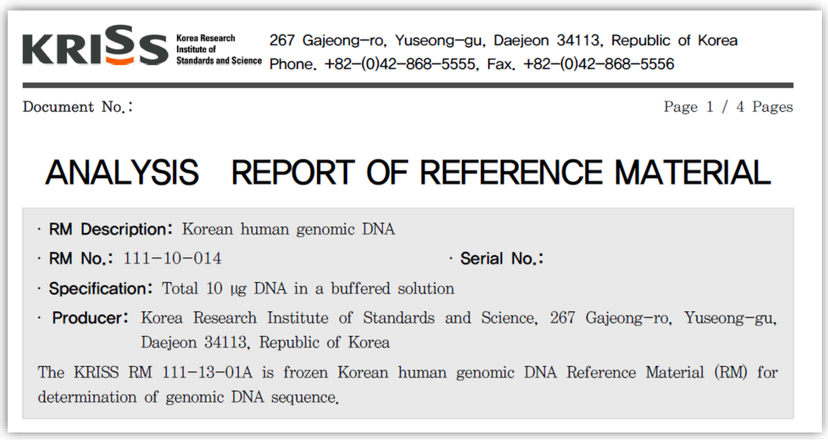


(2) To test the RM's homogeneity, we selected 107 single nucleotide variants (SNV) and small Indels through out the genome and verified their sequences compared to hg19.

→107 SNVs and Indels that AK1 has, compared to hg19, were selected
→ multiplex PCR
→ amplicon sequencing (targeted NGS sequencing)
→ library construction
→ 100 % homogeneity between ten selected vials

Documentation of the KRISS human genomic DNA reference material

Request for RM approval by technical supervisor → Review by the technical committee → Review by the KRISS quality assurance committee → Approval by the quality manager (President of KRISS)



KRISS Korea Research Institute of Standards and Science
267 Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea
Phone. +82-(0)42-868-5555, Fax. +82-(0)42-868-5556

Document No.: Page 1 / 4 Pages

ANALYSIS REPORT OF REFERENCE MATERIAL

- RM Description: Korean human genomic DNA
- RM No.: 111-10-014 · Serial No.:
- Specification: Total 10 μg DNA in a buffered solution
- Producer: Korea Research Institute of Standards and Science, 267 Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea

The KRISS RM 111-13-01A is frozen Korean human genomic DNA Reference Material (RM) for determination of genomic DNA sequence.

Acknowledgements

This study was supported by Korea Ministry of Trade, Industry and Energy and Korea Evaluation Institute of Industrial Technology, under the project name "Developing Korean Reference Genome (2014-2018)", led by Joohon Sung at SNU.