

Original Paper

# Next-Generation Sequencing–Based Cancer Panel Data Conversion Using International Standards to Implement a Clinical Next-Generation Sequencing Research System: Single-Institution Study

Phillip Park<sup>1\*</sup>, MS; Soo-Yong Shin<sup>2,3\*</sup>, PhD; Seog Yun Park<sup>4</sup>, MD; Jeonghee Yun<sup>4</sup>, ABT; Chulmin Shin<sup>1</sup>, MS; Jipmin Jung<sup>1</sup>, MS; Kui Son Choi<sup>1</sup>, PhD; Hyo Soung Cha<sup>1</sup>, PhD

<sup>1</sup>Cancer Data Center, National Cancer Center, Goyang, Republic of Korea

<sup>2</sup>Department of Digital Health, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, Republic of Korea

<sup>3</sup>Big Data Research Center, Samsung Medical Center, Seoul, Republic of Korea

<sup>4</sup>Department of Pathology, National Cancer Center, Goyang, Republic of Korea

\*these authors contributed equally

**Corresponding Author:**

Hyo Soung Cha, PhD  
Cancer Data Center  
National Cancer Center  
323 Ilsan-ro, Ilsandong-gu, Goyang  
Goyang, 10408  
Republic of Korea  
Phone: 82 1073502088  
Email: [kkido@ncc.re.kr](mailto:kkido@ncc.re.kr)

## Abstract

**Background:** The analytical capacity and speed of next-generation sequencing (NGS) technology have been improved. Many genetic variants associated with various diseases have been discovered using NGS. Therefore, applying NGS to clinical practice results in precision or personalized medicine. However, as clinical sequencing reports in electronic health records (EHRs) are not structured according to recommended standards, clinical decision support systems have not been fully utilized. In addition, integrating genomic data with clinical data for translational research remains a great challenge.

**Objective:** To apply international standards to clinical sequencing reports and to develop a clinical research information system to integrate standardized genomic data with clinical data.

**Methods:** We applied the recently published ISO/TS 20428 standard to 367 clinical sequencing reports generated by panel (91 genes) sequencing in EHRs and implemented a clinical NGS research system by extending the clinical data warehouse to integrate the necessary clinical data for each patient. We also developed a user interface with a clinical research portal and an NGS result viewer.

**Results:** A single clinical sequencing report with 28 items was restructured into four database tables and 49 entities. As a result, 367 patients' clinical sequencing data were connected with clinical data in EHRs, such as diagnosis, surgery, and death information. This system can support the development of cohort or case-control datasets as well.

**Conclusions:** The standardized clinical sequencing data are not only for clinical practice and could be further applied to translational research.

(*JMIR Med Inform* 2020;8(4):e14710) doi: [10.2196/14710](https://doi.org/10.2196/14710)

**KEYWORDS**

data standardization; clinical sequencing data; next-generation sequencing; translational research information system

## Introduction

Much research has been conducted to find new biological markers for diagnosis or treatment as next-generation sequencing (NGS) technologies have improved [1]. Recently, as the price and turn-around time of NGS have dramatically reduced, sequencing of patient samples using NGS has been applied in clinical practice [2]. For example, clinical sequencing was mainly applied in cancer patients to determine appropriate treatment by genotyping cancers [3]. Government agencies and private insurance companies in various countries have started to reimburse for clinical sequencing tests. For example, in the United States, if a sequencing laboratory is certified by Clinical Laboratory Improvement Amendments, the sequencing test could be reimbursed [4,5]. Similarly, based on the 100,000 Genomes Project, the National Health Service in the United Kingdom launched a service to provide access to the latest NGS technologies in genomic testing and management [6]. Further, the Korean National Insurance Agency started to reimburse for several panel sequencing tests, including those for cancer and rare diseases, in the beginning of March 2017 [7]. Much additional clinical sequencing has been performed worldwide in clinical practice.

Essentially, clinical sequencing results can be used for diagnosis or to identify appropriate treatment. However, since most of the current clinical sequencing results are not standardized, all clinical sequencing reports are stored in text or pdf format. Therefore, clinical decision support systems cannot utilize the clinical sequencing data through electronic health records (EHRs). In addition, clinical sequencing reports are not interoperable among hospitals owing to the lack of standard adoption. This means that extensive manual manipulation is required for interpretation or use of clinical sequencing reports. Based on the large amount of raw sequencing data and the complicated NGS pipeline from raw data to report generation, clinical sequencing requires well-established standard operating

procedures and highly-trained experts to ensure data quality [8]. To resolve this issue, diverse efforts have been made by standard development organizations. ISO/TC 215 focused on clinical genomics by establishing a subcommittee on genome informatics in 2019. It also published two genomics standards [9,10] and developed six genomics standards [11-16]. The HL7 clinical genomics working group also developed diverse clinical genomics standards [17-23].

Given the research problem mentioned above, this study aimed to develop a system to standardize clinical sequencing data and to provide services suitable for researchers to utilize the data. In this study, we extended a clinical NGS research system (CNRS) in a clinical research data warehouse (CRDW) that structures and standardizes clinical sequencing results by mapping standard terminology from current unstructured text reports.

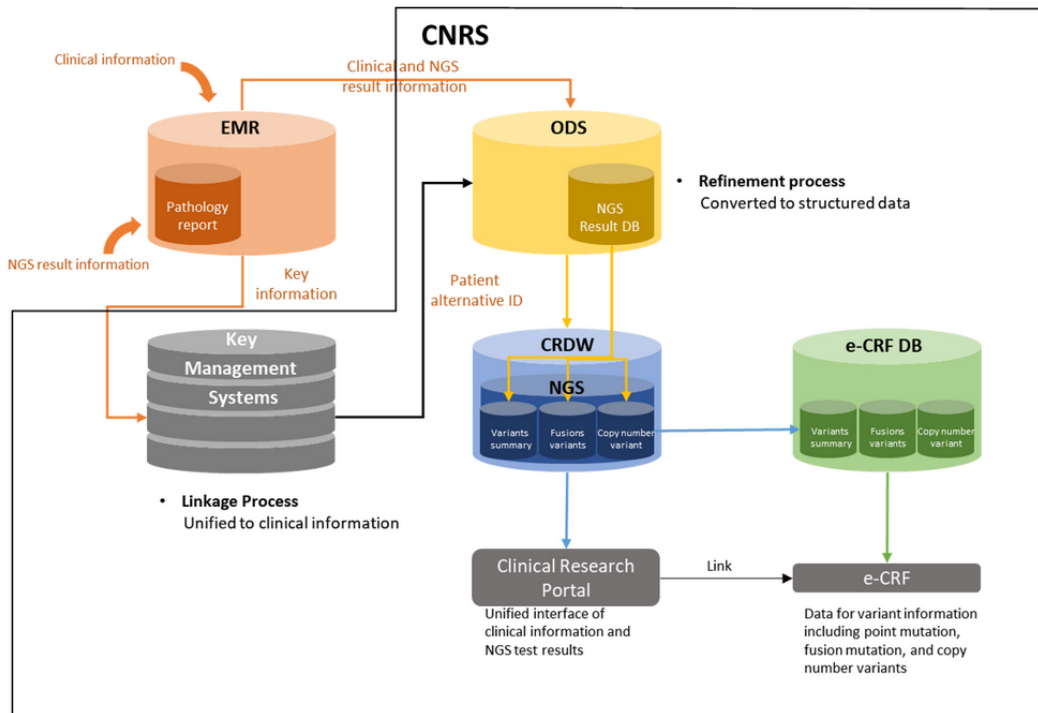
## Methods

### System Architecture and Data Flow

Figure 1 depicts the data flow of the CNRS. The clinical sequencing report is stored as a part of the pathology report in EHRs. The stored clinical sequencing reports are transferred to the operational data store (ODS). In ODS, the data are structured and standardized, and the sequencing information is saved in the CRDW. Through a key management server, the clinical data in the CRDW and the NGS result data in pathology reports are mapped to patient alternative numbers and not patient numbers. In addition, NGS result data in the CRDW are extracted and stored in the NGS viewer database (DB). Researchers can access the necessary deidentified specific genetic variation cohort in the clinical research portal and inspect NGS result data in the NGS result viewer.

The detailed information of each component is presented in subsequent sections.

**Figure 1.** Data flow of the clinical next-generation sequencing (NGS) research system. This system was built for the unified management of the clinical information of each patient and clinical NGS test results. CNRS: clinical next-generation sequencing research system; CRDW: clinical research data warehouse; DB: database; e-CRF: electronic-case report form; EMR: electronic medical record; ODS: operational data store.



**Data Collection of Clinical Sequencing Results**

The Health Insurance Review & Assessment Service in Korea reimburses laboratory-developed panel sequencing tests that include about 100 genes with 14 mandatory genes. As an

example, Table 1 shows the panel genes that are used in the National Cancer Center (NCC), Korea.

The current clinical sequencing report of the NCC is illustrated in Figure 2. Currently, clinical sequencing reports are stored in a single table with 28 attributes in the EHR DB. This table is copied to the ODS on a weekly basis.

**Table 1.** List of panel genes (n=91) used in the National Cancer Center, Korea.

Category	Genes
Mandatory genes (n=14)	<i>ALK, BRAF, BRCA1, BRCA2, EGFR, HER2, IDH1, IDH2, KIT, KRAS, MYC, MYCN, NRAS, and PDGFRA</i>
Additional genes (n=74)	<i>ABL1, AKT1, AKT3, APC, AR, ATM, AXL, CCND1, CDH1, CDK4, CDK6, CDKN2A, CEBPA, CSF1R, CTNNA1, DDR2, ERBB2, ERBB3, ERBB4, ERG, ESR1, ETV1, ETV4, ETV5, EZH2, FANCA, FANCC, FANCF, FANCG, FBXW7, FGFR1, FGFR2, FGFR3, FGFR4, FLT3, FOXL2, GNA11, GNAQ, GNAS, HNF1A, JAK1, JAK2, JAK3, KDR, MAP2K1, MAP2K2, MAP2K4, MET, MLH1, MTOR, NOTCH1, NPM1, NTRK1, NTRK2, NTRK3, PIK3CA, PIK3R1, PPARG, PTEN, PTPN11, RAF1, RB1, RET, ROS1, RUNX1, SMAD4, SMARCB1, SMO, SRC, STK11, TP53, VHL, WTI, and NRG1</i>
Additional fusion genes <sup>a</sup> (n=3)	<i>ALK, ROS1, and RET</i>

<sup>a</sup>Genes in the fusion category are duplicated in the mandatory gene list.

**Figure 2.** Input template of the clinical sequencing report of the National Cancer Center. All boxes are text boxes for free text entry.

NGS test result input\_[Pathology] – Wbest – Web page dialog box

Cancer genome test report ( N 160000038 )

• Main Sampling Site	<input style="width: 80%;" type="text"/>
• Specimen types	<input style="width: 80%;" type="text"/>
• Test method	<input style="width: 80%;" type="text"/>
• Analyzed gene	<input style="width: 80%;" type="text"/>
• Relevancy of sample	<input style="width: 80%;" type="text"/>

• Variants summary

Gene	Exon ID	DNA change	Protein change	Variant info	Allele Frequency	Exonic Effect	Clinical Effect

• Fusion variants

Gene	Fusion A	Fusion B	
Chromosome			
Cytoband			
Break			
Transcript Part			
Locus			
Gene strand			
#Span Read			
#Split Read			
#Total Read			
Distance			

• Diagnosis	<input style="width: 80%;" type="text"/>
• Analytical report	<input style="width: 80%;" type="text"/>
• References	<input style="width: 80%;" type="text"/>

**Structuring/Standardization Process**

Figure 3 demonstrates the data structuring process. All data are structured and standardized according to ISO/TS 20428 during the extract transform and load process [10]. The ISO/TS 20428 standard defines the required and optional fields for sequencing reports, along with the metadata for each field. The required fields include the following 10 categories: clinical sequencing orders, information on the subject of care, information on the

legally authorized person ordering clinical sequencing, performing laboratory, associated diseases and phenotypes, biomaterial information, genetic variations, classification of variants, recommended treatment, and addendum. The optional fields include the following seven categories: medical history, family history, reference genome version, racial genomic information, genetic variation, detailed sequencing information, and references.

**Figure 3.** Structure of the next-generation sequencing (NGS) test results. (A) Clinical NGS result summary of the electronic health record database (DB). (B) Variant summary table in the NGS DB. (C) Gene fusion table in the NGS DB. (D) Copy number variation table in the NGS DB.

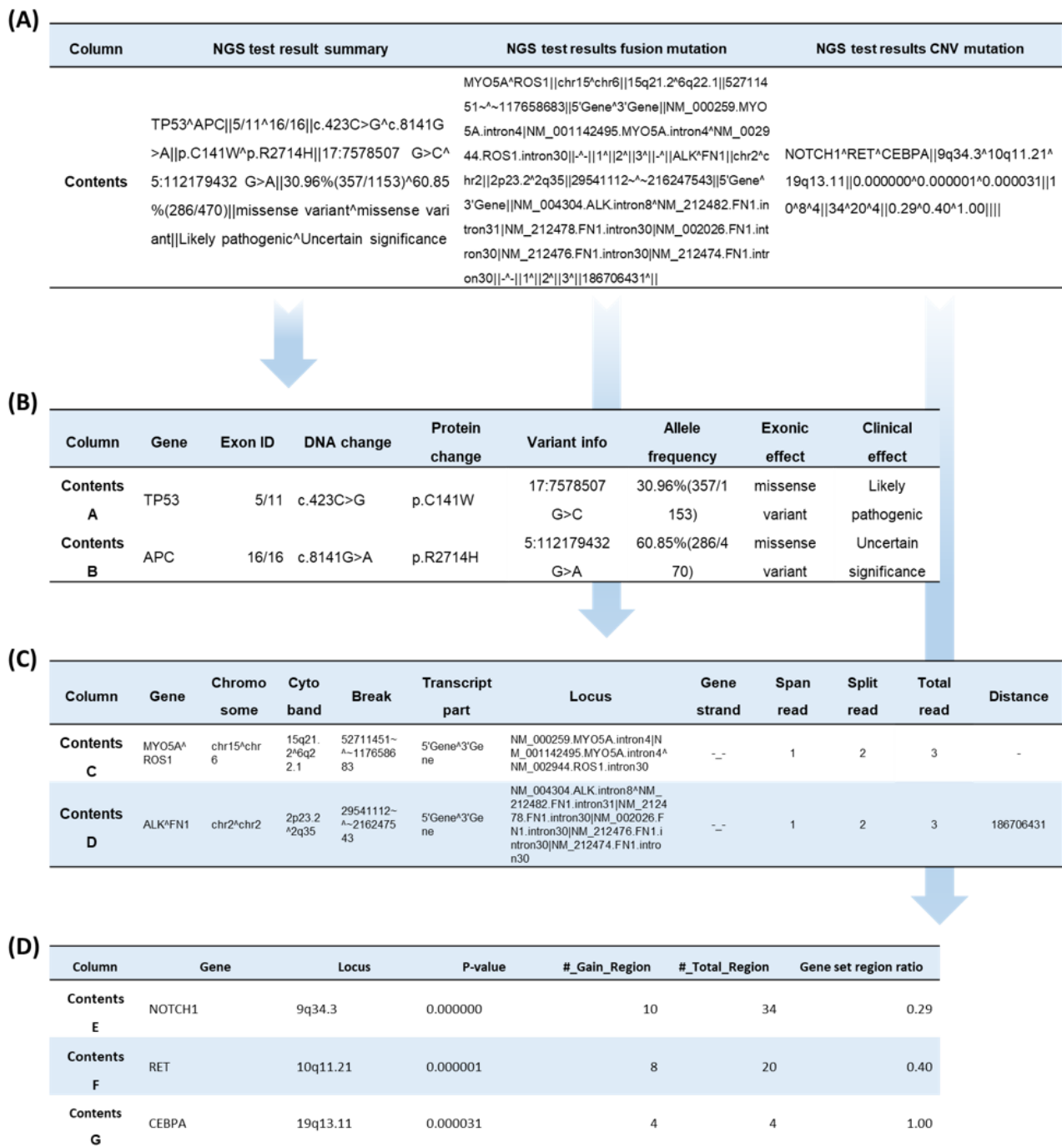


Figure 3 shows that the clinical sequencing data are structured and loaded from the replicated ODS into pathologic and laboratory information, variant summary, fusion variant, and copy number variation (CNV) variant tables. Pathologic and laboratory information has the following nine attributes: identifiers, test order date, quality control results, sample type, report generation date, report generator information, sequencer type, recommended treatment, and references. Variant summary has the following nine attributes: gene name, exon ID, DNA change, protein change, variant information, allele frequency, effects of variants, pathogeny, and clinical relevance. Fusion variant has the following 11 attributes: gene name, chromosome, cyto band, break, transcript part, locus, gene strand, span read,

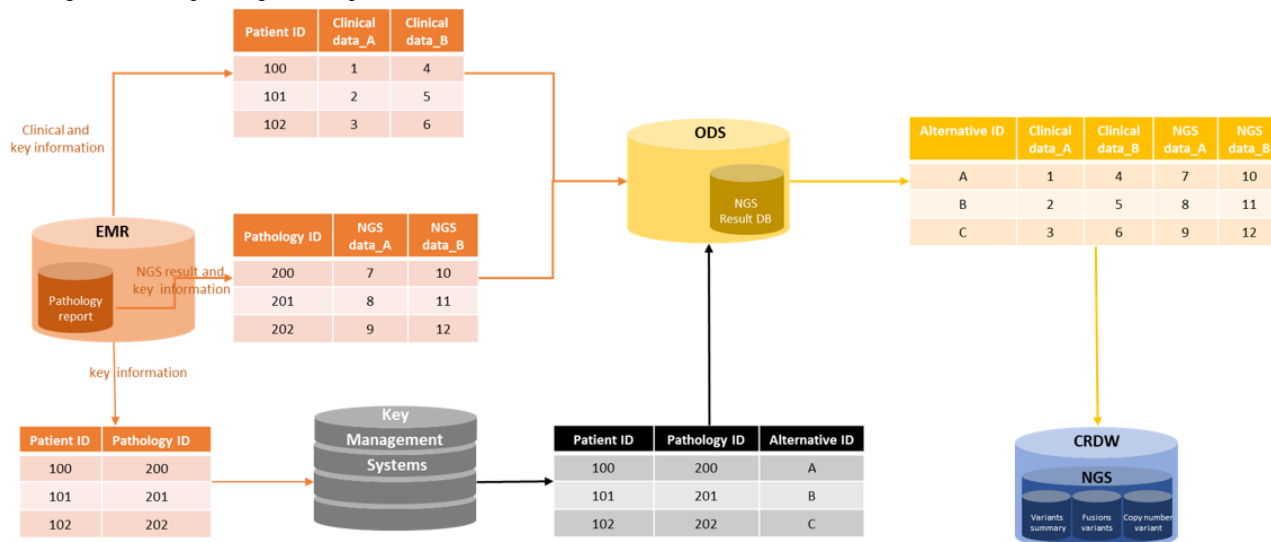
split read, total read, and distance. CNV has the following nine attributes: gene name, locus, P value, gain region, total region, region ratio, gene count, region count, and significant region count.

**Combination With Clinical Data**

Figure 4 shows that the NCC has deidentified the clinical data warehouse as well. The ODS receives clinical data from the table with the primary key as the patient ID and NGS result data from the table with the primary key as pathology ID from EHRs. The key management system receives the pathology ID and patient ID from EHRs and generates an alternative ID. It then sends the ID to the ODS. The ODS deletes the patient ID and

pathology ID and sends the data to the CRDW with the alternative ID. By using an alternative ID, which is a pseudonym for the patient ID, the necessary deidentified clinical data can be combined with the sequencing result data.

**Figure 4.** Overview of the combination with clinical data. CRDW: clinical research data warehouse; DB: database; EMR: electronic medical record; NGS: next-generation sequencing; ODS: operational data store.



### User Services

Users can access the necessary information using a clinical research portal and NGS result viewer. The clinical research portal is a user interface to query or extract clinical and genomic data by changing search options. The NGS result viewer provides functionality to create a structured or standardized clinical sequencing report by converting an original unstructured pathology report using ISO/TS 20428.

This study was approved by the institutional review board (IRB) of the NCC in Korea (NCC2019-0535).

### Results

From April 2017 to February 2019, the CNRS included 367 clinical sequencing results, which consisted of 249 lung cancer cases, 70 ovarian cancer cases, eight breast cancer cases, seven

malignant melanoma cases, seven colon cancer cases, seven stomach cancer cases, six liver cancer cases, five thyroid cancer cases, five kidney cancer cases, two brain cancer cases, and one prostate cancer case. In detail, 51 variants were found and stored among a total of 88 genes. Figure 5 shows the distribution of point mutations by cancer type. Across all cancer types, *TP53* (167/367, 45.5%), *EGRF* (56/367, 15.3%), *KRAS* (34/337, 9.3%), and *BRAC1* (21/337, 5.7%) mutations were common. According to each cancer type, *TP53* (120/249, 48.2%), *EGFR* (51/249, 20.5%), and *KRAS* (27/249, 10.8%) mutations were common in lung cancer; *TP53* (39/70, 55.7%), *BRCA1* (9/70, 12.9%), and *PIK3CA* (5/70, 7.1%) mutations were common in ovarian cancer; and *TP53* (3/8, 37.5%), *BRCA2* (3/8, 37.5%), and *PIK3CA* (3/8, 37.5%) mutations were common in breast cancer. The rates of pathogenic, likely pathogenic, and uncertain significance variants were 55.1%, 34.7%, and 42.9%, respectively.

**Figure 5.** Distribution of point mutations by cancer type in 367 patients. The top 12 genes by frequency are displayed.

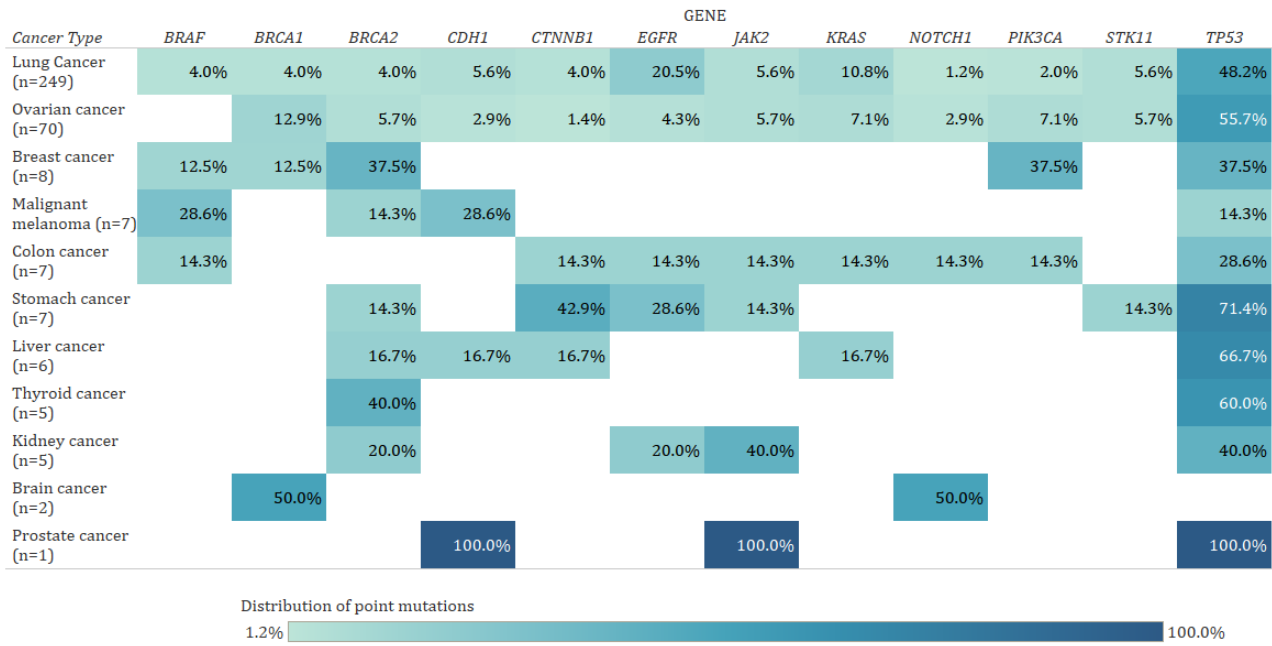


Figure 6 presents the user interface of the clinical research portal. As explained in the previous section, the clinical research portal supports an integrated view of the NGS results and the corresponding clinical data in EHRs. Figure 6 shows the category of the CRDW, which contains clinical data, such as diagnosis, laboratory data, medication, surgery, chemotherapy,

follow-up data, and patient demographic data. Users can choose the appropriate category and then choose desired detailed variables by clicking on them. The NGS results can be visualized (ie, variant summary, CNV, and fusion genes), as illustrated in Figure 7.

**Figure 6.** Interface of the clinical research search portal. The main page of the clinical research search portal comprises two domains. The red rectangle indicates searchable items. The blue rectangle indicates the area where the researcher can select items through drag and drop.

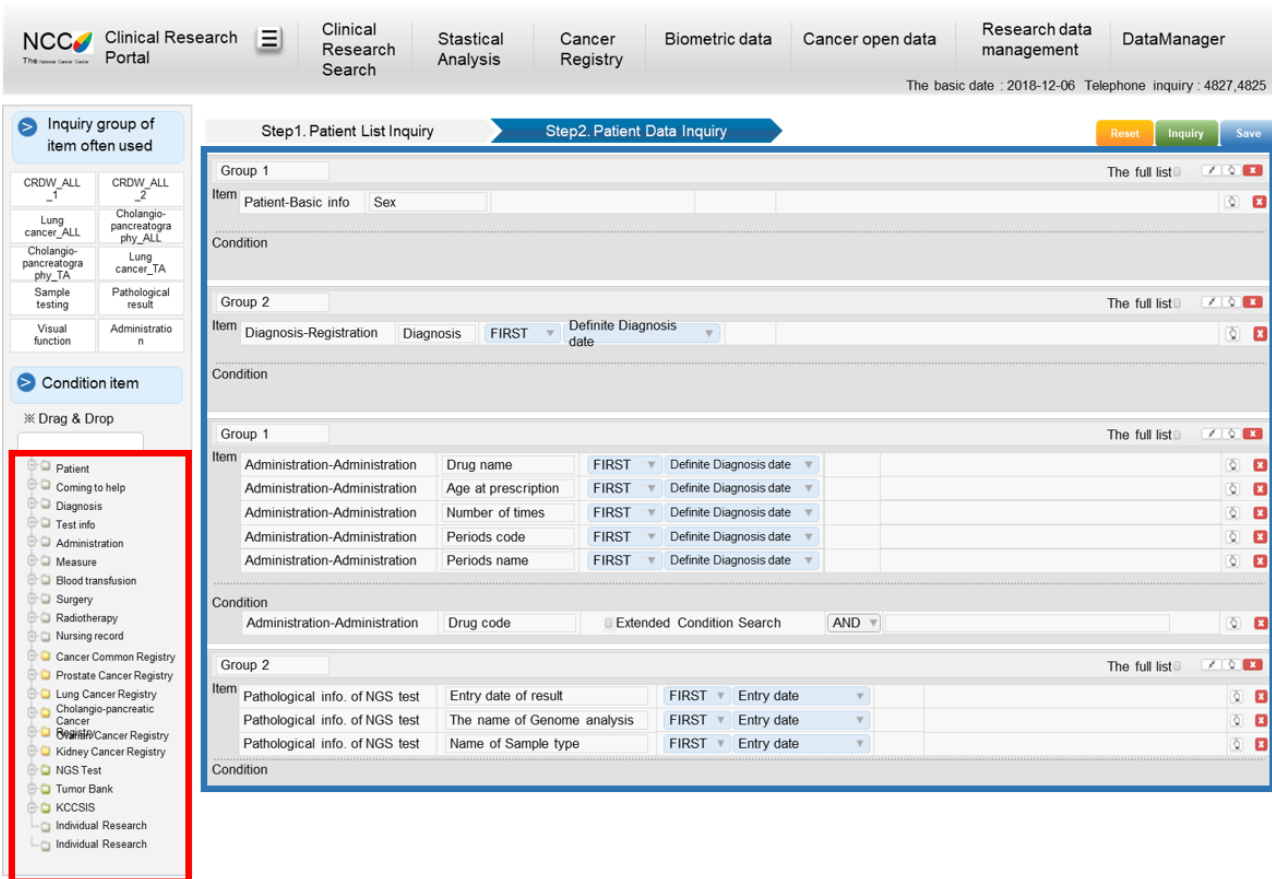


Figure 7. Example of next-generation sequencing results in the clinical research portal.

Data sheet									
No.	Pseudonymized ID	Group 1		Group 2	Group 3	Group 4			
		Sex	Birth	Drug Name	Sample Type	Gene	DNA Change	Protein Change	Exonic Change
1	RN00	F	1958-04-02		Formalin-fixed-paraffin-em...				
2	RN00	M	1950-09-20	Taxol 30mg inj (Paclitaxel)	Formalin-fixed-paraffin-em...				
3	RN01	M	1950-02-05		Formalin-fixed-paraffin-em...				
4	RN02	M	1955-06-01	Carboplatin 450mg (원상17...	Formalin-fixed-paraffin-em...				
5	RN03	M	1948-11-01		Formalin-fixed-paraffin-em...				
6	RN05	F	1942-11-20		Formalin-fixed-paraffin-em...				
7	RN05	F	1942-09-09		Formalin-fixed-paraffin-em...				
8	RN06	M	1950-04-23		Formalin-fixed-paraffin-em...				
9	RN06	M	1952-01-03		Formalin-fixed-paraffin-em...				
10	RN07	M	1949-01-29	Taxol 30mg inj (Paclitaxel)	Formalin-fixed-paraffin-em...				
11	RN07	M	1943-08-16	Taxol 30mg inj (Paclitaxel)	Formalin-fixed-paraffin-em...				
12	RN07	M	1969-01-27	Iressa 250mg tab(Gefitinib)	Formalin-fixed-paraffin-em...				
13	RN08	M	1934-03-09	Iressa 250mg tab(Gefitinib)	Formalin-fixed-paraffin-em...	EGFR	c.2236_2250delGAATTAAGAGAAGCA	p.E746_A750del	conservative inframe deletion
14	RN10	M	1948-03-08	Carboplatin 150mg(원상152...	Formalin-fixed-paraffin-em...				
15	RN11	M	1965-09-14	Erlotinib 150mg (원상16107)	Formalin-fixed-paraffin-em...	EGFR	c.2573T>G	p.L858R	missense variant
16	RN12	M	1953-08-02		Formalin-fixed-paraffin-em...				
17	RN12	M	1949-05-03	Erlotinib 150mg (원상16107)	Formalin-fixed-paraffin-em...	EGFR	c.2573T>G	p.L858R	missense variant
18	RN13	M	1948-03-03	Taxol 30mg inj (Paclitaxel)	Formalin-fixed-paraffin-em...				
19	RN13	M	1951-02-28		Formalin-fixed-paraffin-em...				
20	RN13	F	1960-01-07		Formalin-fixed-paraffin-em...				
21	RN14	F	1948-04-25	Taxol 30mg inj (Paclitaxel)	Formalin-fixed-paraffin-em...				
22	RN16	M	1949-02-28		Formalin-fixed-paraffin-em...				
23	RN16	M	1959-02-25	Erlotinib 150mg (원상16107)	Formalin-fixed-paraffin-em...	EGFR	c.2573T>G	p.L858R	missense variant
24	RN18	M	1953-05-30		Formalin-fixed-paraffin-em...				
25	RN19	M	1961-08-20		Formalin-fixed-paraffin-em...				
26	RN19	F	1961-06-18	Tarceva 100mg tab(Erlotinib)	Formalin-fixed-paraffin-em...	EGFR	c.2235_2249delGGAATTAAGAGAAGC	p.E746_A750del	disruptive inframe deletion
27	RN19	M	1958-06-04	Taxol 30mg inj (Paclitaxel)	Formalin-fixed-paraffin-em...				
28	RN21	M	1951-12-16	Taxol 30mg inj (Paclitaxel)	Formalin-fixed-paraffin-em...				
29	RN25	F	1954-03-20		Formalin-fixed-paraffin-em...				

The NGS result viewer can show the detailed clinical sequencing report of each patient in a structured way, whereas the clinical research portal supports the analysis of aggregated sequencing results. As shown in Figure 8, the clinical sequencing report is

mainly divided into the following three parts: basic test information, sequencing methods and other related information, and variants with reporting results.



**Figure 8.** Example of the next-generation sequencing (NGS) result viewer. The main page of the NGS result viewer is composed of three domains. The first box provides basic test information. The second box explains sequencing methods and other related information. The last box shows mutation data with reporting results.

Institute	NCC	Alteration Number of patient	RN00141***	Name of patient	Park**						
Random number		Status		Date of written consent							
Study schedule	NGS		Visit date	Park**							
CRF name	NGS result information – Version : ver 1.0										
Registration date	2018-04-05 15:44 / CRDW		Modification date	2018-11-28 17:01 / CRDW							
• Pathologic and laboratory information – Version : ver 1.0											
Pathological number		Test order date	2017-10-18								
Information of report generator	***	Report generation date	2017-10-26								
Type of sample	Formalin-fixed-paraffin-embedded tissue		Genetic analysis name	NGS Pan cancer panel 91 genes, version 0 (NGS PCP 91 ver. 0)							
Type of sequencer	Next generation sequencing(NGS) method for PCP		QC results	DNA QC & Library QC: Pass							
Diagnosis details	No clinically significant variants is detected by Next generation sequencing method.		Instrument name	Illumina® MiSeqDx							
• Point Mutation											
	Gene	Exon ID	DNA change	Protein change	Variant info	Allele Frequency	Exonic effect	Clinical effect	MUTATION_YN		
7	ATM	4/63	c.251C>T	p.A84V	11:108099970 C>T	41.80% (130/311)	Missense	Uncertain significance	Y		
• Fusion Mutation											
	Gene	Chromosome	Gene position	Genetic divergence	Transcription site	Gene locus	DNA strand	Span read	Split read	Total read	Distance
• Copy Number Variant											
	Gene	Gene locus	P value	Gain region	Total region	Region ratio	Gene count	Region count	Significant region count		
NGS Inspection Analytical report content		No clinically meaningful mutations were observed. This test does not simultaneously perform an analysis of extracorporeal tissue (e.g., normal tissue around blood of tumor) using NGS assay to identify somatic cell variants of tumors, thus the derived sequencing mutation cannot exclude the possibility of reproductive cell of mosaic mutation. In addition, there is a possibility of untested genetic variation if somatic variation is not detected within 1% and there are a number of pathogenic variants.									

## Discussion

### Principal Findings and Implications

The CNRS converts text-based clinical sequencing reports in EHRs into structured data using international standards. Through the CNRS, the content, as shown in Figure 5, can be easily organized and data can be managed according to international standards. It also provides standardized data through the clinical research search portal; furthermore, using its functions, researchers can set up cohorts with specific mutations and add clinical data columns as needed. Thus, it can be inferred that the CNRS supports researchers in performing translational research by allowing them to easily extract the desired clinical and genomic data from EHRs. For example, non-small cell lung cancer genotyping requires mutation pattern analysis of *KRAS*, *EGFR*, and *BRAF* [1,24]. Similarly, other research requires clinical data such as that on cancer stage, smoking history, and death date for survival analysis [25]. These types of translational studies can be easily performed using the developed CNRS, and this has already been proven by researchers at the NCC.

The CNRS is provided to research projects that have been approved by the IRB. Researchers send data extraction requests to health information managers, and it takes about one to two

weeks to review, refine, and provide clinical data from EHRs. The CNRS could retrieve data through the clinical research portal after receiving IRB approval, and it takes about two or three days from review to delivery after a data extraction request is made. In addition, the NCC has built a genomic cohort linking the NGS DB with cancer registries such as those of lung cancer and ovarian cancer.

There are two unique aspects of the CNRS as developed. One is that the CNRS uses an ISO standard (ISO/TS 20428) to support multicenter research. If other hospitals or research institutes use international standards, the data can be easily integrated into the same format. We also achieved the same results as successfully converting the data and manually cleansing the data previously. The other is that the CNRS can help protect patients' privacy by deidentifying protected health information. To use the patients' data for research purposes, researchers must obtain written consent from patients or deidentify the identifiable data. In this era of big data, deidentification is usually used for a number of reasons. However, if we deidentify the data, it is difficult to link the separate DBs. To overcome this issue, the CNRS adopted a key management server to pseudonymize the patient ID. This means that the CNRS works as an honest broker [26]. To strengthen the protection level, only authorized developers can access this key management server and users can receive the randomly

assigned ID after combing clinical data and sequencing data using the key management system. Therefore, users cannot retrieve real patient IDs in EHRs or pseudonymized IDs in the key management system.

### Limitations

The main contribution of this study is that, for the first time, the ISO/TS 20428 standard was applied to the CRDW to standardize clinical genomic test results. As a result, we also demonstrated that this approach could enable easy search and analysis with clinical data in EHRs. However, it has limitations. We did not verify this system in multiple centers. We hope our approach will help other hospitals or institutions build their own systems.

### Future Work

Our continuing objectives are to extend the categories of clinical and sequencing data in the CNRS and consider the standards proposed by the Global Alliance for Genomics and Health, which has developed diverse practical standard application programming interfaces for international genomic research.

### Conclusion

The CNRS converts the text-based clinical sequencing reports of EHRs into structured data using international standards and provides standardized data. In addition, the CNRS allows researchers to set up cohorts with specific mutations and add clinical data columns as needed. Therefore, it can be inferred that the CNRS enables researchers to conduct translational research by allowing them to easily extract the required clinical and genomic data from EHRs.

### Acknowledgments

This study was supported by a grant from the National Cancer Center (grant no. 1810871-2), the National R&D Program for Cancer Control funded by the Ministry of Health and Welfare, Republic of Korea (grant no. 1631180), and the Technology Innovation Program funded by the Ministry of Trade, Industry & Energy, Republic of Korea (grant no. 20004632).

### Conflicts of Interest

None declared.

### References

1. Lim S, Tan S, Lim W, Lim C. An extracellular matrix-related prognostic and predictive indicator for early-stage non-small cell lung cancer. *Nat Commun* 2017 Nov 23;8(1):1734 [FREE Full text] [doi: [10.1038/s41467-017-01430-6](https://doi.org/10.1038/s41467-017-01430-6)] [Medline: [29170406](https://pubmed.ncbi.nlm.nih.gov/29170406/)]
2. Simon R, Roychowdhury S. Implementing personalized cancer genomics in clinical trials. *Nat Rev Drug Discov* 2013 May;12(5):358-369. [doi: [10.1038/nrd3979](https://doi.org/10.1038/nrd3979)] [Medline: [23629504](https://pubmed.ncbi.nlm.nih.gov/23629504/)]
3. Katsanis S, Katsanis N. Molecular genetic testing and the future of clinical genomics. *Nat Rev Genet* 2013 Jun;14(6):415-426 [FREE Full text] [doi: [10.1038/nrg3493](https://doi.org/10.1038/nrg3493)] [Medline: [23681062](https://pubmed.ncbi.nlm.nih.gov/23681062/)]
4. Mauer C, Pirzadeh-Miller S, Robinson L, Euhus D. The integration of next-generation sequencing panels in the clinical cancer genetics practice: an institutional experience. *Genet Med* 2014 May;16(5):407-412. [doi: [10.1038/gim.2013.160](https://doi.org/10.1038/gim.2013.160)] [Medline: [24113346](https://pubmed.ncbi.nlm.nih.gov/24113346/)]
5. Hehir-Kwa J, Claustres M, Hastings R, van Ravenswaaij-Arts C, Christenhusz G, Genuardi M, et al. Towards a European consensus for reporting incidental findings during clinical NGS testing. *Eur J Hum Genet* 2015 Dec;23(12):1601-1606 [FREE Full text] [doi: [10.1038/ejhg.2015.111](https://doi.org/10.1038/ejhg.2015.111)] [Medline: [26036857](https://pubmed.ncbi.nlm.nih.gov/26036857/)]
6. Ratner M. Next-generation sequencing tests to become routine. *Nat Biotechnol* 2018 Jun 06;36(6):484. [doi: [10.1038/nbt0618-484](https://doi.org/10.1038/nbt0618-484)] [Medline: [29874203](https://pubmed.ncbi.nlm.nih.gov/29874203/)]
7. Lee D. MS26.01 Translation of Clinical Data to Real World - Asia. *Journal of Thoracic Oncology* 2018 Oct;13(10):S296 [FREE Full text] [doi: [10.1016/j.jtho.2018.08.189](https://doi.org/10.1016/j.jtho.2018.08.189)]
8. Park Y, Shin S. Status and Direction of Healthcare Data in Korea for Artificial Intelligence. *Hanyang Med Rev* 2017;37(2):86-92 [FREE Full text] [doi: [10.7599/hmr.2017.37.2.86](https://doi.org/10.7599/hmr.2017.37.2.86)]
9. ISO. ISO 25720:2009 Genomic Sequence Variation Markup Language (GSVML) URL: <https://www.iso.org/standard/43182.html> [accessed 2020-03-10]
10. ISO. ISO/TS 20428:2017 Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records URL: <https://www.iso.org/standard/67981.html> [accessed 2020-03-10]
11. ISO. ISO/CD TS 22692.2 Genomics Informatics— Quality control metrics for DNA sequencing URL: <https://www.iso.org/standard/73693.html> [accessed 2020-03-10]
12. ISO. ISO/DIS 21393 Genomics Informatics — Omics Markup Language (OML) URL: <https://www.iso.org/standard/70855.html> [accessed 2020-03-10]
13. ISO. ISO/WD TS 23357 Clinical genomics data sharing specification for next generation sequencing URL: <https://www.iso.org/standard/75310.html> [accessed 2020-03-10]
14. ISO. ISO/CD TR 21394.2 Genomics Informatics — Whole Genomics Sequence Markup Language (WGML) URL: <https://www.iso.org/standard/75956.html> [accessed 2020-03-10]

15. ISO. ISO/WD TS 22693 Health Informatics — Structured clinical gene fusion report in electronic health records URL: <https://www.iso.org/standard/73694.html> [accessed 2020-03-10]
16. ISO. ISO/WD TS 22690 Health Informatics — Reliability assessment criteria for high-throughput gene-expression data URL: <https://www.iso.org/standard/73691.html> [accessed 2020-03-10]
17. Health Level Seven International. HL7 Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-Qualified Genetic Variation Model (US Realm) URL: [https://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=23](https://www.hl7.org/implement/standards/product_brief.cfm?product_id=23) [accessed 2020-03-10]
18. Health Level Seven International. HL7 Version 3 Implementation Guide: Family History/Pedigree Interoperability, Release 1 URL: [https://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=301](https://www.hl7.org/implement/standards/product_brief.cfm?product_id=301) [accessed 2020-03-10]
19. Health Level Seven International. HL7 Version 2 Implementation Guide: Clinical Genomics; fully LOINC-Qualified Cytogenetic Model, Release 1 - US Realm URL: [https://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=364](https://www.hl7.org/implement/standards/product_brief.cfm?product_id=364) [accessed 2020-03-10]
20. Health Level Seven International. HL7 Implementation Guide for CDA® Release 2: Genetic Testing Reports, Release 1 URL: [https://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=292](https://www.hl7.org/implement/standards/product_brief.cfm?product_id=292) [accessed 2020-03-10]
21. Health Level Seven International. HL7 Domain Analysis Model: Clinical Sequencing, Release 1 URL: [https://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=446](https://www.hl7.org/implement/standards/product_brief.cfm?product_id=446) [accessed 2020-03-10]
22. Health Level Seven International. HL7 Version 3 Standard: Clinical Genomics; Pedigree, Release 1 URL: [https://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=8](https://www.hl7.org/implement/standards/product_brief.cfm?product_id=8) [accessed 2020-03-10]
23. Health Level Seven International. HL7 Domain Analysis Model: Clinical Genomics, Release1 URL: [https://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=479](https://www.hl7.org/implement/standards/product_brief.cfm?product_id=479) [accessed 2020-03-10]
24. Hagemann IS, Devarakonda S, Lockwood CM, Spencer DH, Guebert K, Bredemeyer AJ, et al. Clinical next-generation sequencing in patients with non-small cell lung cancer. *Cancer* 2015 Feb 15;121(4):631-639. [doi: [10.1002/cncr.29089](https://doi.org/10.1002/cncr.29089)] [Medline: [25345567](https://pubmed.ncbi.nlm.nih.gov/25345567/)]
25. Rossi D, Rasi S, Fabbri G, Spina V, Fangazio M, Forconi F, et al. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood* 2012 Jan 12;119(2):521-529 [FREE Full text] [doi: [10.1182/blood-2011-09-379966](https://doi.org/10.1182/blood-2011-09-379966)] [Medline: [22077063](https://pubmed.ncbi.nlm.nih.gov/22077063/)]
26. Choi H, Lee M, Choi C, Lee J, Shin S, Lyu Y, et al. Establishing the role of honest broker: bridging the gap between protecting personal health data and clinical research efficiency. *Peer J* 2015;3:e1506 [FREE Full text] [doi: [10.7717/peerj.1506](https://doi.org/10.7717/peerj.1506)] [Medline: [26713253](https://pubmed.ncbi.nlm.nih.gov/26713253/)]

## Abbreviations

**CNRS:** clinical next-generation sequencing research system

**CNV:** copy number variation

**CRDW:** clinical research data warehouse

**DB:** database

**EHR:** electronic health record

**IRB:** institutional review board

**NCC:** National Cancer Center

**NGS:** next-generation sequencing

**ODS:** operational data store

*Edited by G Eysenbach; submitted 15.05.19; peer-reviewed by YR Park, K Pradeep; comments to author 03.10.19; revised version received 19.11.19; accepted 07.02.20; published 24.04.20*

*Please cite as:*

*Park P, Shin SY, Park SY, Yun J, Shin C, Jung J, Choi KS, Cha HS*

*Next-Generation Sequencing-Based Cancer Panel Data Conversion Using International Standards to Implement a Clinical Next-Generation Sequencing Research System: Single-Institution Study*

*JMIR Med Inform* 2020;8(4):e14710

URL: <http://medinform.jmir.org/2020/4/e14710/>

doi: [10.2196/14710](https://doi.org/10.2196/14710)

PMID: [32329738](https://pubmed.ncbi.nlm.nih.gov/32329738/)

©Phillip Park, Soo-Yong Shin, Seog Yun Park, Jeonghee Yun, Chulmin Shin, Jipmin Jung, Kui Son Choi, Hyo Soung Cha. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org>), 24.04.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which

permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.