iGAPoncloud

ALLinONE

All-In-One service

iGAPonCloud

Introduction of ichrogene

- ichrogene is a global top-genomic analysis company that creates value to realize precision medicine using genome big data.
- We provide accurately predicting services for chronic diseases, cancer, personal characteristics, etc. through genomic big data analysis of more than 600,000 people. Top experts in the fields of genomics, IT, and biotechnology develop reliable services.
- ichrogene also focuses on research and development to allow all customers including individuals, researchers and medical staffs to use genetic information, the most important basis of healthcare services, conveniently without any difficulty.

iDNADoctor

Personal genetic analysis service (DTC, hospitals and clinics)

iSearchme for genome

Global personal genetic information analysis service platform

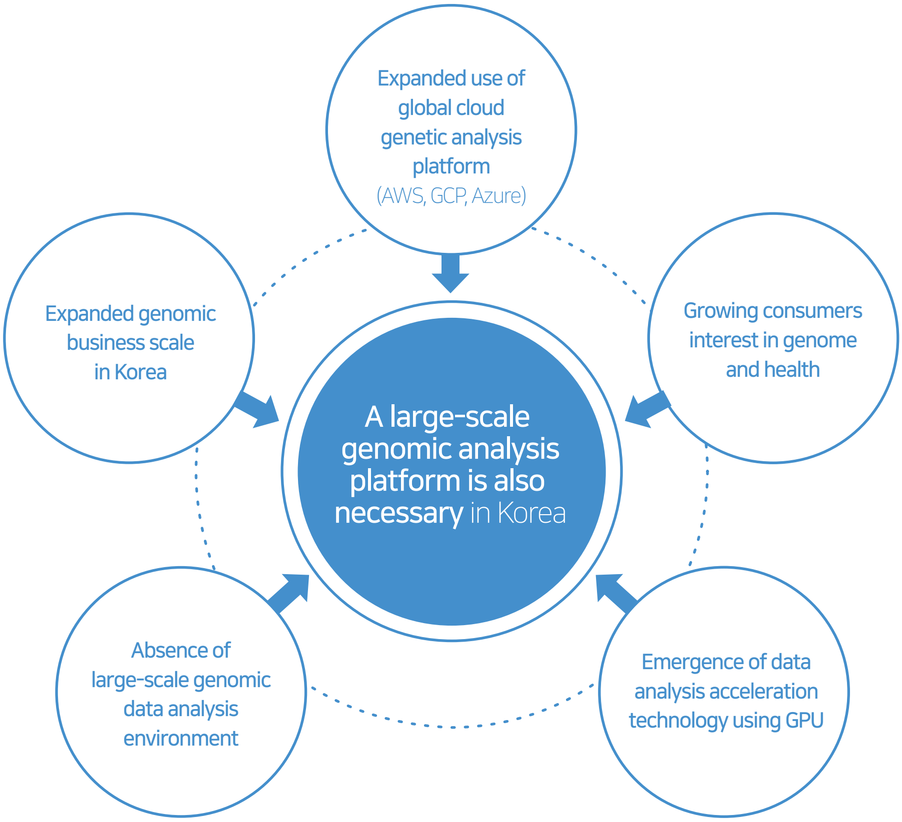
iCareMe

MyData-based customized healthcare service platform

iDNADoctor

Personalized genetic analysis services specializing in chronic diseases (hospitals and clinics)

Needs of a genomic analysis platform



Expanded use of global cloud genetic analysis platform (AWS, GCP, Azure)

Growing consumers interest in genome and health

Emergence of data analysis acceleration technology using GPU

Absence of large-scale genomic data analysis environment

Expanded genomic business scale in Korea

A large-scale genomic analysis platform is also necessary in Korea

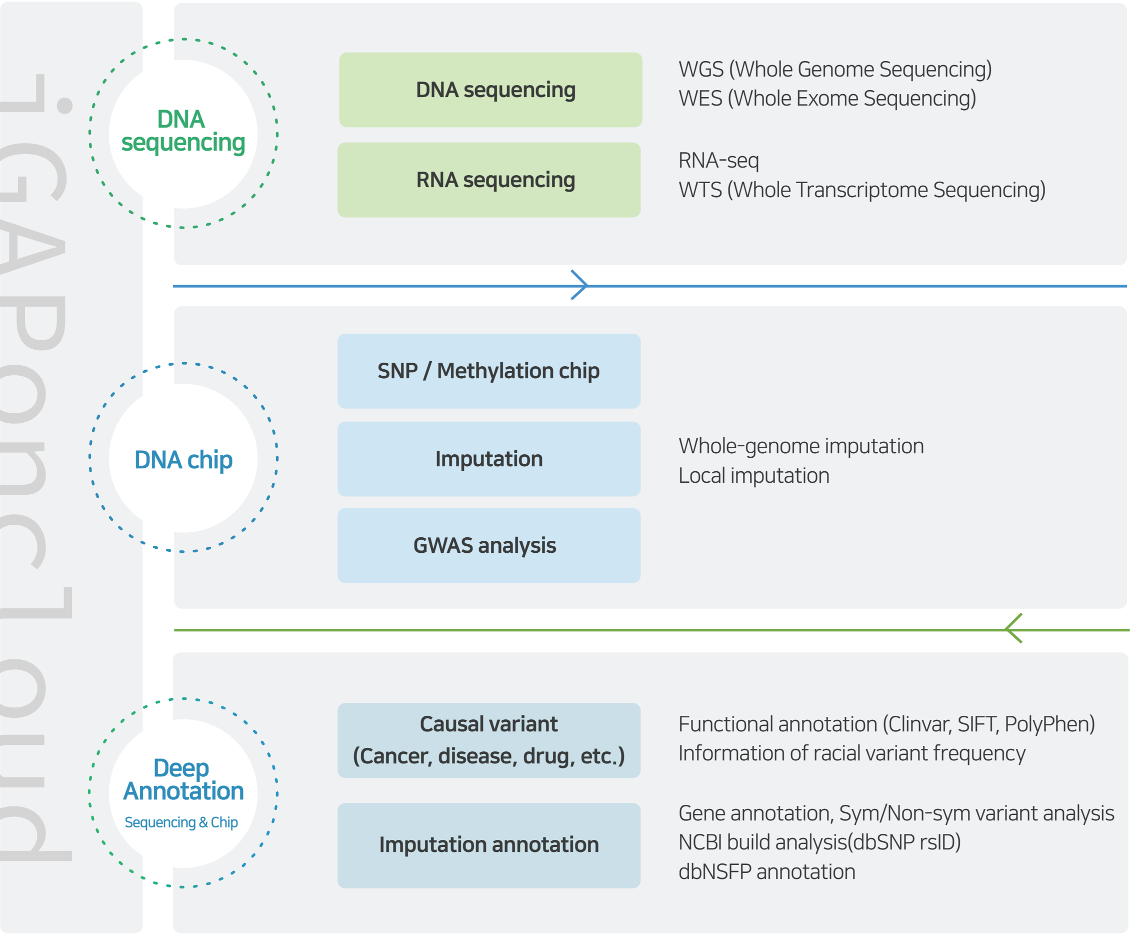
All-In-One service

iGAPonCloud

What is iGAPonCloud?

- iGAPonCloud is an user-friendly integrated genomic data analysis platform that combines the computing power of NAVER cloud plarform, NVIDIA's Clara Parabricks high-speed genomic analysis pipeline, and Ichrogene's genomic analysis solutions.

iGAPonCloud utilization areas



DNA sequencing

- **DNA sequencing**: WGS (Whole Genome Sequencing)
WES (Whole Exome Sequencing)
- **RNA sequencing**: RNA-seq
WTS (Whole Transcriptome Sequencing)

DNA chip

- **SNP / Methylation chip**
- **Imputation**: Whole-genome imputation
Local imputation
- **GWAS analysis**

Deep Annotation
Sequencing & Chip

- **Causal variant (Cancer, disease, drug, etc.)**: Functional annotation (Clinvar, SIFT, PolyPhen)
Information of racial variant frequency
- **Imputation annotation**: Gene annotation, Sym/Non-sym variant analysis
NCBI build analysis(dbSNP rsID)
dbNSFP annotation

Automated Worflow for 3 steps

Utilization of customized analysis results by phase

- We devide the variant identification process into an automated three-step process to generate and provide variant information and clinical reports from raw datas (raw files) to customers.

Step 1
Variant identification

- BAM file generation - Analysis ready files for subanalysis
- VCF file generation - Support of single (individual) vcf and multi-sampled vcf
- Quality control (QC report)

FASTQ
∨
VCF

Step 2
Variant annotation report

- Variant frequency DB (gnomAD, ExAC)
- Variant function DB (SIFT, CADD, etc.)
- Disease-related variant DB (Clinvar, ACMG)

* Various deep annotations

VCF
∨
Clinical report

Step 3
Identification of diseas-causing variant

- Rare variant (Rare disease)
- PRS (Polygenic Risk Score) analysis
- GWAS (Genome-Wide Association Study) analysis

* Integrated analysis with clinical information through correlation analysis

VCF
∨
Association

Target customers

- Genome data production companies, Digital health care companies, Pharmaceutical companies, general researchers in diagnostic test institutions universities and hospitals

Fast and User-friendly Service

Convenient use through user-friendly UI

기본정보

Username	kyh
Analysis date	분석의뢰 등록 시각(자동생성)
Run identifier	분석식별자(자동생성)

분석 설정

Sequencing library type	○ Single End sequencing	○ Paired End sequencing		
NGS Platform	○ Whole Exome sequencing	○ Whole Genome sequencing	○ Targeted sequencing(Panel)	
Analysis range	○ Whole genome	○ Region of interest(required target region import)		
Input files import	/yourDirectoryPath/	ex) /your/directory/here/		
Target region import				
Genome build	○ hg19	○ hg38	○ GRCh37	○ GRCh38
BAM output	○ Yes	○ No		
Output format	○ vcf(variant call format)	○ g.vcf(genomic vcf)	○ compressed	

01 Since it is possible to use NCP's resources as long as needed, you can reduce the licensing cost for NVIDIA Clara Parabricks.

02 Analysis is possible regardless of time and place using the cloud environment.

03 The server image where reference genomes and programs for analysis have been prepared in advance is provided so that analysis can be carried out immediately.

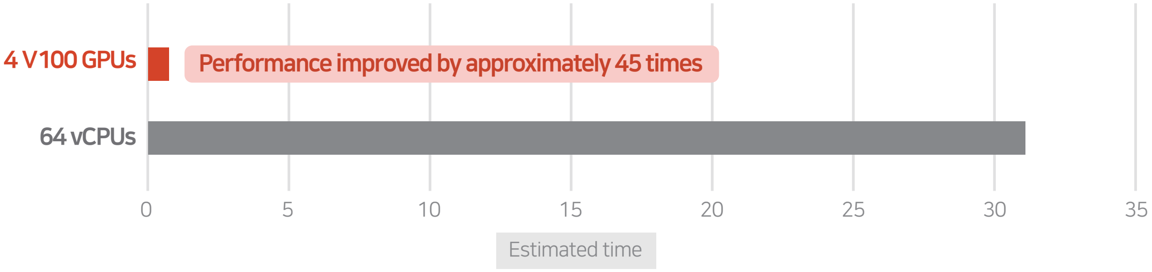
04 Unlike other companies' services that require input of multiple commands into the Linux console, users who are not familiar with the Linux environment also can carry out analysis using the UI.

Significantly reduced analysis time compared to previous analysis solutions

- The time required for genome data analysis has been significantly reduced by introducing Clara Parabricks from NVIDIA, an accelerated computing genome application framework using GPU.

- We saved computing costs by up to 50% by reducing the time by 45 times compared to the time required for previous CPU core-based analyses. (Reduced from approximately 31 hours to 40 minutes)

- Approximately up to 40 whole-genome data (based on WGS, 30X) can be processed per day.



*Source: NVIDIA Clara Parabricks product sheet

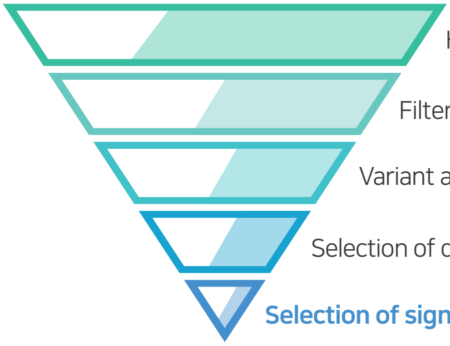
Clinical Use

Provision of deep annotation information for identifying the cause of disease

- We present a guide to the interpretation of cause variation by providing deep annotation information for variant using the VEP program developed and verified by DDD (Deciphering Developmental Disorders), a research consortium for patients with developmental disabilities in the UK.

- We improved accuracy by utilizing various proven disease-related databases (gnomAD, ExAC, Clinvar, ACMG, SIFT, CADD, etc.).

- In addition, we also have enabled analysis between patient groups and control groups in the three-step population-based analysis solutions, contributing to the identification of causal variant in relation to common diseases.



High-speed genomic data analysis (First step analysis)

Filtering based on the accuracy of sequencing and analysis

Variant annotation based on proven DBs such as gnomAD and ExAC

Selection of disease-specific genes and markers

Selection of significant variants that are clinically applicable

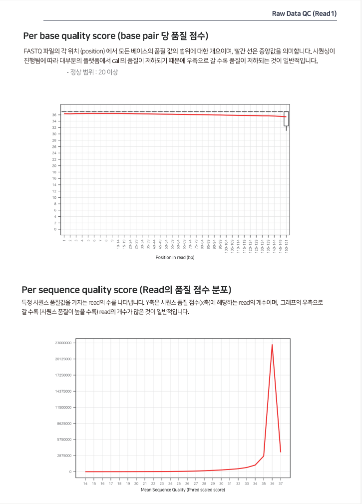
Convenient use through user-friendly UI

- Quality control results for FASTQ, BAM, and VCF files are provided in PDF and CSV formats to make it easier to check the quality of data. (Other companies do not provide such results)

iGAPoncloud
ALLinONE
Report for Whole-Genome Sequencing

ichrogene

시 료 명	EUS001
의뢰기관	아이크로진 생명정보 연구소
의 뢰 자	홍O동
분 석 일	2022-02-07



Raw Data QC (Read1)

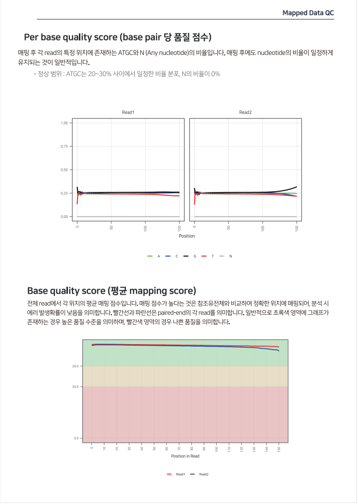
Per base quality score (base pair 당 품질 점수)

FASTQ 파일의 각 위치 (position) 에서 모든 베이스의 품질 값의 범위에 대한 개요이며, 빨간 선은 중앙값을 의미합니다. 시퀀싱이 진행됨에 따라 대부분의 플랫폼에서 call의 품질이 저하되기 때문에 우측으로 갈 수록 품질이 저하되는 것이 일반적입니다.

- 정상 범위 : 20 이상

Per sequence quality score (Read의 품질 점수 분포)

특정 시퀀스 품질값을 가지는 read의 수를 나타냅니다. Y축은 시퀀스 품질 점수(x축)에 해당하는 read의 개수이며, 그래프의 우측으로 갈 수록 (시퀀스 품질이 높을 수록) read의 개수가 많은 것이 일반적입니다.



Mapped Data QC

Per base quality score (base pair 당 품질 점수)

매핑 후 각 read의 특정 위치에 존재하는 ATGC와 N (Any nucleotide)의 비율입니다. 매핑 후에도 nucleotide의 비율이 일정하게 유지되는 것이 일반적입니다.

- 정상 범위 : ATGC는 20~30% 사이에서 일정한 비율 분포, N의 비율이 0%

Base quality score (평균 mapping score)

전체 read에서 각 위치의 평균 매핑 점수입니다. 매핑 점수가 높다는 것은 참조유전체와 비교하여 정확한 위치에 매핑되어, 분석 시 에러 발생확률이 낮음을 의미합니다. 빨간선과 파란선은 paired-end의 각 read를 의미합니다. 일반적으로 초록색 영역에 그래프가 존재하는 경우 높은 품질 수준을 의미하며, 빨간색 영역의 경우 나쁜 품질을 의미합니다.

Use of iGAPonCloud Service

Procedure to use all-in-one service

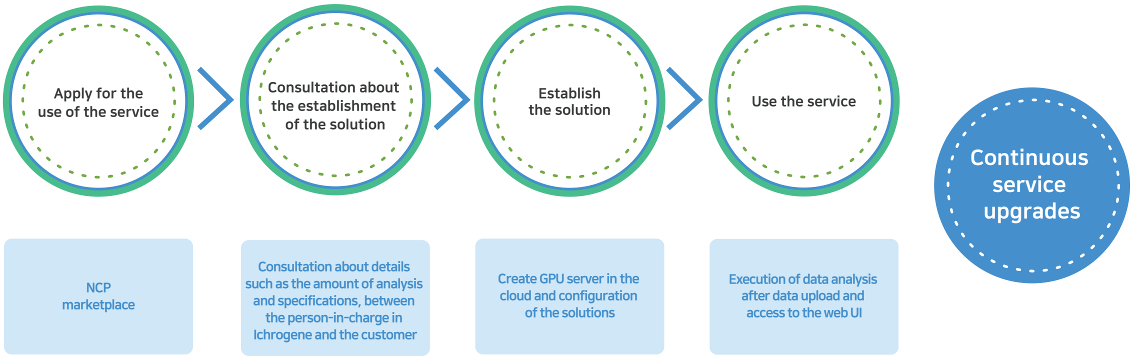
01 Apply for the use of the service in the marketplace within NAVER's cloud platform.

02 Proposes a solution suitable for customer needs through consultation between the person in charge of iGAPonCloud of Ichrogene and the customer.

03 Distributes a server image where a high-speed genome analysis environment and web UI are applied to allow the customer to access the UI through a web browser and perform analysis.

04 Customers who are familiar with Linux can directly access the Linux console to carry out analysis.

05 Continuous service upgrades allow customers to carry out analysis always using the latest solutions.



Apply for the use of the service

NCP marketplace

Consultation about the establishment of the solution

Consultation about details such as the amount of analysis and specifications, between the person-in-charge in Ichrogene and the customer

Establish the solution

Create GPU server in the cloud and configuration of the solutions

Use the service

Execution of data analysis after data upload and access to the web UI

Continuous service upgrades

Procedure to use all-in-one service

Q **How can I access web UI?**

A All servers have their own IP address, so costomers can access web UI by entering the IP address on the web browser. (ex: https://172.10.10.10:8080)

Q **Can I use the service only through web UI?**

A You can carry out the same task on CUI by accessing the terminal program (e.g., Putty) using the IP address mentioned above. However, since the access parameter automation part is excluded in the CUI mode, you need to make the json file yourself.

Q **Can I carry out multiple tasks at the same time?**

A In order to maximize service performance, even if a number of tasks are registered, only one task is carried out at a time. We recommend creating more servers for coustomers who want to work in parallel.

Q **What items are included in the QC report?**

A Most of the information provided by typical FastQC programs (fastq, bam, vcf) is included.

Q **If I stop the server, will I still be charged?**

A According to the NAVER Cloud platform policy, you will be charged continuously even if the GPU server is stopped. When the task is completed, be sure to notify us after downloading the related data.

ichrogene, Inc
Head office : 1205-1206 77, Changnyong-daero 256beon-gil, Yeongtong-gu, Suwon-si, Gyeonggi-do, Republic of Korea
Gwangju Branch : 602 193-12, Geumnam-ro, Dong-gu, Gwangju, Republic of Korea
ichrogene USA, Inc : 1952 Gallows Street Suite #303, Vienna, VA, 22182 USA

+82-31-243-1063 I www.ichrogene.co.kr