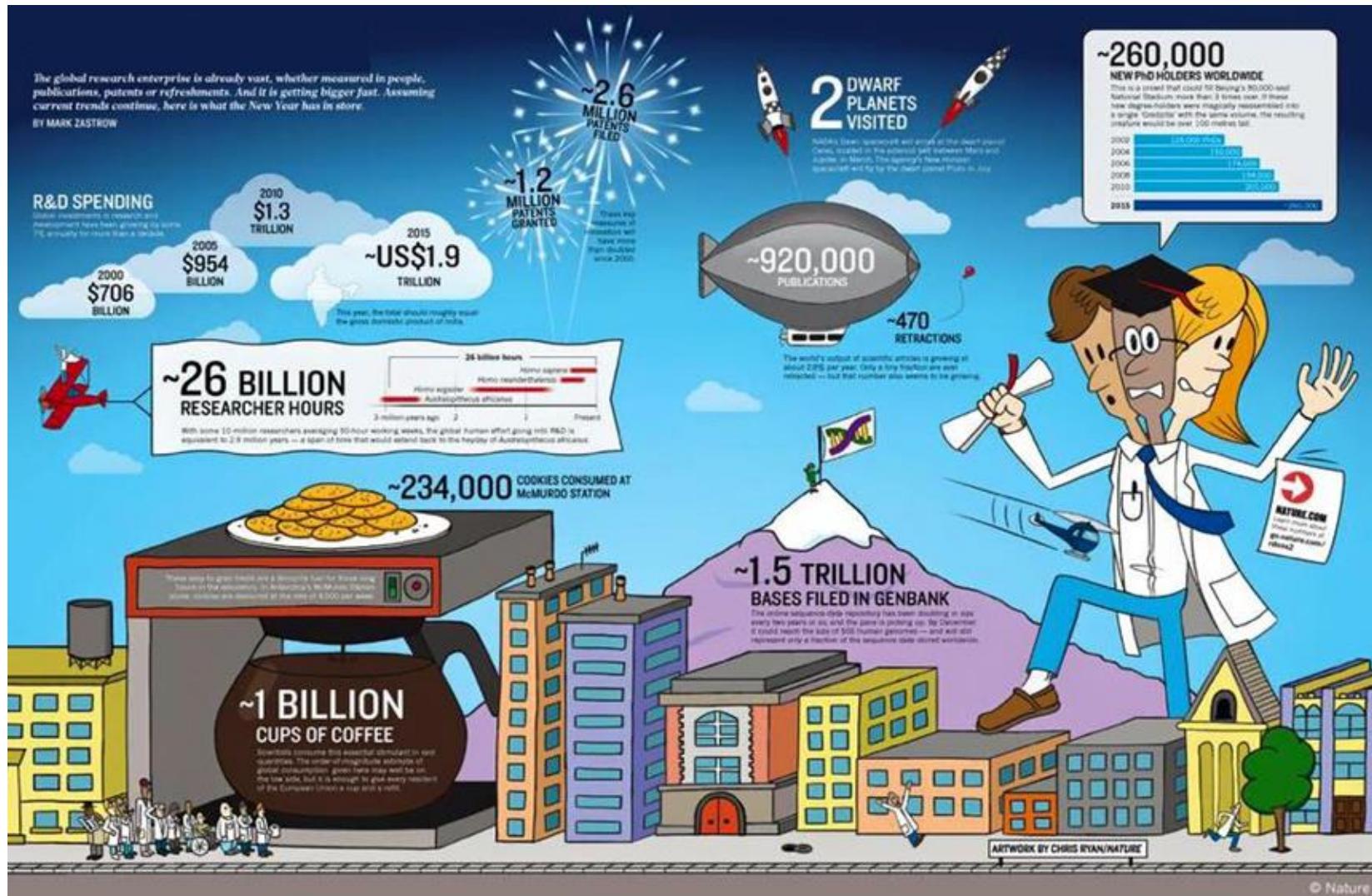




Introduction of NGS

Science Gossip



J.P. Morgan Bio-Health Conference(2018Y)



Immunotherapies

Dementia(Alzheimer)

Microbiome

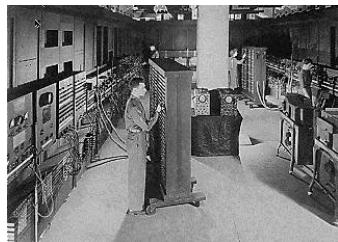
Mr Ahn's day



IT Vs BIT Industry

**Information
Technology
(IT)**

1950~



1980~



IBM

하드웨어

1985~



YAHOO!



Google

정보검색

facebook



SNS

**Genome
Technology**



illumina®



life technologies™



테라젠이텍스
THERAGEN ETEX CO.,LTD.

Knome®

DNAexus



Google

23andMe

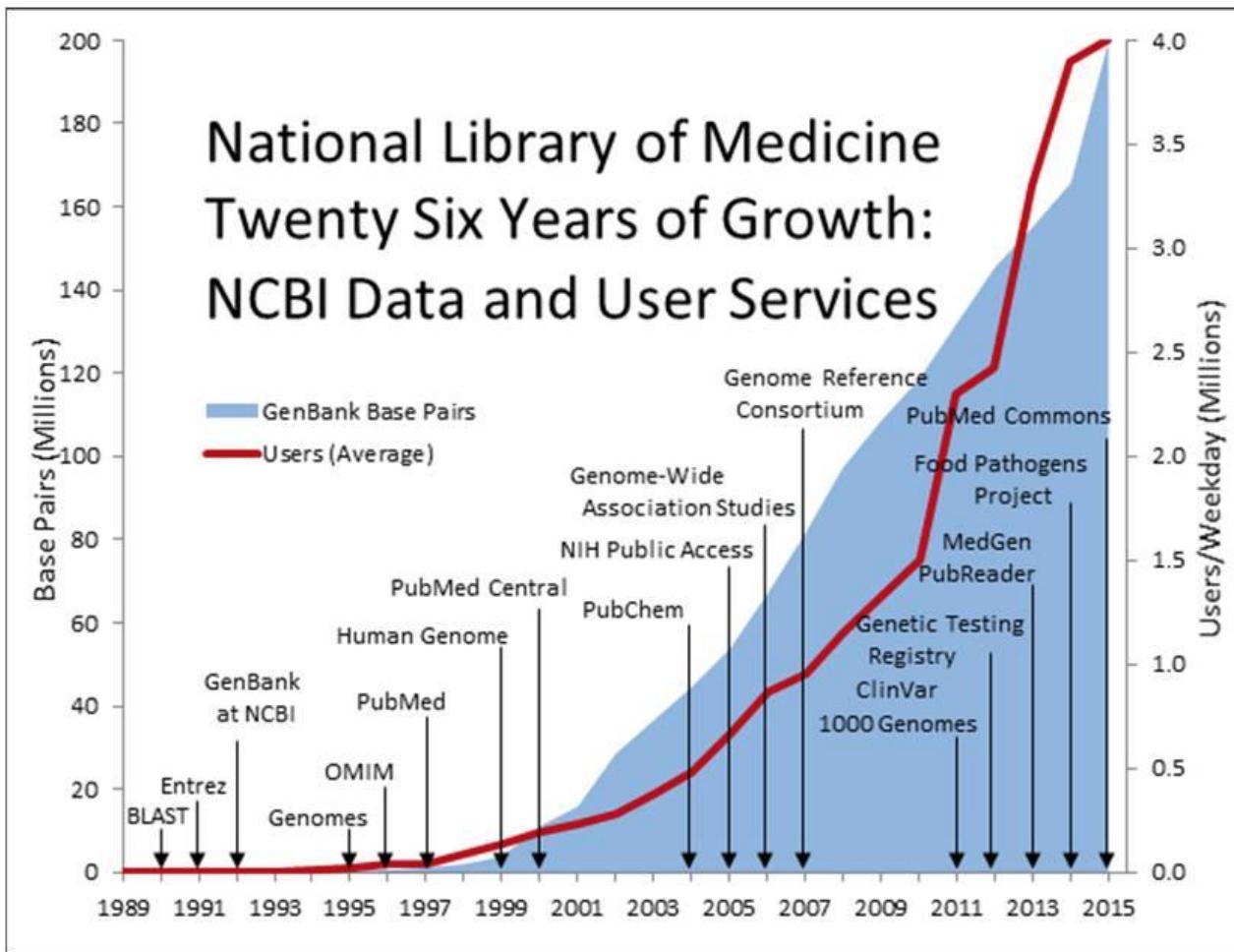
Microsoft Amalga
Unified Intelligence System

2009~

2012~

2020~

NCBI Data & User Services(26 Years)

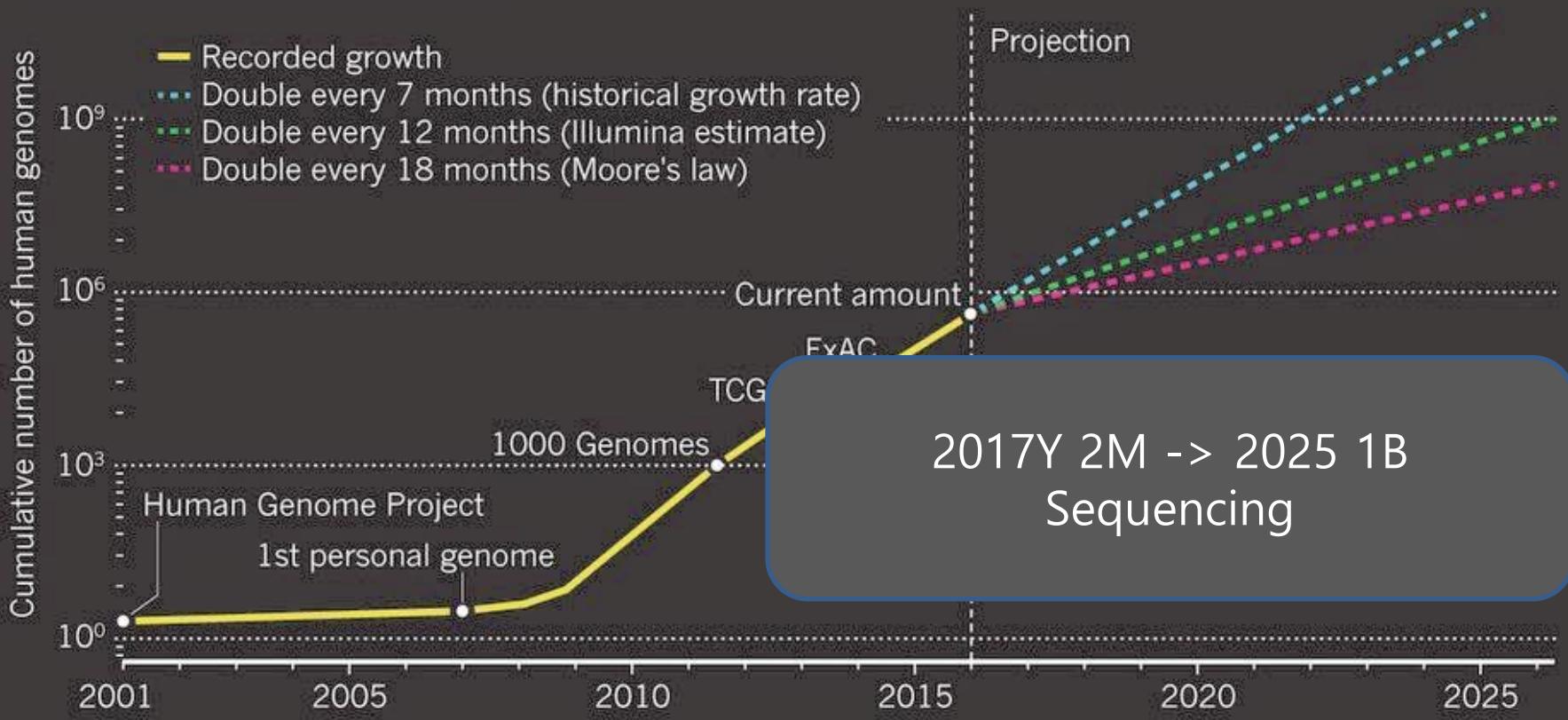


- Genbank: 37만 종, 2억개 서열
- Bio/Medical 2,600만 편 논문
- 9,000만 개 화학구조
- 120만 개 Bioassay
- 23만 건의 clinical trial 등록
- 6만 종 유래, 7000만 개 단백질
- 2년마다 2배씩 증가

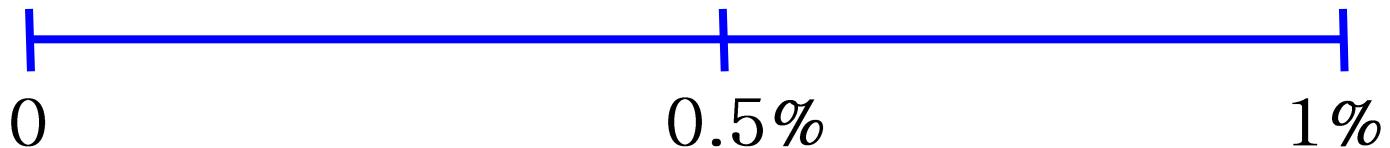
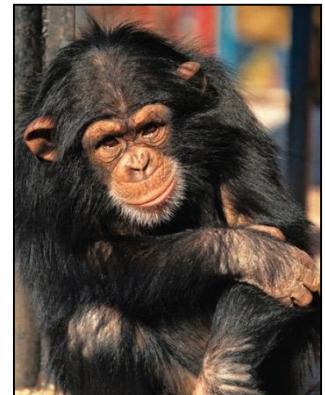
Human Genome Sequencing

DNA SEQUENCING SOARS

Human genomes are being sequenced at an ever-increasing rate. The 1000 Genomes Project has aggregated hundreds of genomes; The Cancer Genome Atlas (TCGA) has gathered several thousand; and the Exome Aggregation Consortium (ExAC) has sequenced more than 60,000 exomes. Dotted lines show three possible future growth curves.



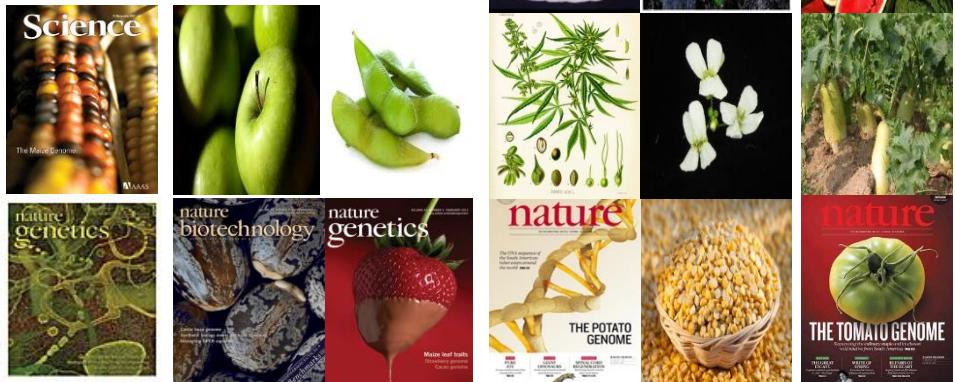
It's Different.



Whole Genome Denovo with NGS



2014



2000

2002

2006

2007

2008

2009

2010

2011

2012

2001

200

200

2007

2009

2010

2011

2012



Illumina Value(\$50B=SK Hynix)

Illumina, Inc. (ILMN)  Add to watchlist

NasdaqGS - NasdaqGS Real Time Price. Currency in USD

301.59 **-18.34 (-5.73%)**

301.70 **+2.74 (0.91%)**

At close: 4:00PM EST

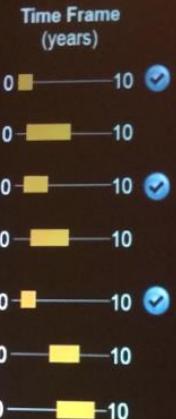
After hours: 5:38PM EST



Illumina Goal

Where Are We Headed?

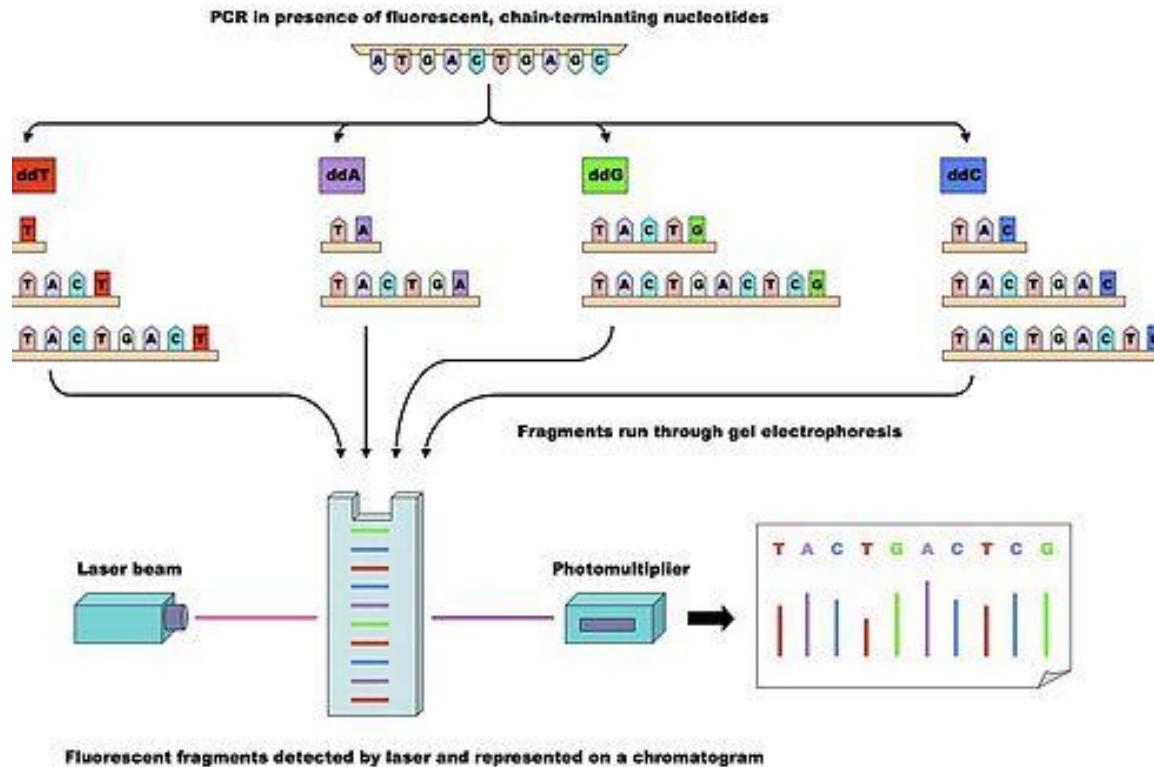
1. Sequencing begins to directly save lives
2. Tumor samples routinely sequenced/Standard of Care
3. Nations begin sequencing their populations
4. Sequences accessible in electronic medical record
5. Complete human sequence under \$1,000
6. Infants routinely sequenced at birth
7. Cancer managed as chronic disease



1. 게놈 기술로 사람 생명 구하기. (니콜라스 볼커)
2. 루틴하게 암 검체 진단하기. (파운데이션 메디슨)
3. 국가적으로 자국민을 대상으로한 대규모 게놈 프로젝트 시작하기. (10만게놈 프로젝트, 오마바 PMI)
4. 게놈 데이터를 병원 EMR을 통해 활용하기.
(미국내 병원들에 시도중)
5. 백만원에 사람 게놈 해독하기. (HiSeqX10)
6. 신생아들 대상으로 루틴하게 검사하기.
7. 암 게놈 기술을 통해 암을 만성질환처럼 관리하기

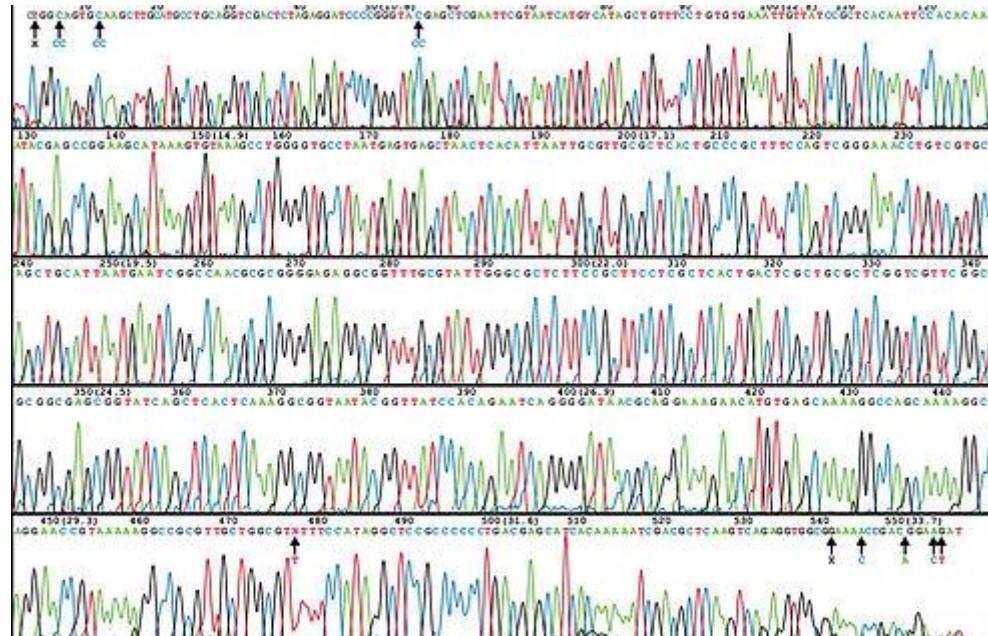
2008년 제시 하였던 Illumina의 초기 road map 중 3개가 이미 실현이 되었고 5년 안에 추가적으로 4개가 실현 될 것으로 예상

1st : Sanger Sequencing



그림에 나와있는 3'-ATGACTGAGC-5'와 같은 template DNA의 서열을 분석하는 경우에, 일반적인 d{A,C,G,T}TP (NTP)만 넣어주면 DNA polymerase는 이에 상보적인 DNA를 합성하게 된다. 하지만 dNTP 외에 소량의 dd{A,C,G,T}TP (ddNTP)를 섞어주게 되면, DNA polymerase가 template DNA에 상보적인 서열을 합성해 나가다가, 중간중간에 ddNTP가 끼어 들어간 DNA 문자가 합성되게 된다. 그러한 문자는 더 이상 길어지지 않고 합성이 중단되게 된다. 이들 ddNTP에는 각각을 구별할 수 있는 형광물질이 결합되어 있기에, 새로이 합성된 DNA들의 마지막 염기 종류에 따라 서로 다른 형광을 띠게 된다. 전기영동법에 의해 크기 순으로 나열할 수 있으며, 레이저 빛을 각 전기영동 밴드에 쪼면, 형광물질에 따라 특이적인 파장의 빛을 발하게 되며, 이를 순서대로 읽으면 원래 염기서열과 상보적인 5'-TACTGACTG-3'을 얻게 된다.

1st : Sanger Sequencing



전기영동으로 다양한 크기의 DNA 분자를 분리하는 방법은, DNA 분자에 있는 인산결합이 음전하를 띠고 있기에 아래쪽에 양극을 배치하여 전기적으로 당기게 된다. 일반적으로 수 시간이 걸리게 되는데, 처음에는 크기에 따라 잘 구별되던 DNA 분자들이 시간이 경과하면서 확산 작용에 의해 번지게 된다. 따라서, 늦게 나오는 전기영동 밴드는 더 넓게 퍼진 상태가 되며, 심해지면 이웃 밴드와 겹치게 되어 분해능이 떨어지게 된다. 오른쪽의 일반적인 전기영동 크로마토그램을 보면 이와 같은 현상을 볼 수 있다.

2nd : NGS(Next Generation Sequencing)

Next Generation Sequencing (NGS)

VS.

Sanger Sequencing

NGS

- High depth of coverage
- Fast turnaround time for sequencing larger regions
- Cost-effective when sequencing larger regions
- Not as effective at sequencing repetitive regions

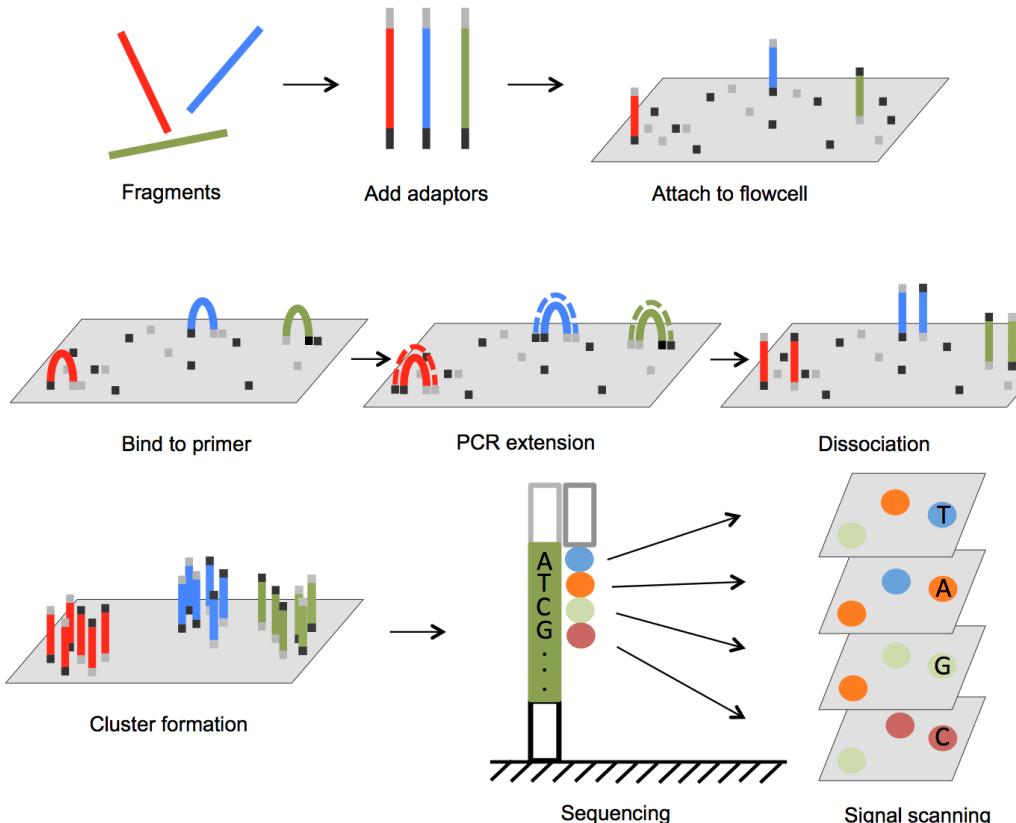
Sanger

- Low depth of coverage
- Slow turnaround time for sequencing larger regions
- Expensive when sequencing larger regions
- More effective at sequencing repetitive regions

ACACTGGAACTGAGACACGGTCCAGTC
GGGGAAACCCCTGAAGCAGCAACGCCGCGT
TTAGGGAAAGAACCATGACGGTACCTACG
GTAATACGGAGGGTGCAGCGTTACTCGG
AAGTCTTTGTGA MUTATION ATCT
ATCTAGAGTGAGGGAGAGGCAGATGGAA
GGAATACCCATTGCGAAGGCGATCTGCT
GGCAAACAGGATTAGATAACCTGGTAGT
CTAGTCACGGCAGTAATGCACCTAACGG



NGS – HiSeq(Illumina)

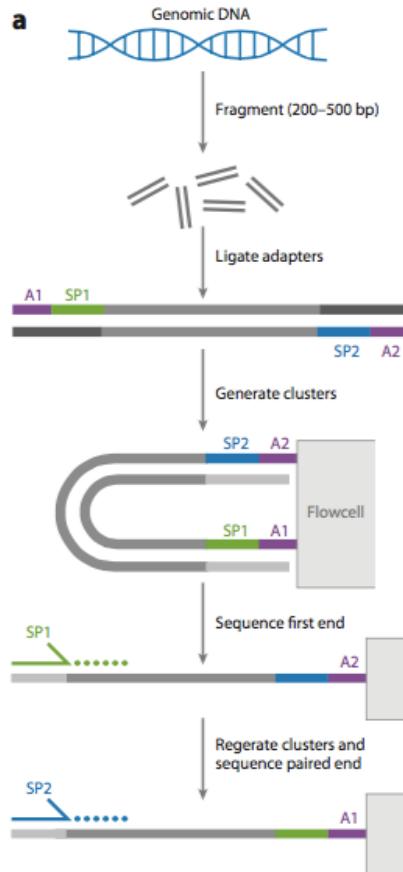


시퀀싱하고자하는 서로 다른 수십억 개의 single-strand DNA 조각들의 양쪽 끝에 double-stranded linker(adapter)를 붙여준다. Flowcell 바닥에는 이들 adaptor 서열에 상보적인 DNA를 고정하여 놓고, adaptor가 붙어 있는 조각들을 flowcell에 고루 퍼지게 하면, 각 조각은 여기저기에 달라붙게 된다. 양 끝이 모두 결합되면 브릿지처럼 된다. 이때 생성된 이중 가닥 부분을 primer로 사용하여 Polymerase chain reaction(PCR)을 하여 증폭 시킨다. PCR이 진행되면 새로 합성된 DNA 가닥들은 바닥에 고정되고 고정된 DNA 가닥은 반대 방향 adapter에 의해 primer와 결합하여 다시 합성된다. 이 과정이 반복되면서 증폭이 끝날 때는 하나의 DNA 조각에 약 1000개 정도의 동일한 PCR 결과물이 하나의 cluster를 이루게 된다.

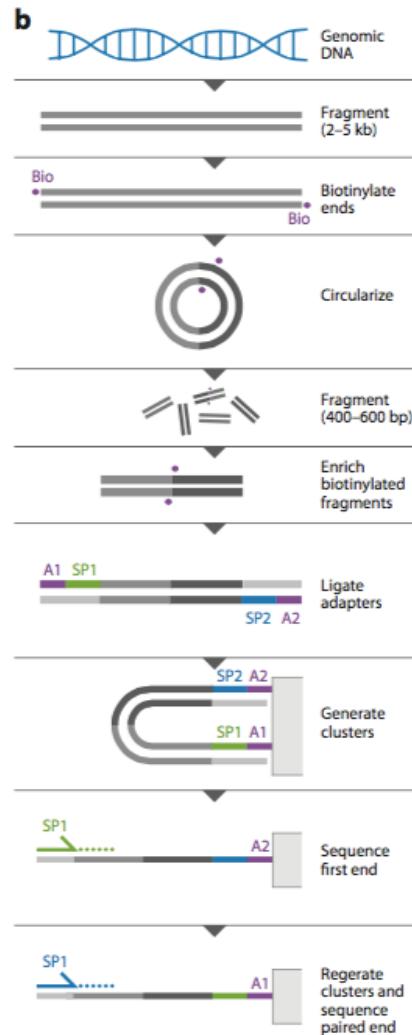
즉, 애초에 사용한 DNA 조각 하나로부터 하나의 클러스터가 형성되며, 이들은 다음 단계에서 사용하는 형광법으로 검출하기 충분한 양이 되는 것이다. 이때, 한쪽 가닥은 없애고 나머지 가닥만을 시퀀싱한다. 이 cluster들에 염기의 종류에 따라 각각 다른 빛을 내는 형광 물질이 결합된 dNTP가 DNA polymerase에 의해 한 염기씩 합성될 때, 염기마다 형광 물질 빛을 읽은 뒤 기록한다. 기록한 형광 물질은 washing으로 제거한 뒤 새로운 dNTP가 합성되면 위 과정을 반복한다. 여기서 중요한 것은, 한 염기씩 붙이고 형광을 검출한 후에, 그 형광을 없애고 그 다음 염기를 붙여야 한다는 점이며, 이 과정이 착오없이 차근차근 진행되어야 한다는 점이다.

NGS – HiSeq(Illumina)

paired-end



mate-pair



- Paired-end sequencing

- DNA 조각의 양 끝단을 해독하는 기술
- DNA 조각의 크기를 조절함으로써 해독되는 서열 길이보다 넓은 영역의 정보를 얻을 수 있음.
- 800bp 길이의 DNA를 사용 가능. (최신 장비는 500bp 이하)

- mate-pair

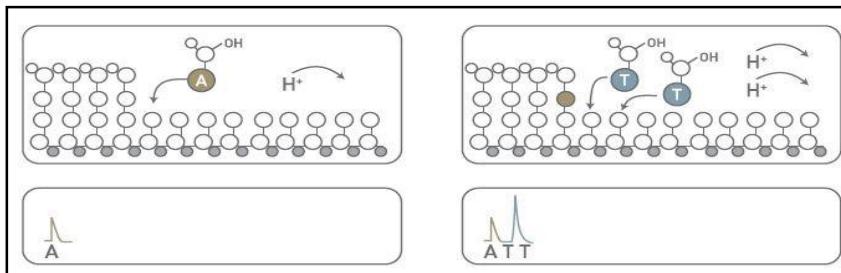
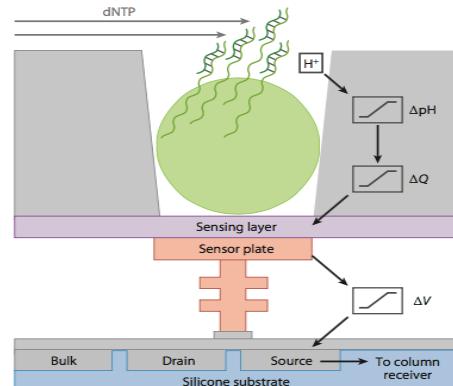
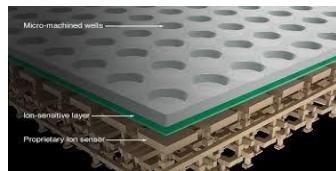
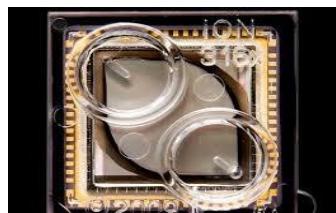
- 1kb 이상의 DNA 조각의 양 끝단을 해독하는 기술

NGS의 짧은 서열 길이에 따른 단점을 보완할 수 있는 기술

SV, Gene Fusion, Alternative Splicing 등 다양한 분석에 폭넓게 사용됨.

2.5G Sequencer(Ion PGM/Proton)

Semi-conductor Sequencing



<http://tdelia-irsc.weebly.com/ion-torrent-dna-sequencing.html>



• 해독 원리

- dNTP 를 이용한 합성과정에서 발생하는 pH 변화를 semi-conductor 에서 감지하여 서열을 해독
- dATP를 반응시켜보고 수소 이온이 발생하는지 측정해 본다. 시그널이 나오지 않으면 남은 dATP를 씻어내고 다른 dNTP들을 차례로 반응시켜본다

• 해독 능력

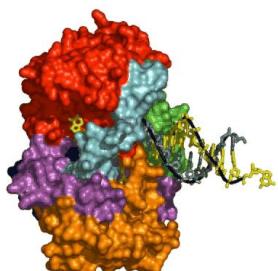
- 35bp~400bp 길이
- 5시간 내에 1G~10G의 해독이 이루어짐.
- Homopolymer 문제가 있음.

• Ion torrent PGM Dx

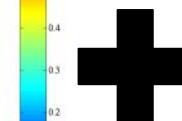
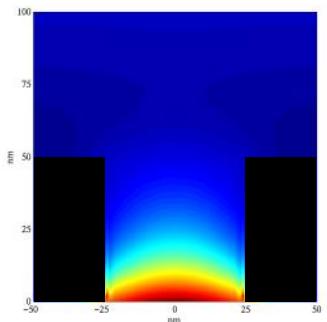
- 미국 FDA class II 허가, 국내 체외 진단용 의료기기 허가

3rd : SMRT RSII(PacBio)

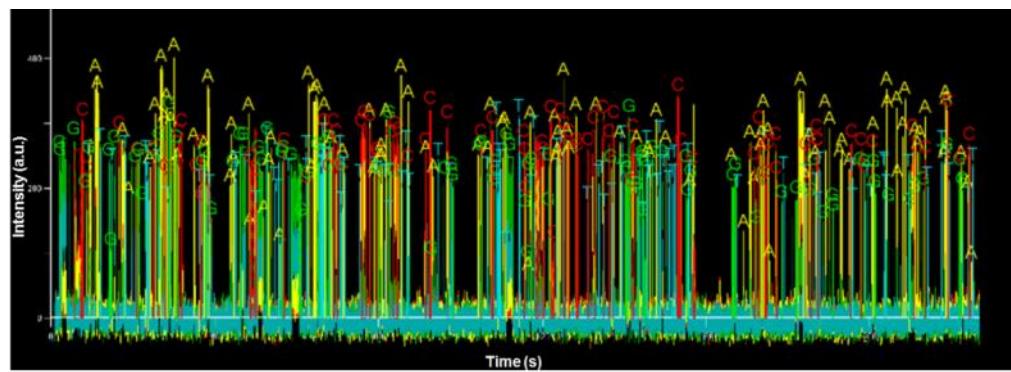
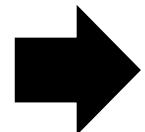
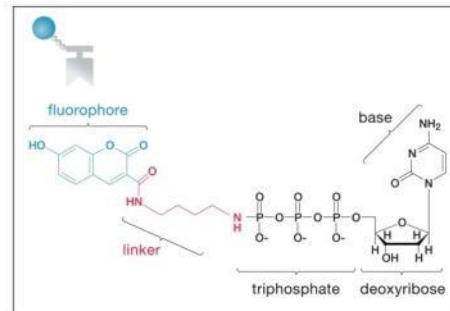
DNA Polymerase/ Template Complex



ZMW Confinement



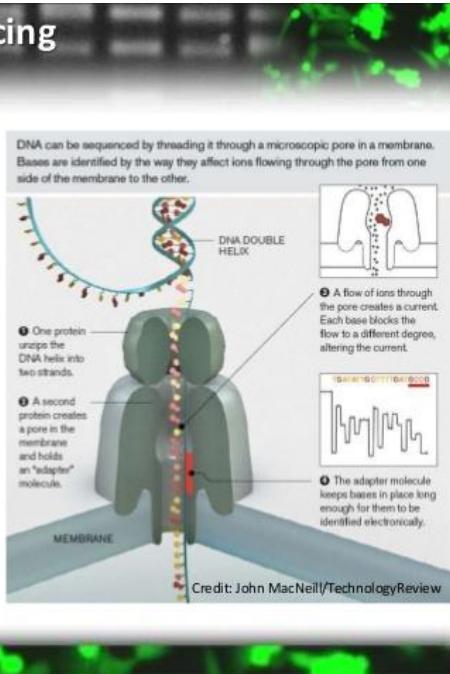
Phospholinked Nucleotides



3rd : Oxford Nanopore Sequencing

Future: Nanopore Sequencing

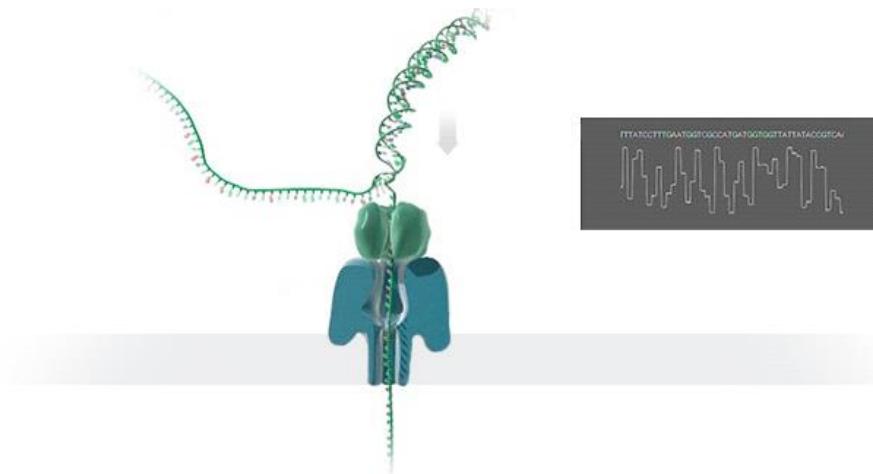
- Leading candidate is Oxford Nanopore
- Concept
 - Detect flow of electrons through the pore
 - Each base causes a detectable change in the current
 - Results in direct sequencing
 - Theoretically could be used to sequence RNA and protein too
- Advantages
 - Long read length
 - Plug and play
 - Easily scalable
- Disadvantages
 - No hard data yet
 - No specific release date



- 막 단백질에 나노미터 수준의 채널을 형성한 후에 이곳에 DNA 단일 가닥을 통과시킬 때, 염기 서열에 따라 막 간의 전위 차가 변하는 것을 착안하여 시퀀싱

- 바깥 쪽에서 이중 나선의 DNA를 helicase 효소를 이용하여 풀어주면서 단일 가닥이 일정한 속도로 채널 속으로 이동하도록 구성되어 있으며, 채널 안에 약 4 염기 정도가 들어가기 때문에 실제로 측정한 전위 차는 한 염기에 해당하는 것이 아니라 4 염기 단위로 얻어지는 것이다

- template DNA에 상보적인 화학 반응을 하지 않는다는 점이고 DNA를 있는 그대로 측정한다



NGS 장비 - Illumina

Illumina SBS tech-based Sequencer Systems

	Miseq (Dx)	Hiseq2500/2000	Hiseq4000/X-10
Illumina SBS tech-based Sequencer Systems			
적합한 서비스	Amplicon Seq, Targeted Reseq, 16S metagenomics, ctDNA seq	WES, RNA seq, Methyl-seq, sm RNA seq, mtDNA seq, Single cell genomics	WES, WGS, RNA seq, Single cell genomics
분석 sample (形/量)	DNA: 300 ng 이상, FFPE : ϕ 20 μm 이상, ctDNA 분석용 plasma: 400 μl 이상,	Blood: 400 μl 이상, Tissue: 150 μg 이상, *Single cell genomic 가능	
Dx 승인 여부	미국 '13년 11월 19일 한국 '16년 5월 11일	한국, 미국 인증 無	한국, 미국 인증 無
Output Range	0.5Gb~15Gb	250~300Gb / 720~800Gb	650~750Gb / 1300~1500Gb
Q value	Q30 (99.999%)	Q30 (99.999%)	Q30 (99.999%)
Initializing time	2h	10h	10h
Run time	5h~3.5 days	2.5~6 days	3.5~6 days
Analysis time	9h	1.5~2 days	2~3 days

Newest NGS Platform

Illumina
HiSeq X/4000



“Personal Genomics
– \$1,000 Genome”

Thermo Fisher
Ion S5, Proton



“Benchtop
– Clinical”

PacBio
RS II, Sequel



“15kb length
– De novo
Assembly”

Oxford
Nanopore
MinION



“ultra portable
– Virus”

Mobile Sequencing : SmidgION



SmidgION: nanopore sensing for use with mobile devices

Using the same core technology as the handheld MinION device, we are now starting to develop an even smaller device.

In early development



VolTRAX™: rapid, programmable, portable, disposable sample processor

\$100 Human Genome Era

NovaSeq



>

HiSeq X-10



	HiSeq X	NovaSeq	Note
Running time	3days	1.8days	
Throughput	1.8Tb	6Tb	
Tb/Year	219Tb	1368Tb	6.25 Times

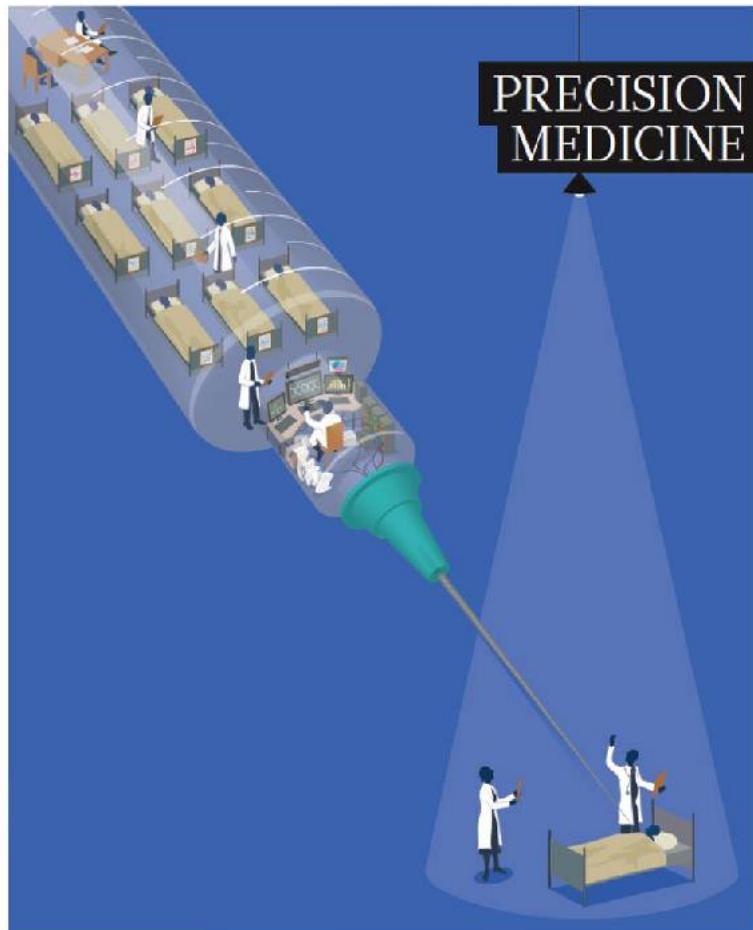
NGS Platform Features

Platform	MiniSeq	MiSeq	Next Seq 500	HiSeq 2500 RR	HiSeq 2500 V3	HiSeq 2500 V4	HiSeq 4000	HiSeq X	NovaSeq 6000 2017	NovaSeq 6000 2018	Ion Proton P1	RS P6-C4	Sequel	Smidg ION 2017	Mini ION R 9.4	Prome thl ON	Prome thION th eoretical	QiaGen Gene Reader	BGI SEQ 500	BGI SEQ 50
Reads: (M)	25	25	400	600	3000	4000	5000	6000	10000	20000	165	5.5	38.5	--	0.05	--	--	400	1600	1600
Read length: (paired-end*)	150*	300*	150*	100*	100*	125*	150*	150*	150*	150*	200	15K	12K	--	10K	10K	--	--	100*	50
Run time: (d)	1	2	1.2	1.125	11	6	3.5	3	1.66	1.66	--	4.3	4.3	--	2	--	--	--	1	0.4
Yield: (Gb)	7.5	15	120	120	600	1000	1500	1800	3000	6000	10	12	84	5.5	2.75	3100	6400	80	200	8
Rate: (Gb/d)	7.5	7.5	100	106.6	55	166	400	600	1800	3600	--	2.8	19.5	--	1.375	--	--	--	200	20
Reagents: (\$ K)	--	1	4.41	6.145	23.47	29.9	29.9	12.75	--	--	1	2.4	11.2	--	0.5	2.5	2.5	0.5	--	--
per-Gb: (\$)	--	93	36.75	51.2	39.1	29.9	20	7	--	--	100	200	80	--	180	20	20	--	--	--
hg-30x: (\$)	--	11160	4410	6144	4692	3588	2400	840	--	--	12000	24000	9600	--	21800	2400	2400	--	--	--
Machine: (\$)	49.5K	99K	250K	740K	690K	690K	900K	1M	985K	985K	243K	695K	350K	--	1K	75K	--	--	200K	--

https://docs.google.com/spreadsheets/d/1GMMfhyLK0-q8XkIo3YxIWaZA5vVMuhU1kg41g4xLkXc/edit?hl=en_GB&fbclid=IwAR1w8J14LmcwJZJ2QRll924hwsQaZzwMV5xsbcAnctpXMnKdSeoEjNosJ0s#gid=1569422585

Precision Medicine(정밀의학)

nature INSIGHT



Personalized medicine:
Time for one-person trials



N-of-1 clinical trial



Unknown Case(Rare Disease)



Nicholas S. Volker



Robert Stone



Maya Nieder



박채원*

장 천공, 패혈증, 고열

2세 때 최초 증상
4년간 160회 수술
800일 이상 입원

운동 및 의사소통 장애

1세 때 최초 증상
14년간 원인파악 실패
부모는 지능이상으로
판단하고 의사소통 포기

심각한 발달 장애

2세 때 최초 증상
6차례 유전자 검사 실패

복시, 이명, 고열

15세 때 최초 증상
진단을 받지 못함
동네 안과, “이상 없음”
검진센터, “이상 없음”

Decided Treatment After Genomics Test

장 천공, 패혈증, 고열



Nicholas S. Volker

운동 및 의사소통 장애



Robert Stone

심각한 발달 장애



Maya Nieder

복시, 이명, 고열



박채원*



면역유전자 XIAP 이상 발견
제대혈 수혈, 증상 호전
일반음식 섭취, 야외활동 가능



Dystonia 16 증후군 진단
(전세계에서 9번째 사례),
근육운동을 돋는 치료제 처방,
현재 보조기구를 통해
부모와 의사소통 가능



유아기 발달에 관여하는 유
전자 이상 발견 (세계 최초
사례), 현재 연구 기관에서
해당 질병 연구 중



다발성경화증, 탈수초증상,
시신경척수염 판정 전전.
현재 의식불명, “병명없음”
3개월간 3천만원 사용
(병명 없어 의료비 지원 불가)

Clinical Study Case in Domestic



선천성 근무력증 11살 보경이가
태어나 처음 걷게 된 과정

2003년 희귀 유전성 질환인
선천성 근무력증으로 출생

두 살 때 근육 조직 검사 결과
근무력증 진단

지난해 열 살 때까지 줄곧 놀거나
누군가 앉혀주면 앉아 생활해옴

서울대병원, 미국 워싱턴대 의대
유전체 분석 연구소에 검사 의뢰



근육질환 관련 DOK7 유전자
변이 발견

올해 초부터 신경과 근육 연결
작용 활성 신경 물질(아세틸콜린)
강화제 투여

6월부터 조금씩 일어서기 시작

10월 현재 화장실 걸어서 가고
계단 오르기 시작

자료: 서울대병원 소아신경과

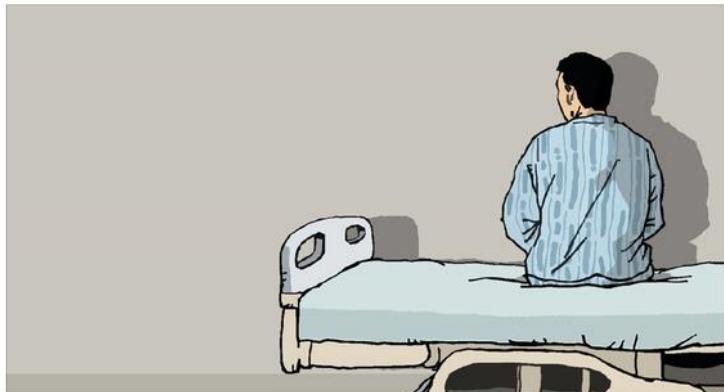
NGS를 이용한 원인
유전자의 변이 발견

Clinical Study : Assume

오진으로 '잃어버린 13년'...약 바꾸고 이를 만에 일어나

[중앙일보] 입력 2017.12.06 11:24 수정 2017.12.06 12:47

김정석 기자



[중앙포토]

3살 때 뇌성마비 판정을 받고 13년을 누워 지낸 여성이 약을 바꾸고 이를 만에 일어나 걸었다. 병원의 오진 때문이었다. 간단한 치료만으로 호전될 수 있는 병이었지만 잘못된 처방과 치료로 이 여성은 자신 평생의 절반 이상을 누워 지냈다.

로켓배송
1+1
D-FORLA
D-FORLA
D-FORLA
D-FORLA

coupons
로켓페이로 계좌이체하면
추가 2% 캐시!
추천기사

1조 더 덜여 무안공항 경유
호남KTX '드자'로 훈다
“한국인은 어린애 같다”
그 13년 뒤 시진핑의 공세

이청연, 인천교육감직
상실...징역 6년형 확정
中 환구시보 “전쟁나도 첫 공격 대상은

'세가와병'이라고 불리는 이 병은 신경전달물질의 합성에 관여하는 효소의 이상으로 도파민 생성이 감소해 발생한다. 주로 소아에게 발생하며 소량의 도파민 약물을 투약하면 특별한 합병증 없이 치료가 가능하다. 때문에 조기진단과 치료가 매우 중요하다

우리가 NGS 기술을 생각해 봤더라면... ㅠㅠ

NIPT(Non-Invasive Prenatal Test)

OPEN ACCESS Freely available online



The Feasibility Study of Non-Invasive Fetal Trisomy 18 and 21 Detection with Semiconductor Sequencing Platform

Young Joo Jeon^{1,2,*}, Yulin Zhou^{3,*}, Yihan Li^{4,*}, Qiwei Guo³, Jinchun Chen⁴, Shengmao Quan⁵, Ahong Zhang⁵, Hailing Zheng⁵, Xingqiang Zhu⁵, Jin Lin⁵, Huan Xu⁵, Ayang Wu⁶, Sin-Gi Park^{1,2}, Byung Chul Kim^{7,8}, Hee Jae Joo^{1,2}, Hongliang Chen^{5*}, Jong Bhak^{1,2,6,9}

1TheragenEx Bio Institute, Suwon, Republic of Korea, **2**Genome Care, Seoul, South Korea, **3**Molecular Diagnostics Laboratory, Department of Medical Genetics, Prenatal Diagnosis Center, Xiamen Maternal and Child Health Hospital, Xiamen, Fujian, China, **4**Xiamen Hospital of Traditional Chinese Medicine, Fujian University of Traditional Chinese Medicine, Xiamen, Fujian, China, **5**Xiamen Vangenes BioTech, Xiamen, Fujian, China, **6**Zhangzhou Affiliated Hospital of Fujian Medical University, Zhangzhou, Fujian, China, **7**Clinomics, Ulsan, Republic of Korea, **8**Personal Genomics Institute, Genome Research Foundation, Suwon, Republic of Korea, **9**BioMedical Engineering, UNIST, Ulsan, Republic of Korea

Abstract

Objective: Recent non-invasive prenatal testing (NIPT) technologies are based on next-generation sequencing (NGS). NGS allows rapid and effective clinical diagnoses to be determined with two common sequencing systems: Illumina and Ion Torrent platforms. The majority of NIPT technology is associated with Illumina platform. We investigated whether fetal trisomy 18 and 21 were sensitively and specifically detectable by semiconductor sequencer: Ion Proton.

Methods: From March 2012 to October 2013, we enrolled 155 pregnant women with fetuses who were diagnosed as high risk of fetal defects at Xiamen Maternal & Child Health Care Hospital (Xiamen, Fujian, China). Adapter-ligated DNA libraries were analyzed by the Ion Proton™ System (Life Technologies, Grand Island, NY, USA) with an average 0.3x sequencing coverage per nucleotide. Average total raw reads per sample was 6.5 million and mean rate of uniquely mapped reads was 59.0%. The results of this study were derived from BWA mapping. Z-score was used for fetal trisomy 18 and 21 detection.

Results: Interactive dot diagrams showed the minimal z-score values to discriminate negative versus positive cases of fetal trisomy 18 and 21. For fetal trisomy 18, the minimal z-score value of 2.459 showed 100% positive predictive and negative predictive values. The minimal z-score of 2.566 was used to classify negative versus positive cases of fetal trisomy 21.

Conclusion: These results provide the evidence that fetal trisomy 18 and 21 detection can be performed with semiconductor sequencer. Our data also suggest that a prospective study should be performed with a larger cohort of clinically diverse obstetrics patients.

Citation: Jeon YJ, Zhou Y, Li Y, Guo Q, Chen J, et al. (2014) The Feasibility Study of Non-Invasive Fetal Trisomy 18 and 21 Detection with Semiconductor Sequencing Platform. PLoS ONE 9(10): e110240. doi:10.1371/journal.pone.0110240

Editor: Sinuhe Hahn, University Hospital Basel, Switzerland

Received: June 11, 2014; **Accepted:** September 11, 2014; **Published:** October 20, 2014

Copyright: © 2014 Jeon et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper. Sequence data has been deposited in the NIH short read archive (SRA) with the following BioProject accession number: SRP044689

Funding: This work was supported by the Industrial Strategic Technology Development Program, 10040231, 'Bioinformatics platform development for next generation bioinformation analysis' funded by the Ministry of Knowledge Economy (MKE, Korea). Co-authors Young Joo Jeon, Sin-Gi Park, Hee Jae Joo, and Jong Bhak are employed by TheragenEx Bio Institute. TheragenEx Bio Institute provided support in the form of salaries for authors YJJ, SGP, HJJ, and JB, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section. Co-authors Young Joo Jeon, Sin-Gi Park, Hee Jae Joo, and Jong Bhak are employed by Genome Care. Genome Care provided support in the form of salaries for authors YJJ, SGP, HJJ, and JB, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section. Co-author Byung Chul Kim is employed by Clinomics. Clinomics provided support in the form of salaries for author BCK, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of this author are articulated in the 'author contributions' section.

Competing interests: We have the following interests: co-authors Young Joo Jeon, Sin-Gi Park, Hee Jae Joo, and Jong Bhak are employed by TheragenEx Bio Institute. Co-authors Young Joo Jeon, Sin-Gi Park, Hee Jae Joo, and Jong Bhak are employed by Genome Care. Co-authors Shengmao Quan, Ahong Zhang, Hailing Zheng, Xingqiang Zhu, Jin Lin, Huan Xu, and Hongliang Chen are employed by Xiamen Vangenes BioTech. Co-author Byung Chul Kim is employed by Clinomics. There are no patents, products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials.

* Email: hongliangchen@gmail.com (HJ); jongbhak@genomics.org (JB)

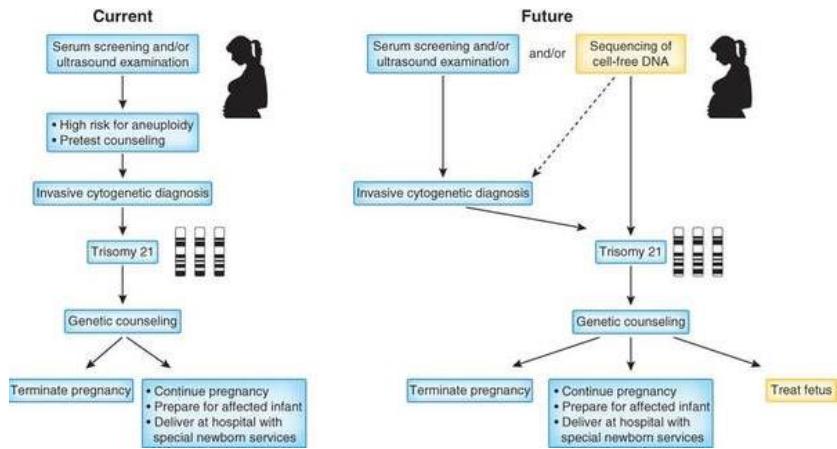
† These authors contributed equally to this work.

• 국내에서는 처음으로 비침습적 태아검사(NIPT) 기술을 이용한 임상 결과를 정리한 논문을 PLoS One에 출판

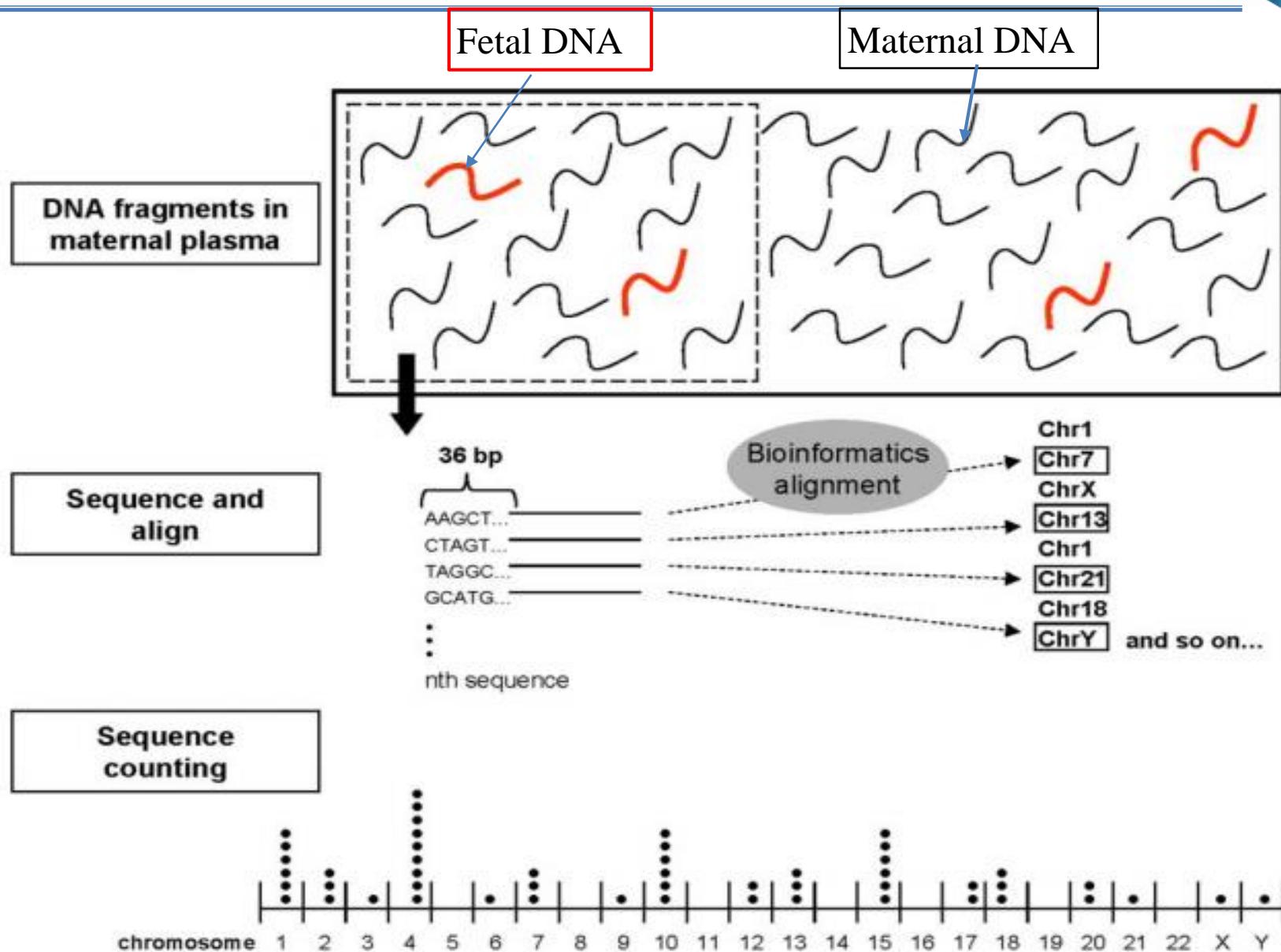
- 155명의 산모 혈액에서 fetal cell-free DNA를 얻어 분석
- 18번, 21번 염색체의 trisomy 여부를 정확하게 판독하는 실험 및 분석 기법을 개발
- 약 0.3X WGS 시퀀싱만으로도 오류 없이 trisomy를 검출

Proton I 칩 하나

- 10개의 검체를 분석이 가능, 비용이 매우 저렴
- 해독 및 분석이 하루 만에 모두 완료



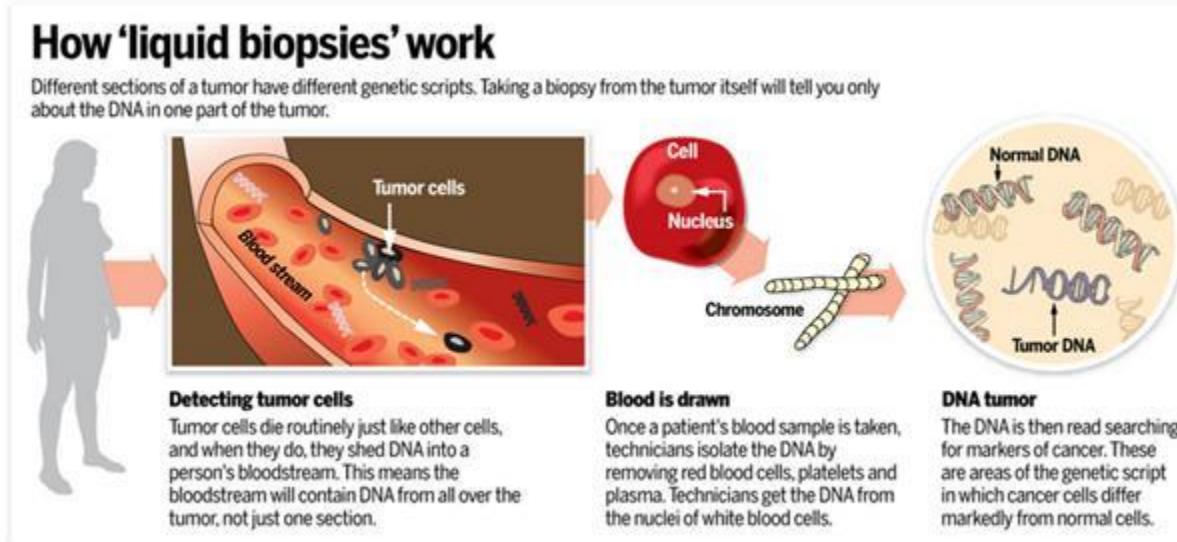
Advantage of TheragenEtex at NIPT



ctDNA from Plasma : Liquid Biopsy

■ 액체 생체검사(liquid biopsy) : Relapse or Meta screening in Cancer

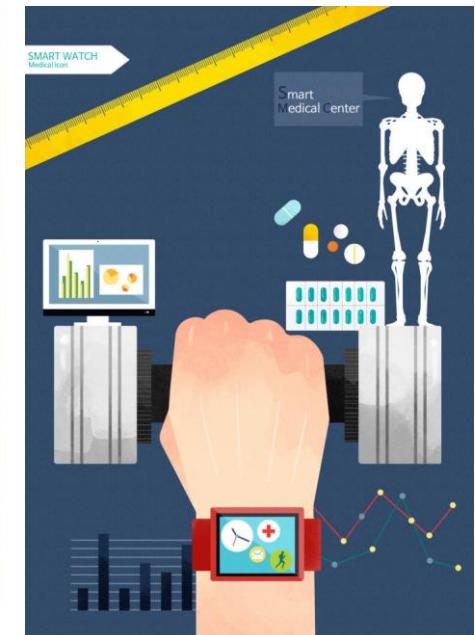
1. 액체 생체검사는 비침습적인(non-invasive) 방법으로, 기존의 침습적인 진단 및 검사방법의 대안으로 주목
 - 기존의 암 진단방법은 조직 샘플 채취 및 대장 내시경 검사 등의 침습적인 방법에 의존
 - 액체 생체검사는 혈액 등의 체액 검사만으로 신체 부위별 혈액 내에 존재하는 암세포 유래 DNA를 분석하여 암 발생 및 전이 등에 대한 상세한 관찰 가능
2. 의사들은 환자로부터 비교적 간편하게 체액을 추출하고 암 전이 여부를 신속하게 확인 가능
 - * 유전체 분석기술의 급속한 발전과 비용 절감이 액체 생체검사의 실용화를 견인



혈액을 이용한 암 연구



Digital Health Care

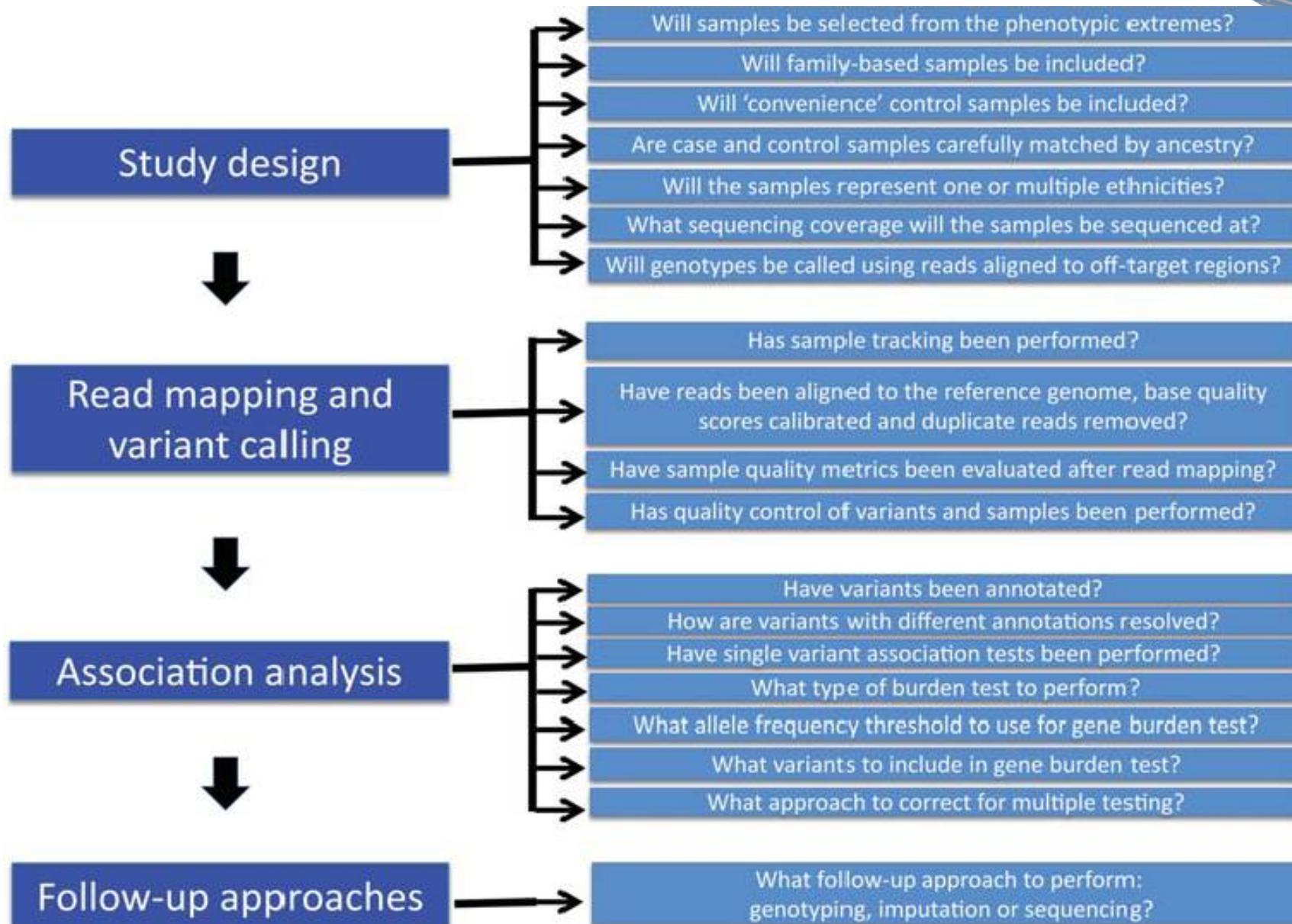


Healthcare Platform

Introduction of NGS



Sequencing study



NGS : Throughput(Coverage & Depth)

- Depth
 - 총 해독량을 대상 영역의 길이로 나눈 값.
 - 대상 영역이 평균적으로 읽어진 횟수를 예측하는 값.
- 예제
 - 사람 Genome 3Gbp
 - 해독량 90Gbp
 - Depth = $90/3 = 30x$
 - 사람의 유전체를 평균 30번 읽음.
 - Exome (Target 54M)
 - 해독량 5.4G
 - Depth $5.4/0.054 = 100x$
 - Target 영역을 100번 읽음.



NGS World

DNA-
seq

Next-Generation

Sequencing

Transcriptome
alternative
splicing

RNA-
seq

Novel Genes
de novo
assembly

whole genome de novo
assembly

Whole Genome
Sequencing

Metagenome

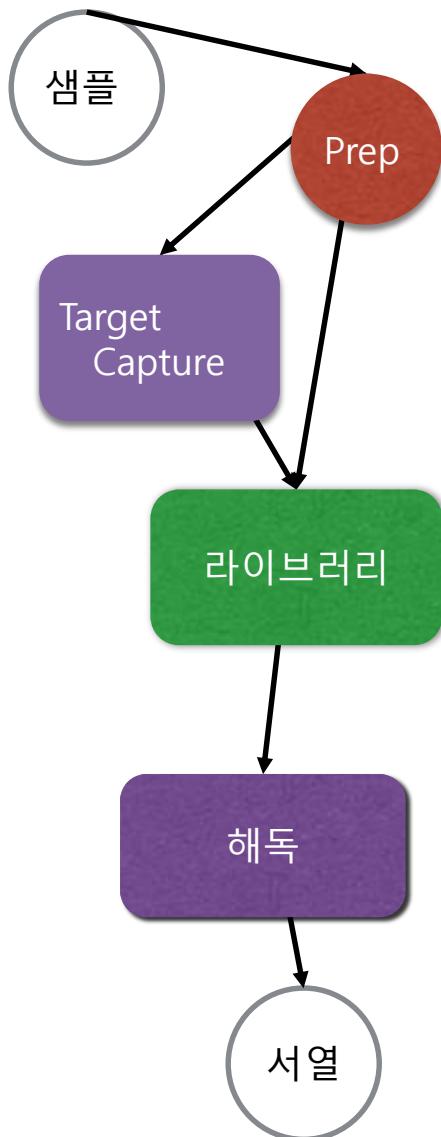
Epigen
ome

ChIP-seq
MeDIP-seq
BiSulfite-seq

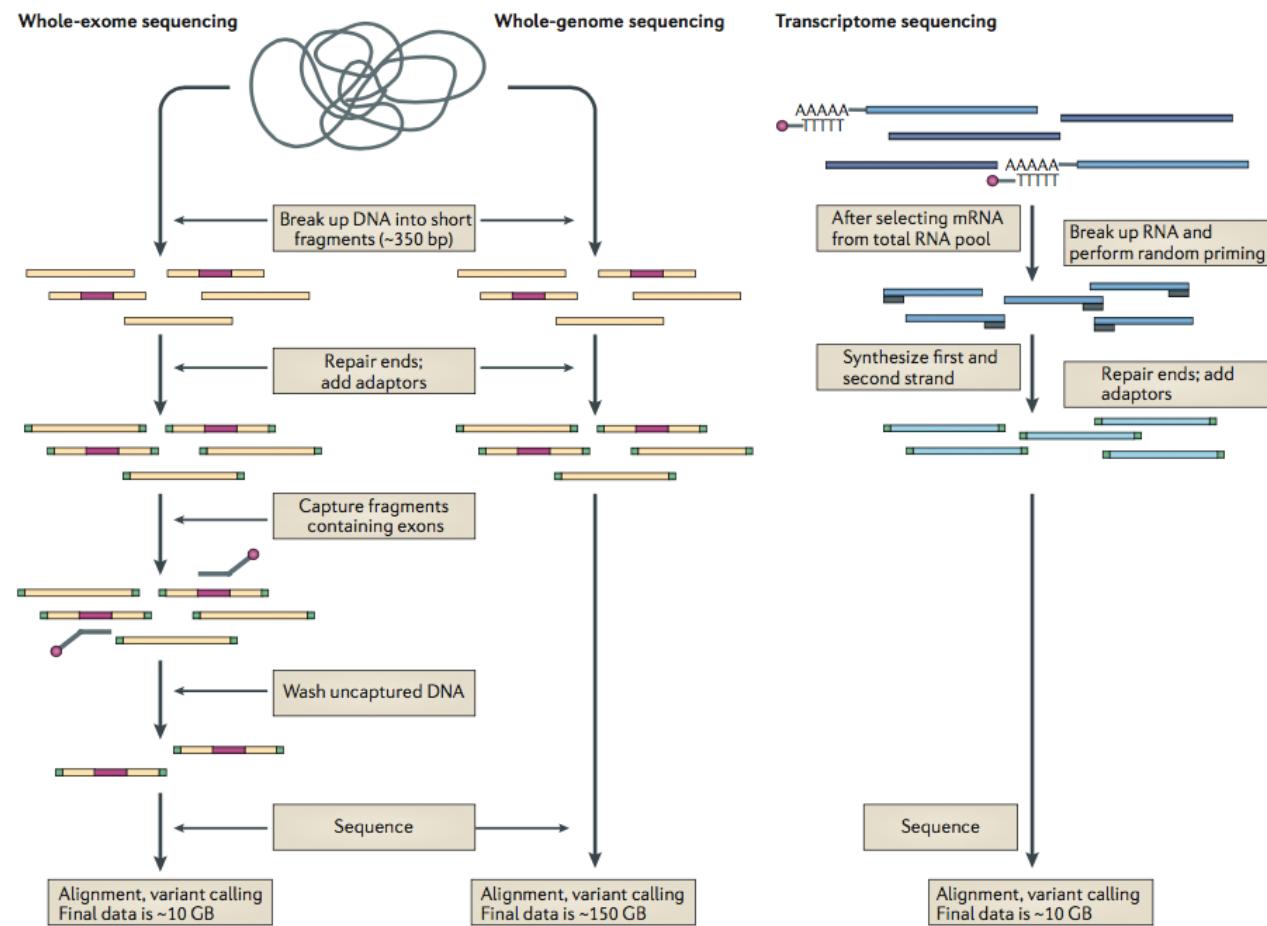
DNA-seq Vs RNA-seq

	RNA-seq	DNA-seq
Methods	Reference-based, de novo assembly	WES, WGS re-sequencing, WGS de novo
Goal	Expression, Differentially Expressed Genes, Novel transcript, Alternative splicing form, Gene fusion	SNPs, Indels, SV
Measure	Mapped Read Count	Base accuracy

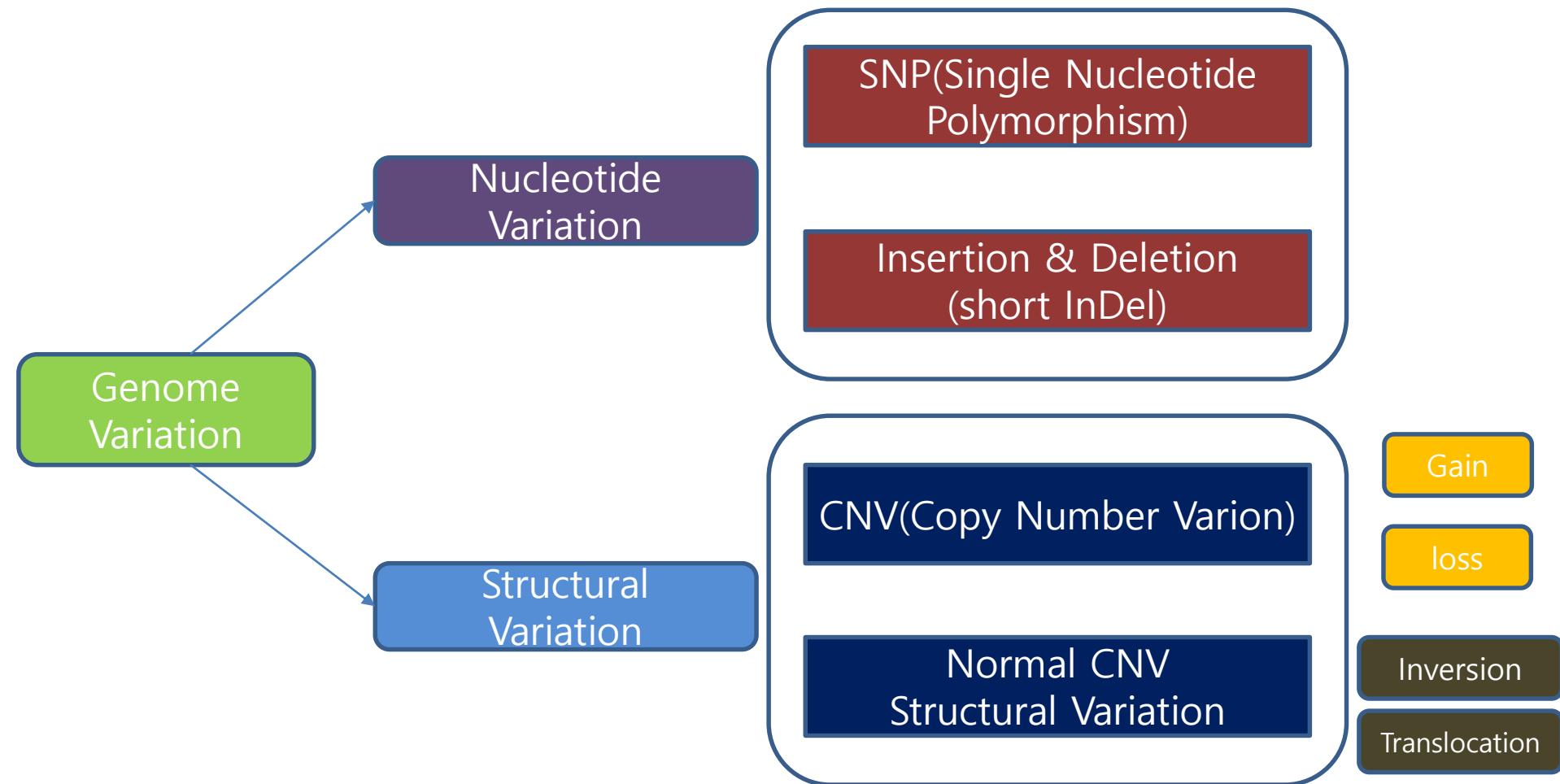
NGS workflow



Simplified workflows for whole-exome, whole-genome and transcriptome sequencing.



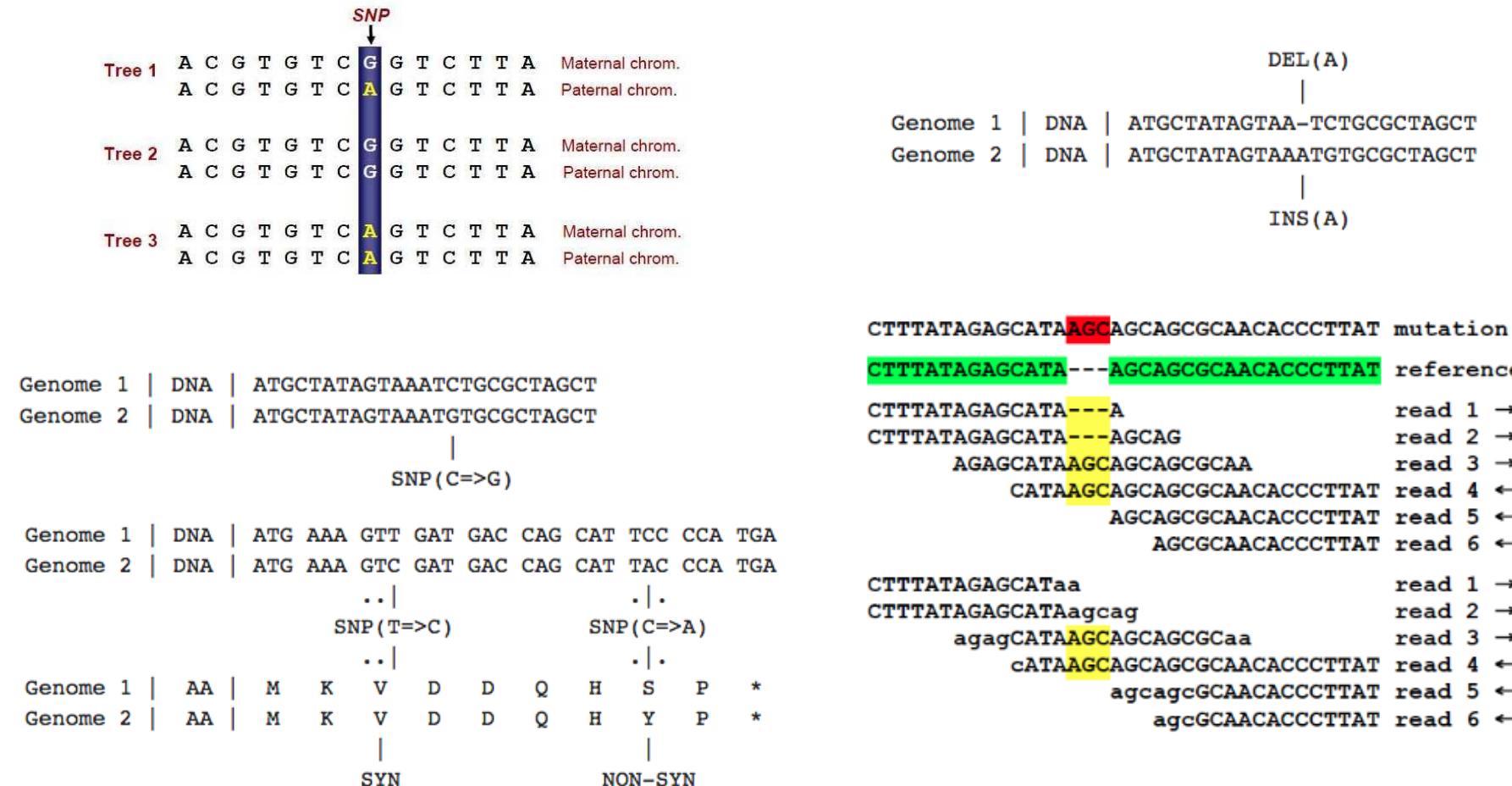
NGS : DNA Work(What is Genetic Variation)



Based on Genotyping Research

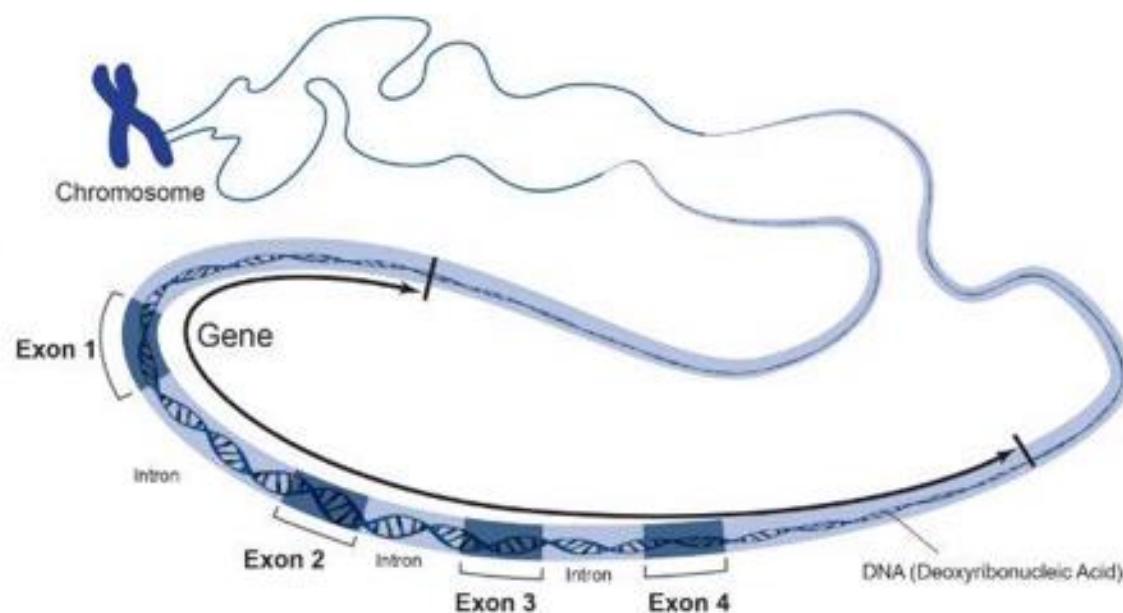
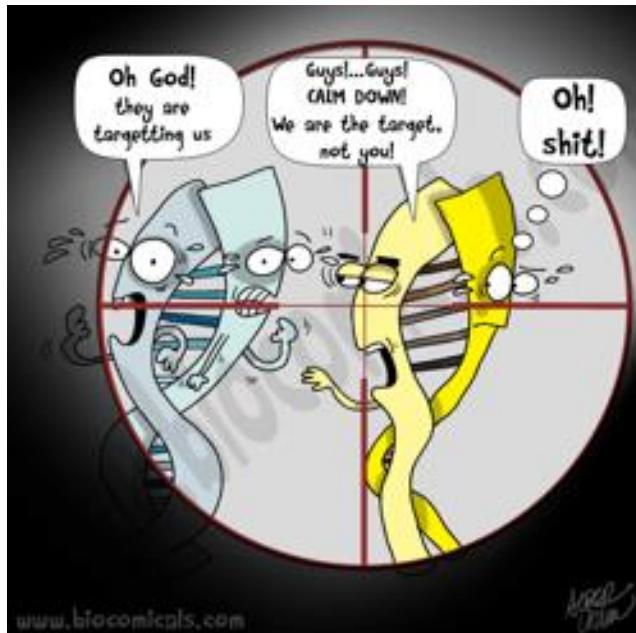
NGS : Nucleotide Variation

Single nucleotide polymorphism (SNP, pronounced as "snip") Insertion (INS) & Deletion (DEL)



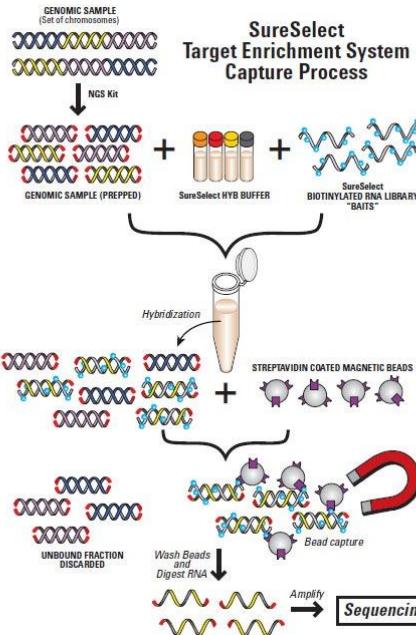
NGS : WES(Whole Exome Sequencing)

- Genome 중 Coding 영역인 Exon 영역만을 해독하는 방법.
- Human Genome 에는 약 180,000개의 Exon 이 존재. (Genome 1%)
- 적은 데이터 생산으로 의미있는 데이터를 찾아내는 것이 가능.

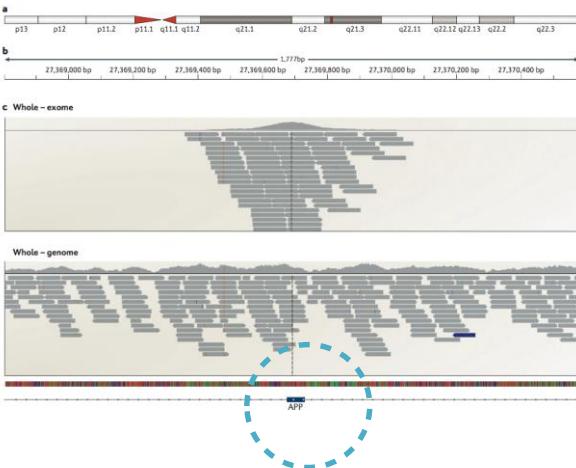


NGS : WES(Whole Exome Sequencing)

Exome Protocol 예제



Exome Sequencing 결과

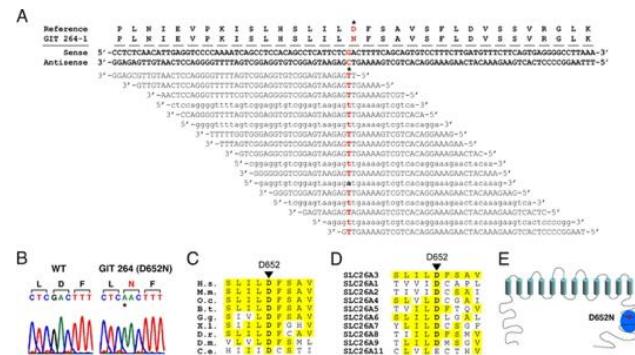


Jose Bras et al., Nature Reviews Neuroscience 13, 453-464 (July 2012)

Exome Sequencing 응용

Genetic diagnosis by whole exome capture and massively parallel DNA sequencing

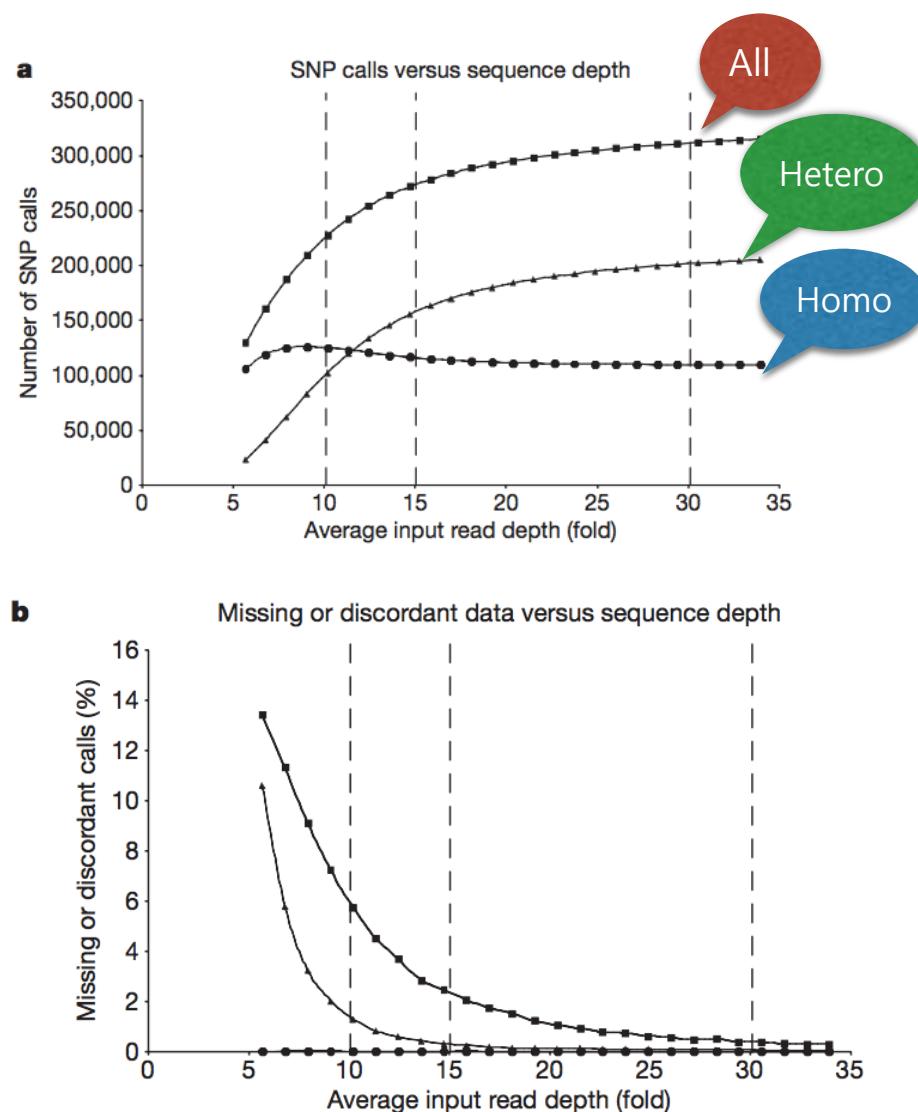
Murim Choi^a, Ute I. Scholl^a, Weizhen Ji^a, Tiewen Liu^a, Irina R. Tikhonova^b, Paul Zumbo^b, Ahmet Nayir^c, Aysin Bakkaloglu^d, Seza Ozan^d, Sami Sanjad^d, Carol Nelson-Williams^e, Anita Farhi^e, Shrikant Mane^b, and Richard P. Lifton^{a,1}



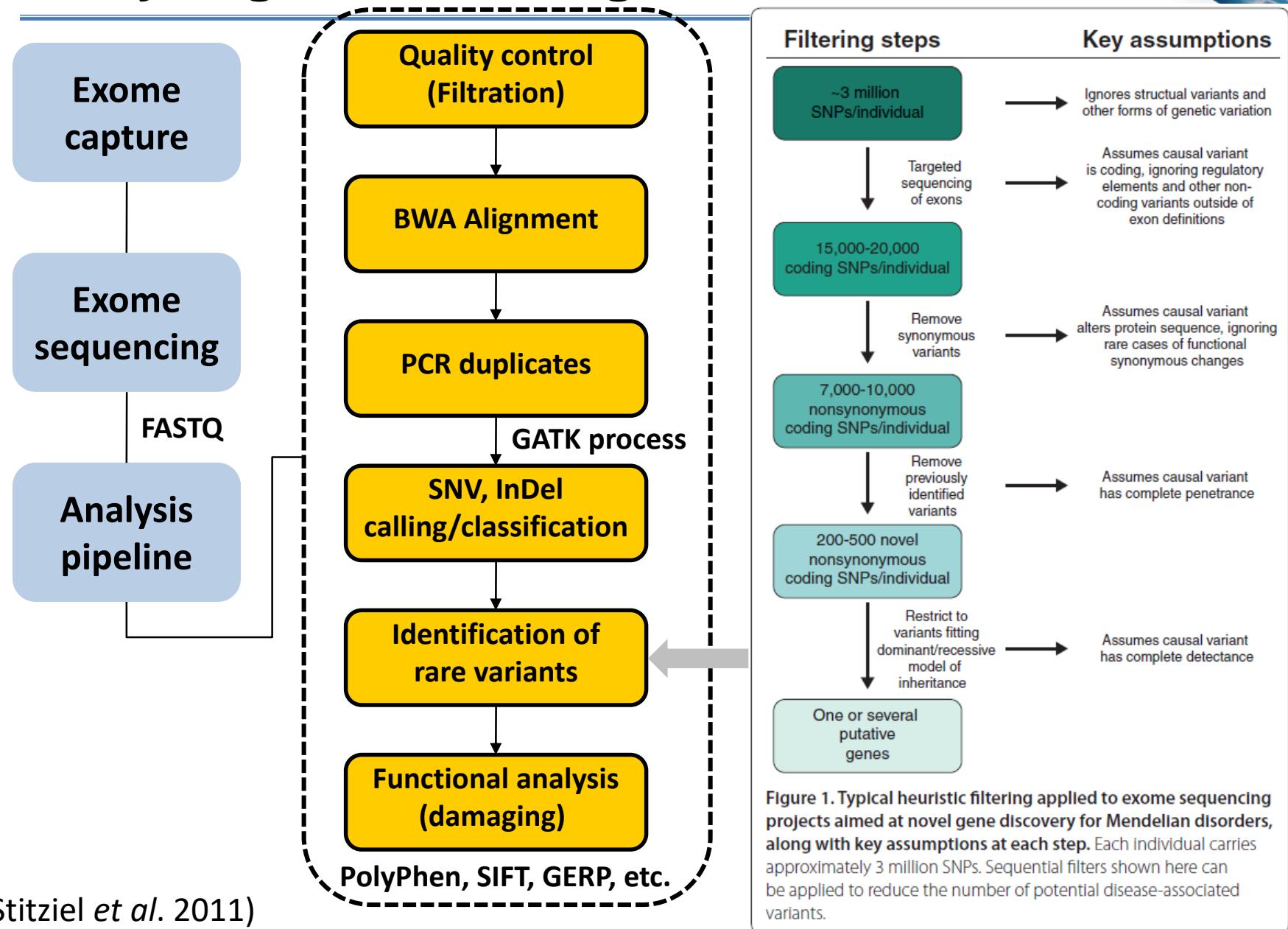
Homozygous missense mutation at highly conserved position in SLC26A3 in GIT 264-1

NGS : Throughput(Coverage & Depth)

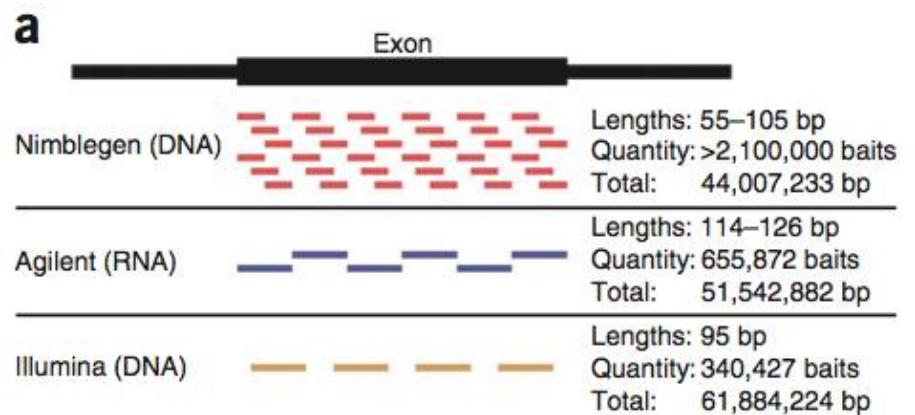
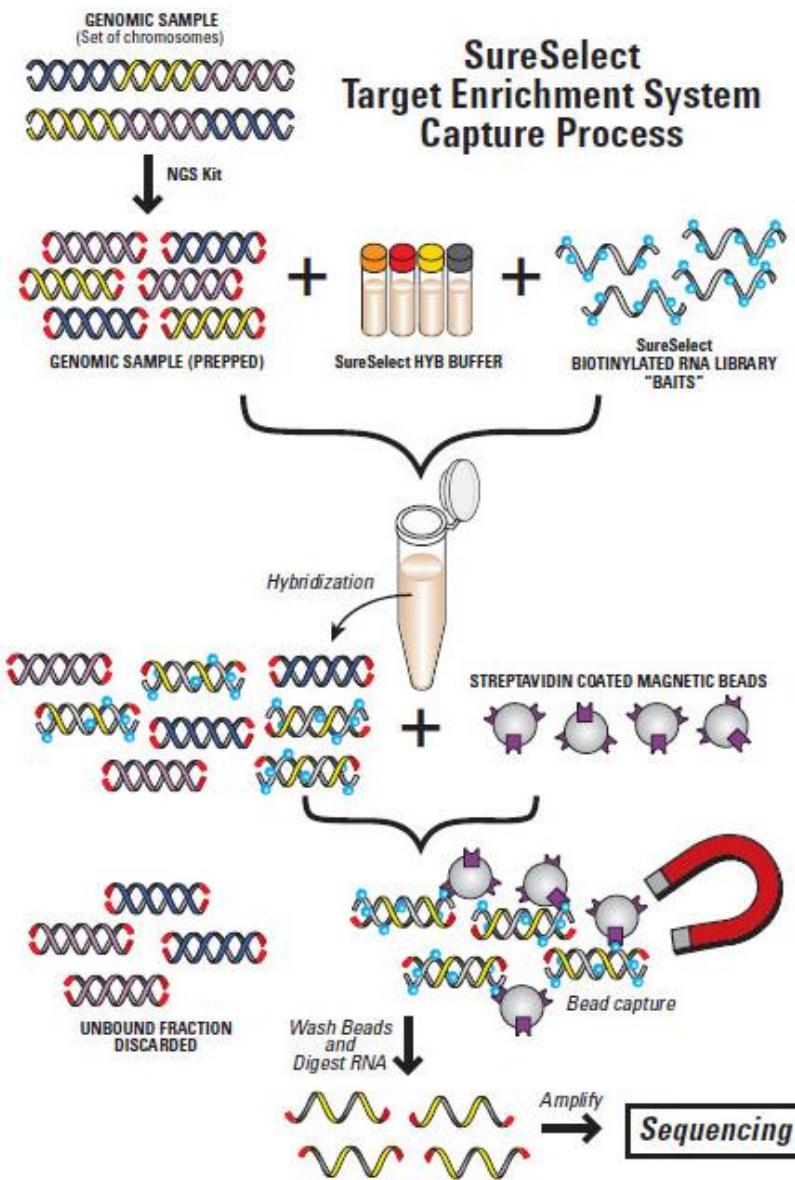
- NGS의 문제점
 - NGS 의 Base Calling 에러는 10^3 수준으로 높음.
 - Sequencing 이 이루어 지지 않는 영역이 존재할 가능성이 있음.
 - 짧은 서열 길이에 따른 Alignment 오류 가능성
- 해결 방법
 - 동일한 위치를 반복적으로 읽는 방법으로 해결 가능
 - Homo SNP = 5x 이상.
 - Hetero SNP = 30x 이상.



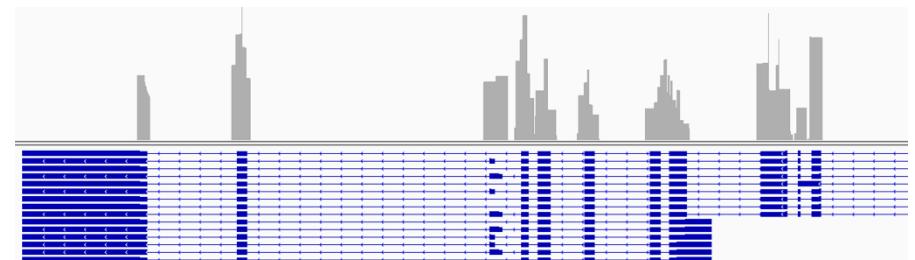
Analyzing variant using Exome



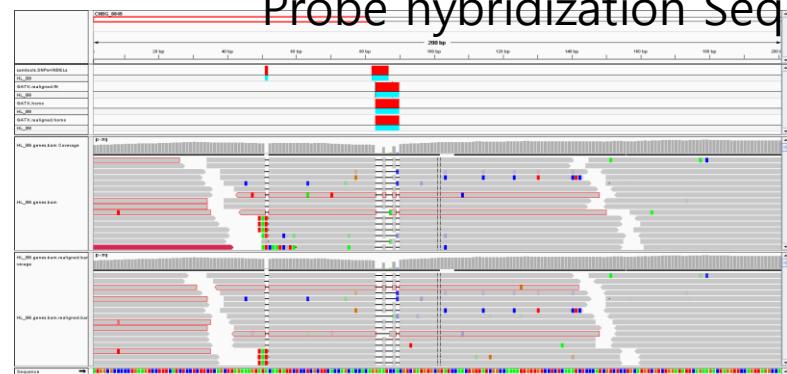
Exome Capture



Amplicon Sequencing

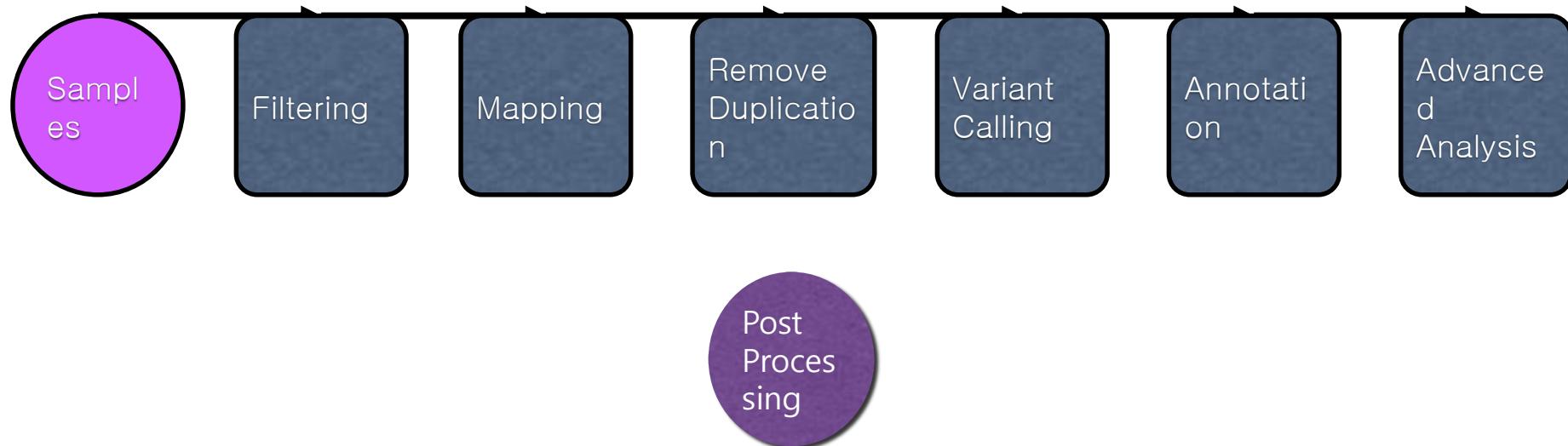


Probe hybridization Sequencing



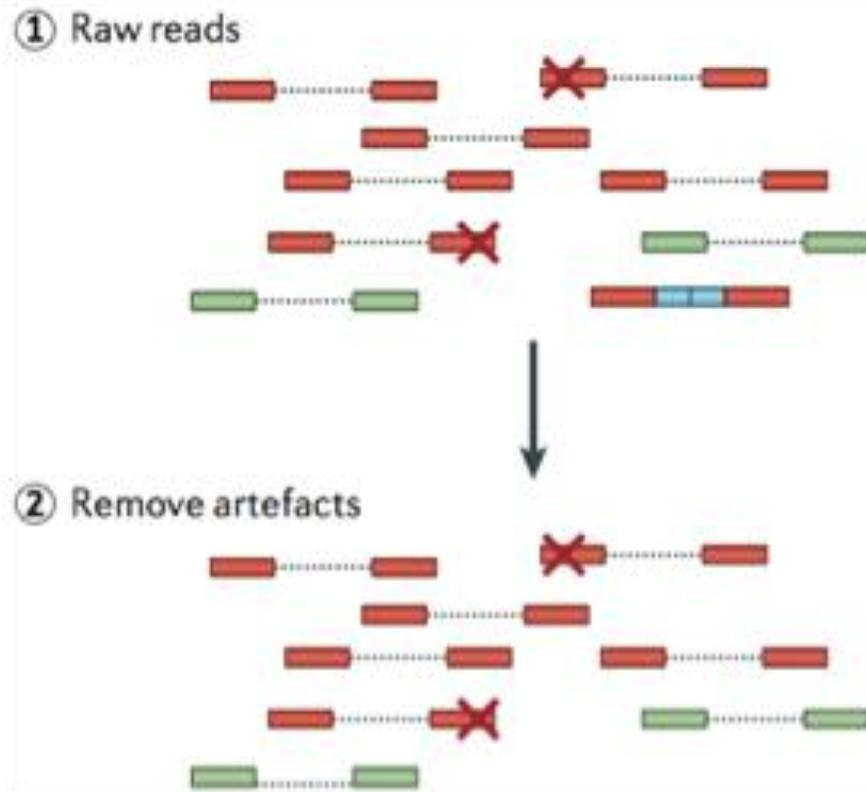
NGS : DNA-Seq Variant Calling

- Variant Calling
 - 해독 데이터에서 SNV, InDel 을 찾아내는 분석 방법
 - DNA 데이터를 분석하는 가장 기본적인 방법.
- Analysis Process



NGS : Data Filtering

- 분석에 영향을 줄 수 있는 낮은 품질의 해독 데이터를 제거하는 작업.
- 생산된 데이터의 품질이 좋을 경우 Filtering 을 진행하지 않는 경우도 있음.
- 제거 대상
 - Improving mapping accuracy & calling accuracy
 - Removing artifacts
 - Sequencing adaptor
 - Low quality reads



```
@HWI-ST621:439:D1D3WACXX:3:1101:3170:2170 1:N:0:GGCTAC
GTGGTTTCGTTAACCTGTTGAATCAAAGCCCTGTTCGATCCAACACAGTTGGACTTGTGAAACTGTCGGTGAAACTGGTCCAGACAACCTCGGATT
+
@=?DDDAD?CBDHHEH>EH?<@CB<FFEAE?;E??CF??DBGCGE@DDF99.=<1=CG>>GEB2;;<?7@;>=8A;?CC@C#####

```

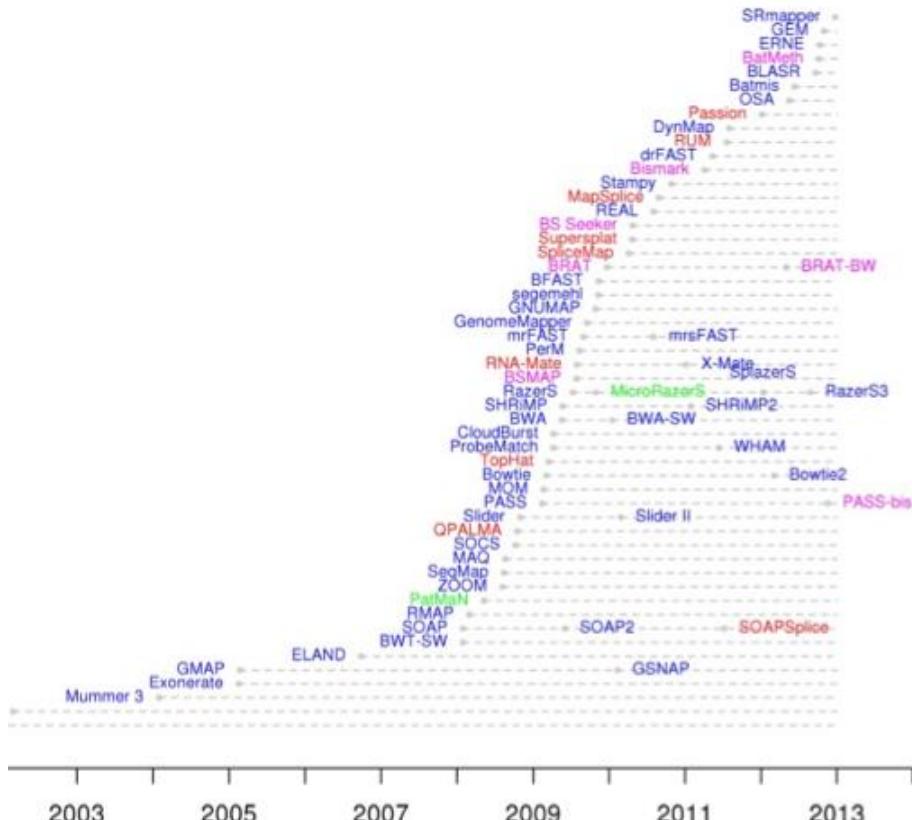


```
@HWI-ST621:439:D1D3WACXX:3:1101:3170:2170 1:N:0:GGCTACC
GTGGTTTCGTTAACCTGTTGAATCAAAGCCCTGTTCGATCCAACACAGTTGGACTTGTGAAACTGTCGGTGAAACT
+
@=?DDDAD?CBDHHEH>EH?<@CB<FFEAE?;E??CF??DBGCGE@DDF99.=<1=CG>>GEB2;;<?7@;>=8A;?CC@C
```

NGS : Data Mapping

- Filtering 된 High Quality Read를 참조 유전체에 정렬하여 위치를 찾는 단계
- Application
 - Blast
 - bwa mem method
 - de facto standard
 - Issac
 - Illumina에서 개발한 Mapper
 - 빠른 속도가 장점

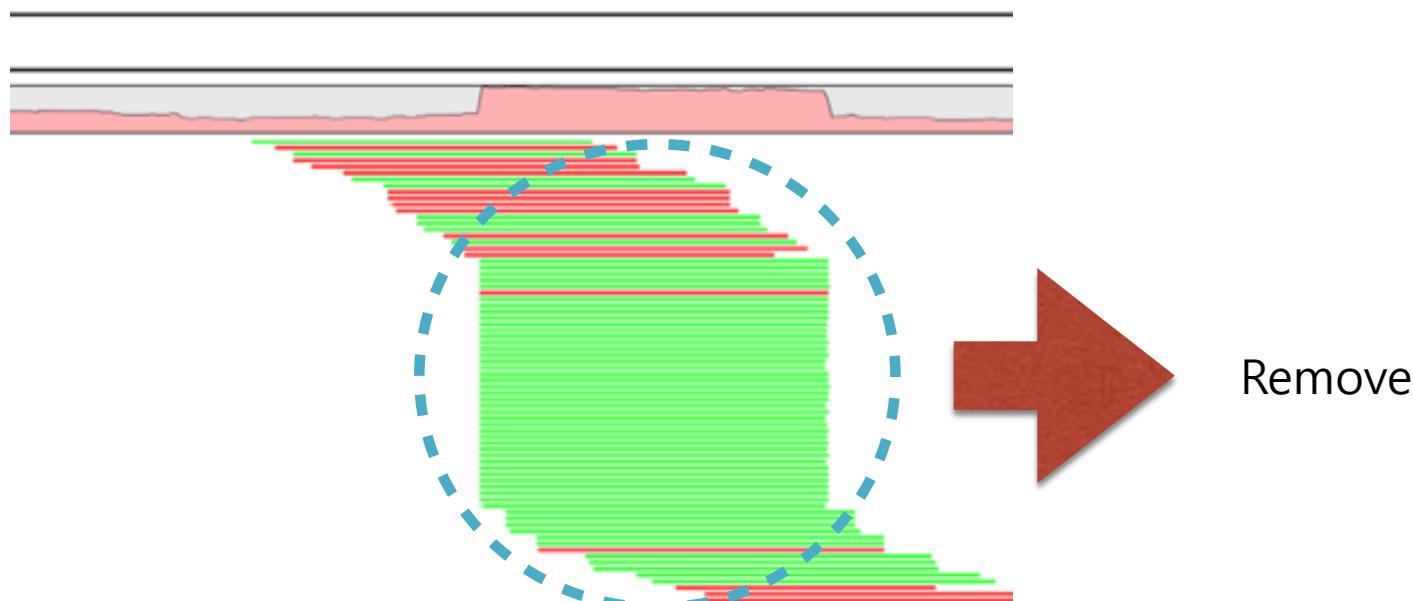
다양한 Mapping 도구



NGS : Remove Duplication

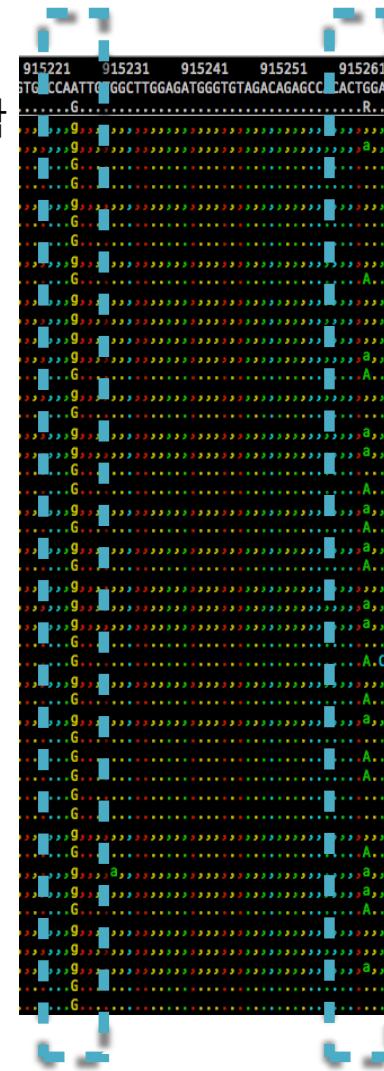
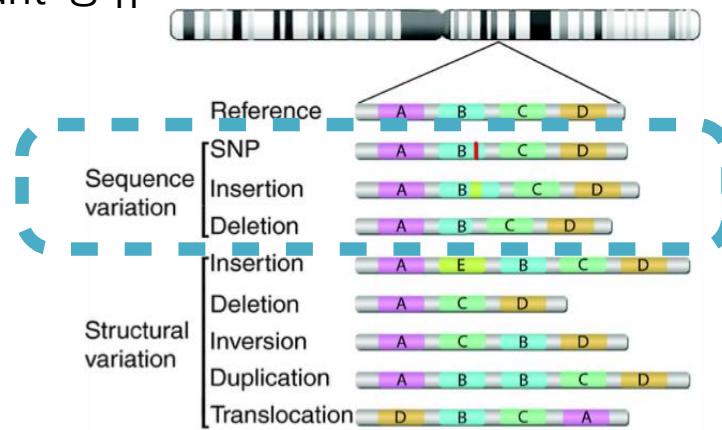
- 라이브러리 준비과정에서 특정 유전자 조각이 이상증폭되어 중복으로 나타나는 경우 분석 결과에 영향을 주게 됨.
- 이상 증폭된 DNA 조각이 Error를 가지고 있을 경우 실제 Variant Calling 단계에서 Calling 될 수 있음.
 - Low depth cutoff (>5)를 넘어설 수 있음.
- 이를 예방하기 위해서 Mapping 정보를 이용하여 동일한 위치에 Mapping 된 Read 중 가장 품질 좋은 것 만을 선택

Duplication 예시



NGS : Variant Calling

- Mapping 결과를 해석하여 SNP 영역을 찾는 작업
- Calling 정확도를 높이기 위해 Cutoff 을 적용함
 - Low depth cutoff (>5 depth)
- Applications
 - Samtools, GATK, Isaac
- Calling 알고리즘마다 다른 결과가 나오지만, 최근에는 거의 평준화 됨.
- Protocol
 - GATK Best Practice 가 가장 널리 사용됨.
- Variant 종류



GATK

The Genome Analysis Toolkit or GATK is a software package developed at the Broad Institute to analyse next-generation resequencing data. The toolkit offers a wide variety of tools, with a primary focus on variant discovery and genotyping as well as strong emphasis on data quality assurance. Its robust architecture, powerful processing engine and high-performance computing features make it capable of taking on projects of any size.

[Learn more...](#)

About Overview of the GATK and the people behind it

Guide Detailed documentation, guidelines and tutorials

Community Forum for questions and announcements

Events Materials from live and online events

Latest version: 2.7-1
Release Notes
[Download now!](#)
For-profit users: click here

Samtools

SAM (Sequence Alignment/Map) format is a generic format for storing large nucleotide sequence alignments. SAM aims to be a format that:

- Is flexible enough to store all the alignment information generated by various alignment programs;
- Is simple enough to be easily generated by alignment programs or converted from existing alignment formats;
- Is compact in file size;
- Allows most of operations on the alignment to work on a stream without loading the whole alignment into memory;
- Allows the file to be indexed by genomic position to efficiently retrieve all reads aligning to a locus.

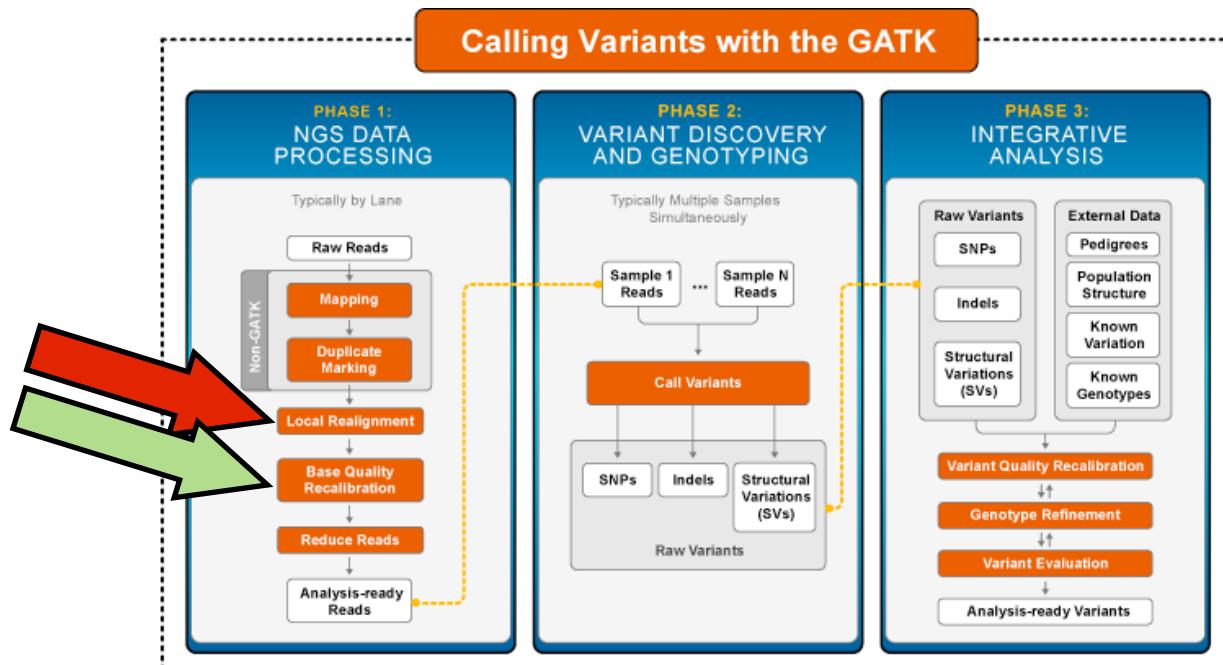
SAM Tools provide various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format.

SAM tools is hosted by [SourceForge.net](#). The project page is [here](#). The source code releases are available from the [download page](#). You can check out the most recent source code from the [GitHub project page](#) with:

```
git clone git://github.com/samtools/samtools.git
```

NGS : Variant Calling(GATK)

- Broad Institute에서 개발한 Genome Analysis Tool Kit (GATK)를 이용한 Variant Calling Method
- de facto standard로 사용되고 있음.
- 2단계의 Post Processing 과정을 담고 있음.
 - Local Realignment
 - Base Recalibration
- 2가지 Variant Calling
 - Unified Genotyper
 - Haplotype



NGS : Annotation

Filtering

- 질병 원인 변이 분석을 위해 Calling 된 Variant 중 일 반적으로 나타나는 변이를 제거하는 단계
- 방법
 - Depth 를 이용한 filtering
 - Common Variant Filtering
- Common Variant Filtering
 - 일반적으로 나타나는 변이는 질병의 원인이 아닌 것으로 보고 제거
 - Common Variant (인구의 10% 이상에서 나타나는 변이)
 - 인종별, 민족별 변이 특화된 DB 를 이용하는 것이 좋은 결과를 낼 수 있다.
- 변이 데이터 베이스
 - dbSNP (dbSNP asia),
 - 1000 Genome (1000 Genome Asia)
 - KPGP, In-House

Annotation

- 질병 원인 변이 분석을 위해서 기존에 알려진 원인 유전자에 대한 정보를 더하는 작업.
- 방법
 - Database 기반의 Annotation
 - Mutation 의 영향력 예측을 통한 Annotation
- 질병 데이터베이스
 - Cosmic
 - OMIM
 - ClinVar
 - My Cancer Genome
- Mutation 의 영향력
 - Non-sense mutation, Missense mutation
 - Polyphen

Advantages of RNA-seq

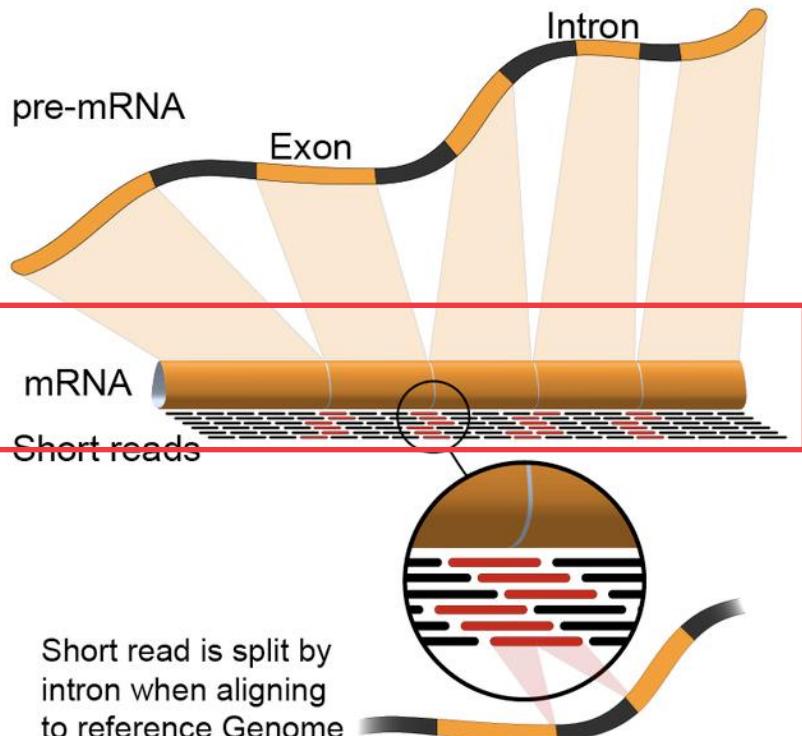
Table 1 | Advantages of RNA-Seq compared with other transcriptomics methods

Technology	Tiling microarray	cDNA or EST sequencing	RNA-Seq
Technology specifications			
Principle	Hybridization	Sanger sequencing	High-throughput sequencing
Resolution	From several to 100 bp	Single base	Single base
Throughput	High	Low	High
Reliance on genomic sequence	Yes	No	In some cases
Background noise	High	Low	Low
Application			
Simultaneously map transcribed regions and gene expression	Yes	Limited for gene expression	Yes
Dynamic range to quantify gene expression level	Up to a few-hundredfold	Not practical	>8,000-fold
Ability to distinguish different isoforms	Limited	Yes	Yes
Ability to distinguish allelic expression	Limited	Yes	Yes
Practical issues			
Required amount of RNA	High	High	Low
Cost for mapping transcriptomes of large genomes	High	High	Relatively low

RNA-Seq: a revolutionary tool for transcriptomics. Zhong Wang, Mark Gerstein and Michael Snyder, NATURE REVIEWS, 2009, 10:57

NGS : RNA-seq

모든 RNA의 서열을 읽어서 그 비율을 비교하는 것

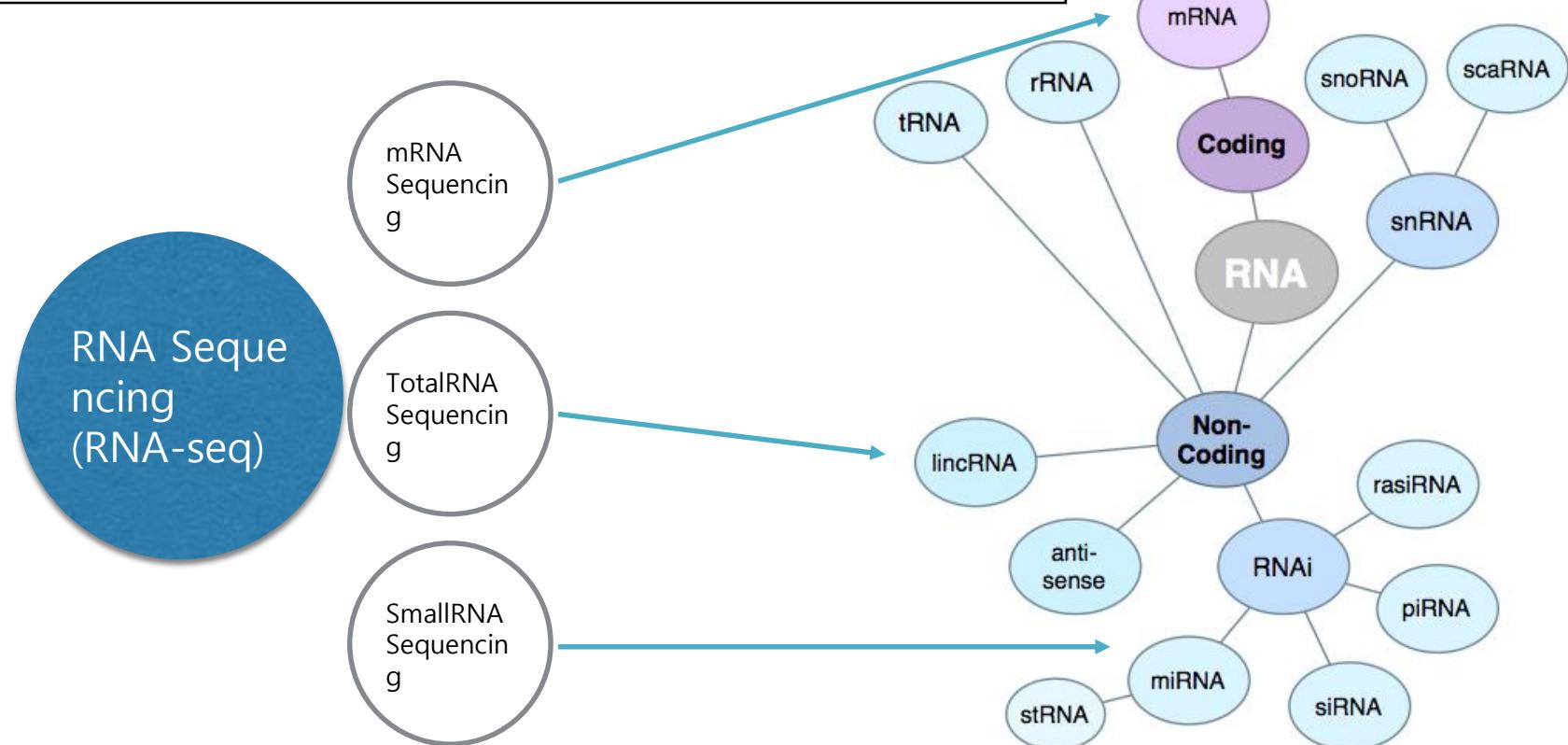


<http://en.wikipedia.org/wiki/RNA-Seq>

-
- a. Relative expression level in sample
 - b. Differentially expressed gene
 - c. Identification of alternative spliced transcripts
 - d. Prediction of novel transcripts
 - e. Gene Fusion

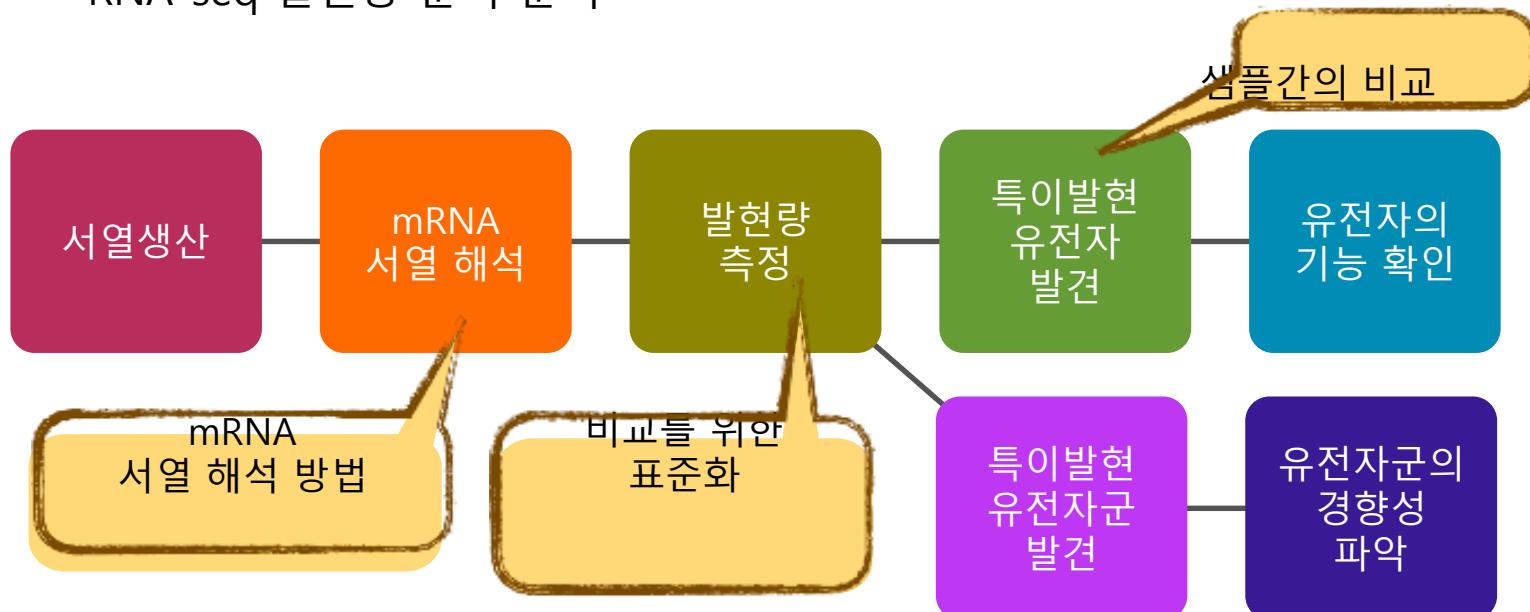
NGS : RNA-seq

RNA 종류에 따라 해독하는 방법이 다릅니다.



NGS : RNA-seq

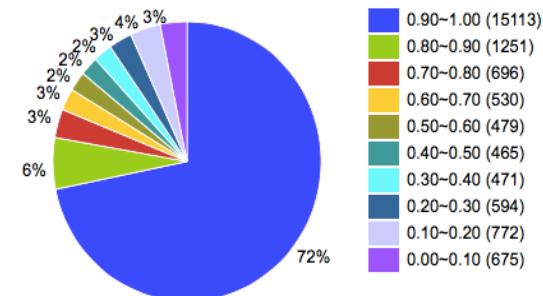
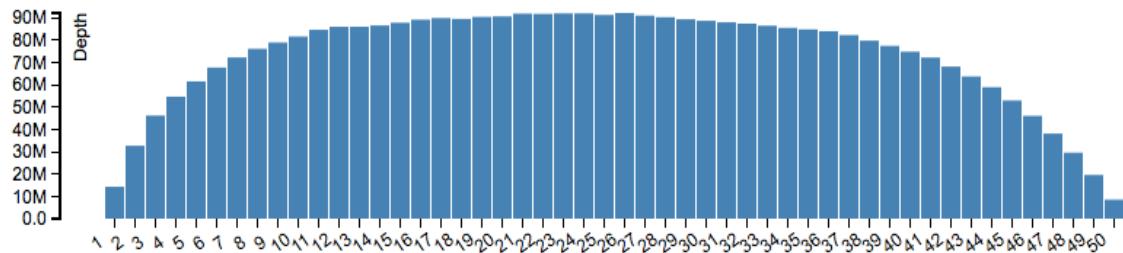
- RNA-seq 은 Spliced Mapper 가 필요합니다.
 - TopHat, STAR, GSNAP
 - Splicing Mapping 는 Mapping 오류가 발생할 확률이 상대적으로 높음.
- RNA-seq 분석은 DNA-seq 에 비해서 변수가 아주 많습니다.
- RNA-seq 발현량 분석 순서



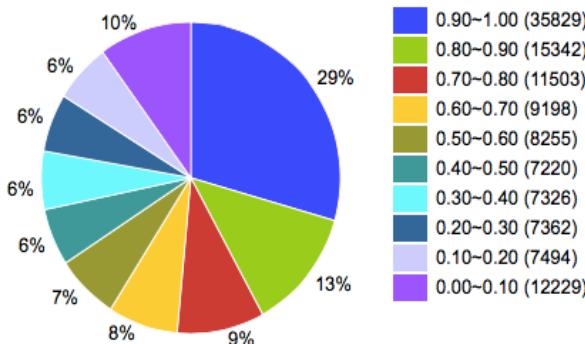
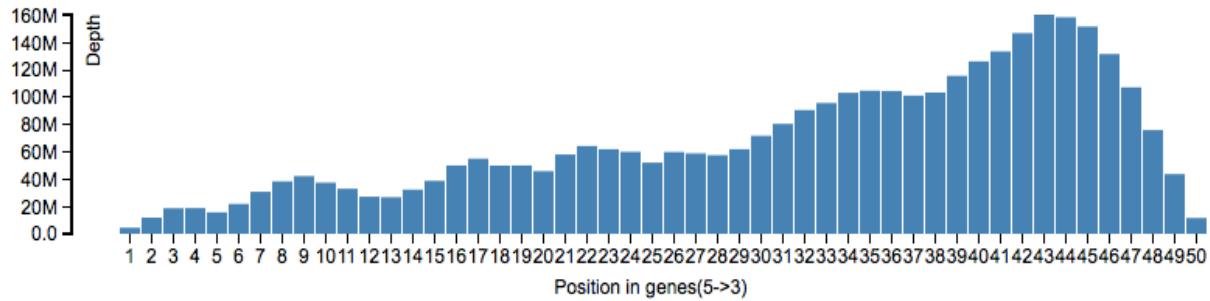
NGS : RNA Quality

- RNA 분석에서 가장 중요한 것은 RNA-quality.

High Quality RNA



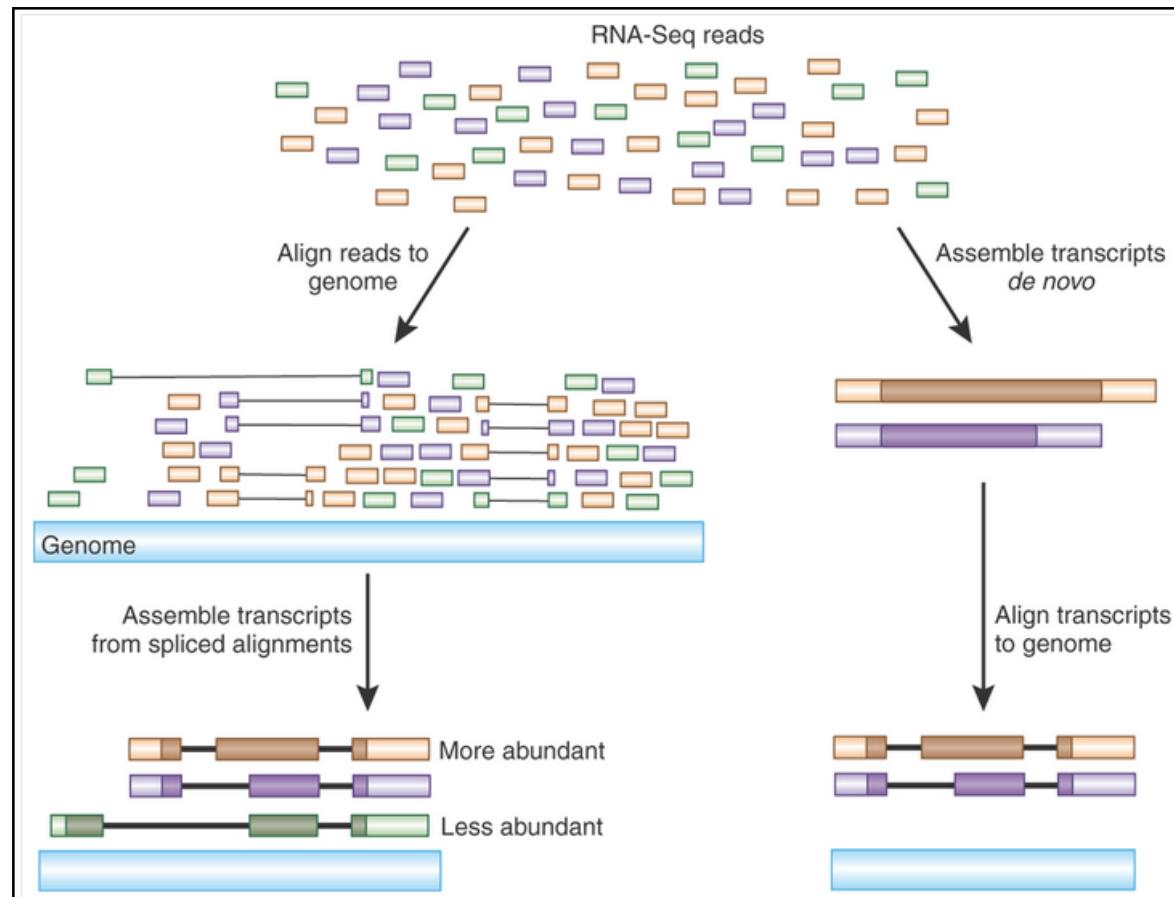
Low Quality RNA



NGS : RNA-seq

Reference-base

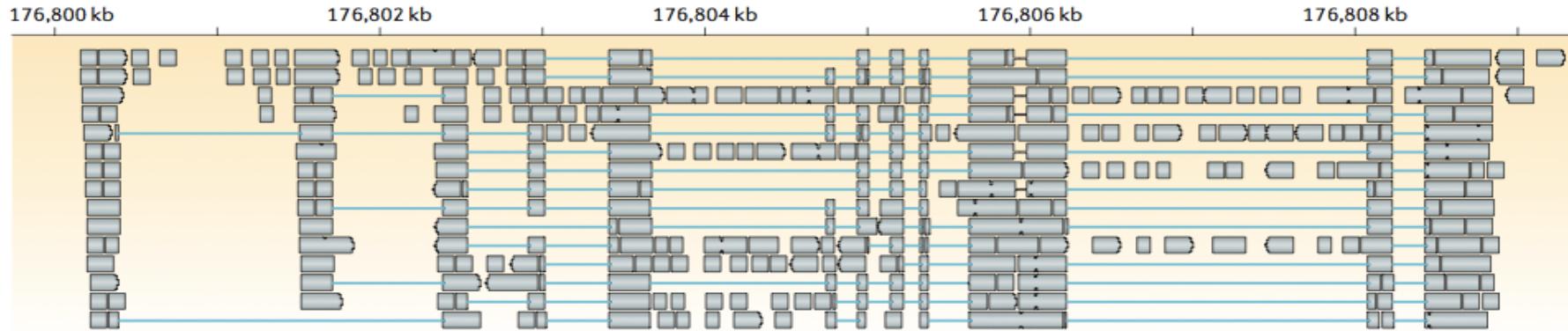
De novo



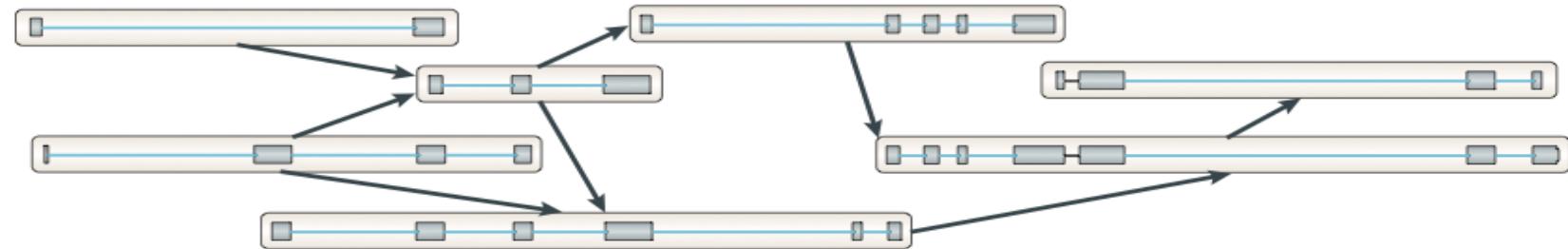
Advancing RNA-Seq analysis, Brian J Haas & Michael C Zody,
Nature Biotechnology 28, 421–423 (2010) doi:10.1038/nbt0510-421

NGS : RNA-seq(Reference based)

a Splice-align reads to the genome



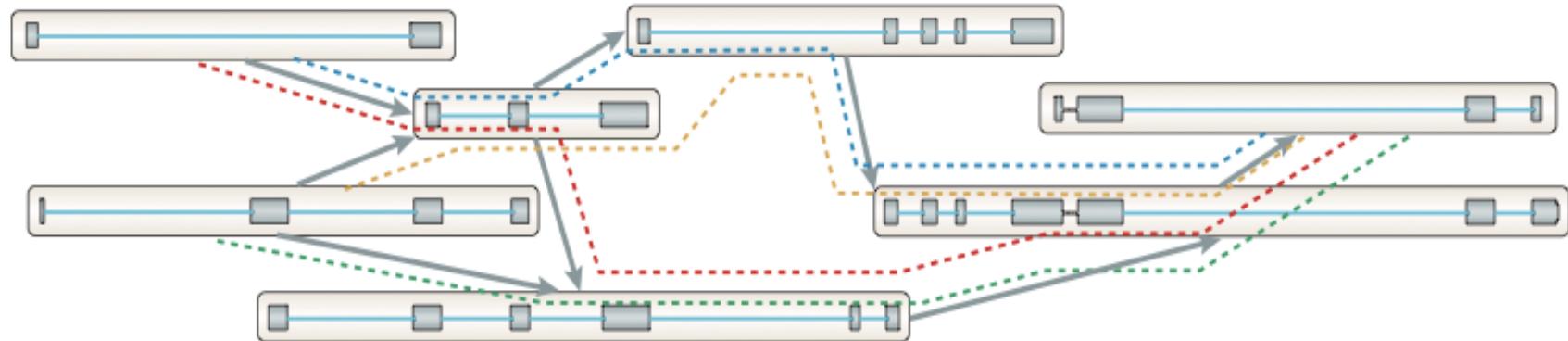
b Build a graph representing alternative splicing events



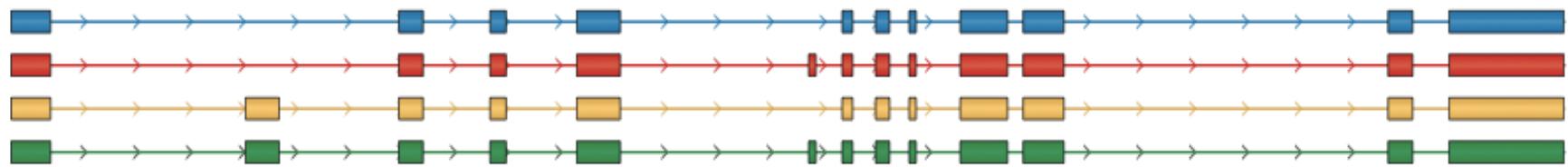
Martin JA, Wang Z. Next-generation transcriptome assembly. Nat Rev Genet. 2011 Sep 7;12(10):671-82.

NGS : RNA-seq(Reference-based)

c Traverse the graph to assemble variants

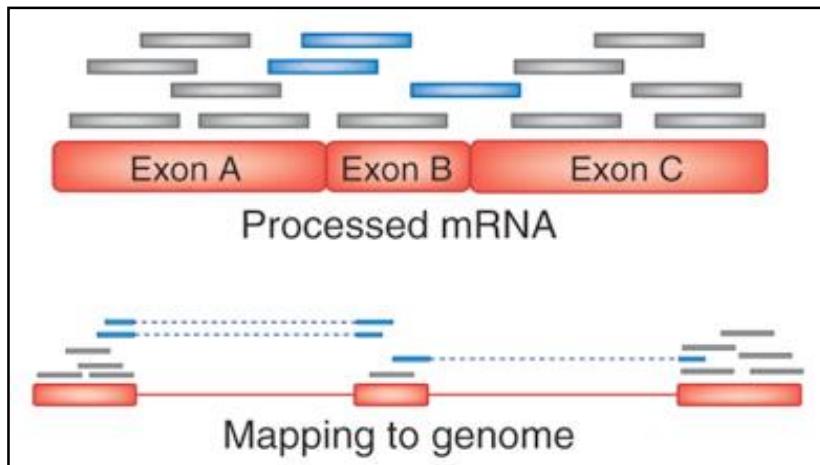


d Assembled isoforms

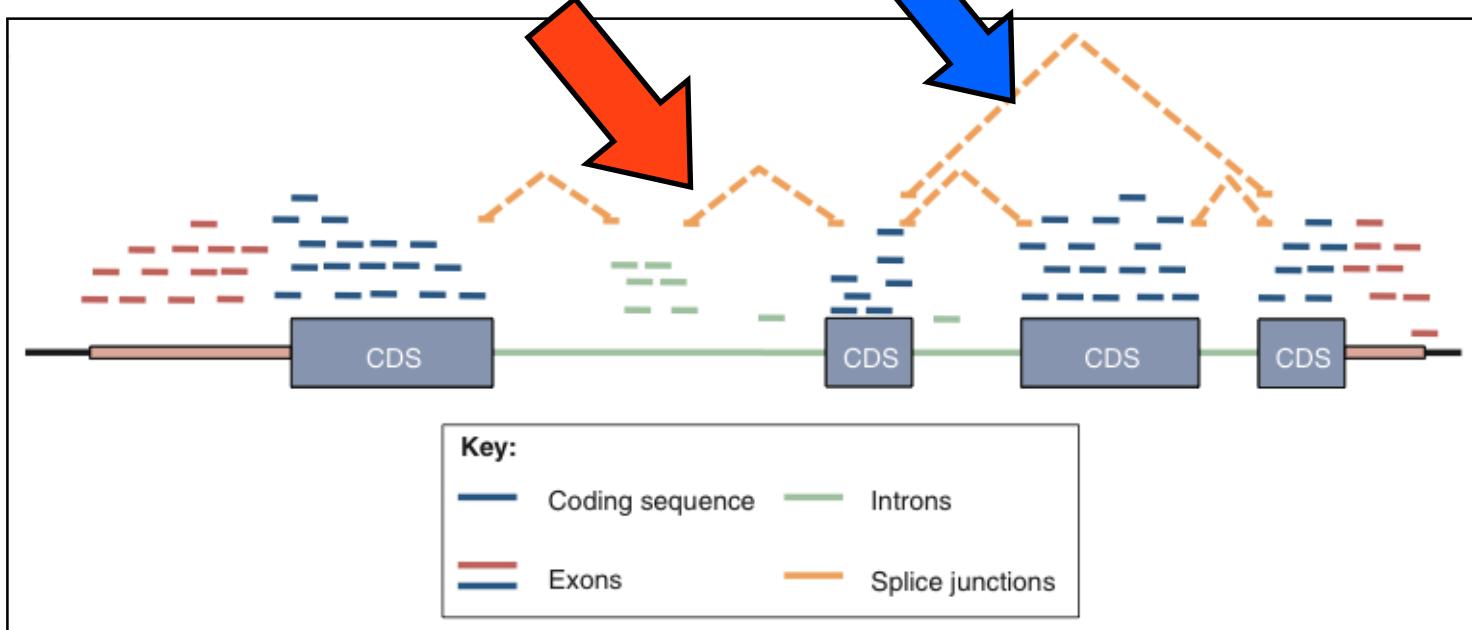


Martin JA, Wang Z. Next-generation transcriptome assembly. Nat Rev Genet. 2011 Sep 7;12(10):671-82.

NGS : RNA-seq(Reference-based)



Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. *Genome Biol.* 2010;11(12):220.



Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. *Nat Biotechnol.* 2009 May;27(5):455-7.

RNA-Sequencing data (발현 값)

RPKM Example

Gene A 600 bases

Gene B 1100 bases

Gene C 1400 bases

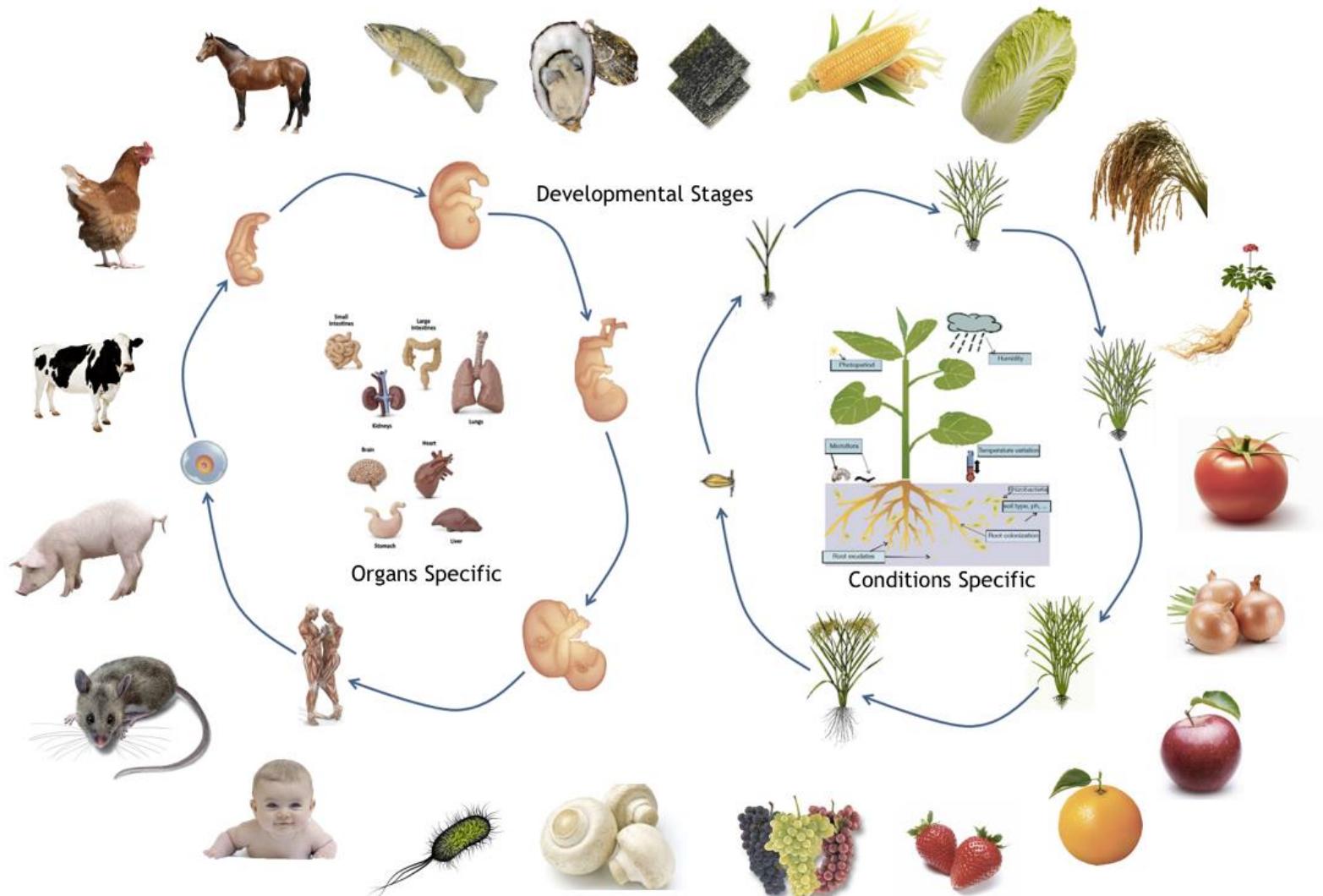
$$\text{RPKM} = 12/(0.6*6) = 3.33$$

$$\text{RPKM} = 24/(1.1*6) = 3.64$$

$$\text{RPKM} = 11/(1.4*6) = 1.31$$



NGS : RNA-seq(Aplications)



NGS : RNA-seq

TotalOmics
for Bio Institute

Home Overview Samples Expression DEG DEG consensus GO Clustering Pathway Variant Fusion
Target Genes Help

Transcriptome Analysis Report (Reference-based)

Reference Genome

Species	Homo sapiens
Source	Ensembl
Genes	63,194
Release	72
Transcripts	213,141

Data and Analysis

Samples: TRS(TBL_TRS), TSS(TBL_TSS), SSS(TBL_SSS), TRS2(TBL_TRS2), ORG(TBL_ORG)

- TRS vs TSS (trs_tss_trs_tss)
- TRS vs SSS (trs_tss_trs_sss)
- TSS vs SSS (trs_tss_tss_sss)
- TRS vs TRS2 (trs_tss_trs2_trs2)
- TRS vs ORG (trs_tss_trs_ogr)
- TRS2 vs ORG (trs2_tss_trs2_ogr)

mailto:kotomeda@theragen.kr, +82-31-888-9446
Copyright © 2013 Theragen BIO Institute. All Rights Reserved.
2nd Floor, B-dong, AICT, Gwanganji Techvalley, IJ-dong, Suwon, gyeonggi-do, Republic of Korea (443-270)

Theragen TotalOmics
for Bio Institute

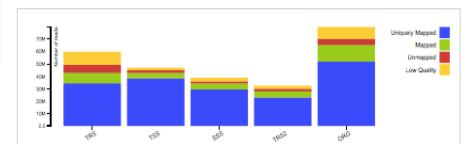
Home **Overview** Samples Expression DEG DEG consensus GO Clustering Pathway Variant Fusion
Target Genes Help

Overview

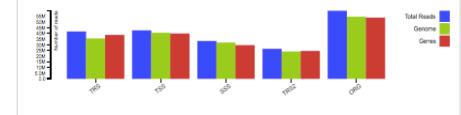
Files

- Gene Expression [Excel]
- Transcript Expression [Excel]
- Geneset [GTF]
- CDNA Sequence [Fasta]

Sequencing



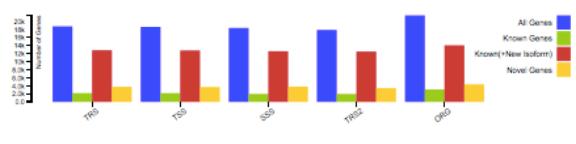
Read Alignment



Gene Coverage



Expression

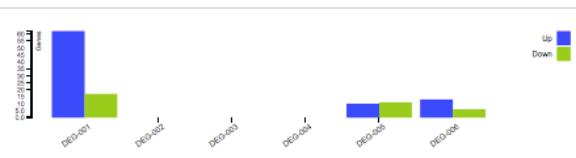


Sample Correlation

Pearson correlation coefficient between FPKM of genes of samples

	TRS (TBL_TRS)	TSS (TBL_TSS)	SSS (TBL_SSS)	TRS2 (TBL_TRS2)	ORG (TBL_ORG)
TRS	-	0.903	0.910	0.998	0.880
TSS	0.903	-	0.949	0.907	0.827
SSS	0.910	0.949	-	0.914	0.855
TRS2	0.998	0.907	0.914	-	0.878
ORG	0.880	0.827	0.855	0.878	-

DEG



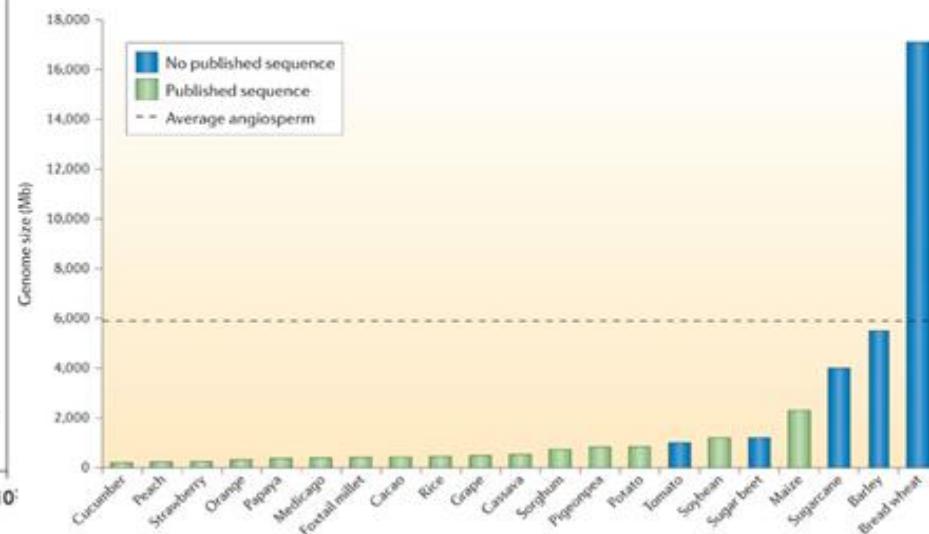
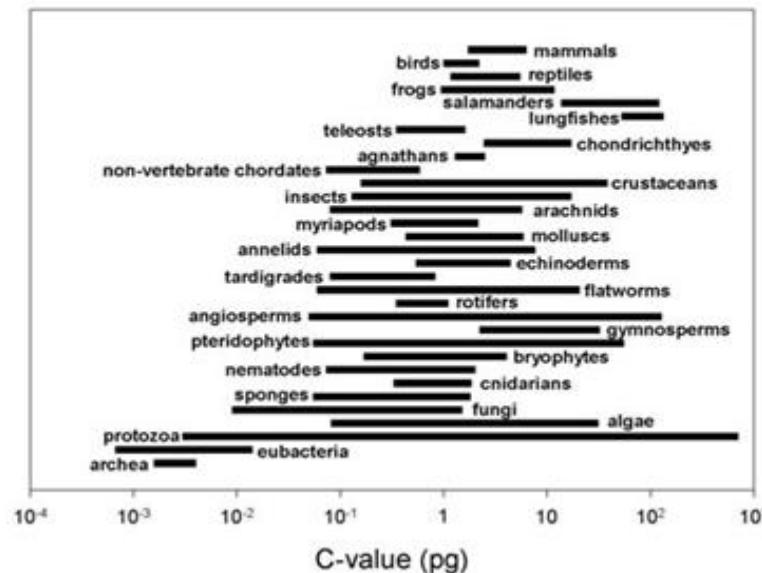
No	Group 1 (G1)	Group 2 (G2)	Genes	Excel	
			Sum	Up (G2 only)	Down (G1 only)
1	TRS	TSS	79	62 (7)	17 (12)
2	TRS	SSS	0	0 (0)	0 (0)
3	TSS	SSS	0	0 (0)	0 (0)
4	TRS	TRS2	0	0 (0)	0 (0)

Prediction of Genome size

- 동, 식물의 유전체 사이즈는 종별로 매우 다양하며, 그 편차도 매우 큼
- 정확한 유전체 사이즈 및 그 복잡도를 파악하지 못하면 해독소요 비용이 매우 높아지게 됨

Animal genome size average

(1pg = 978Mb)



Genome Assembly

De-novo

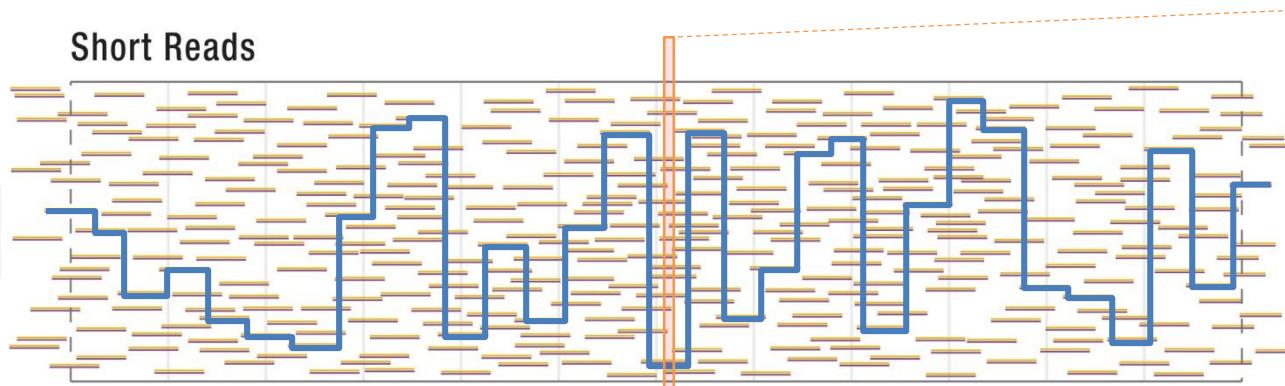


Reference

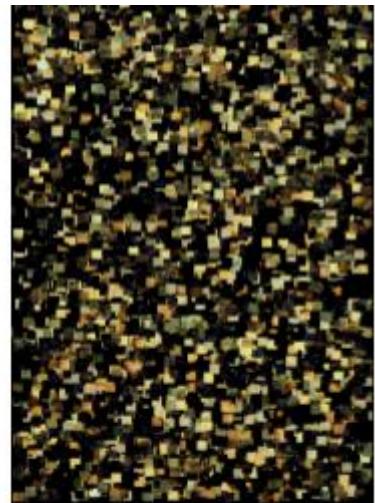


Read length - Impacts

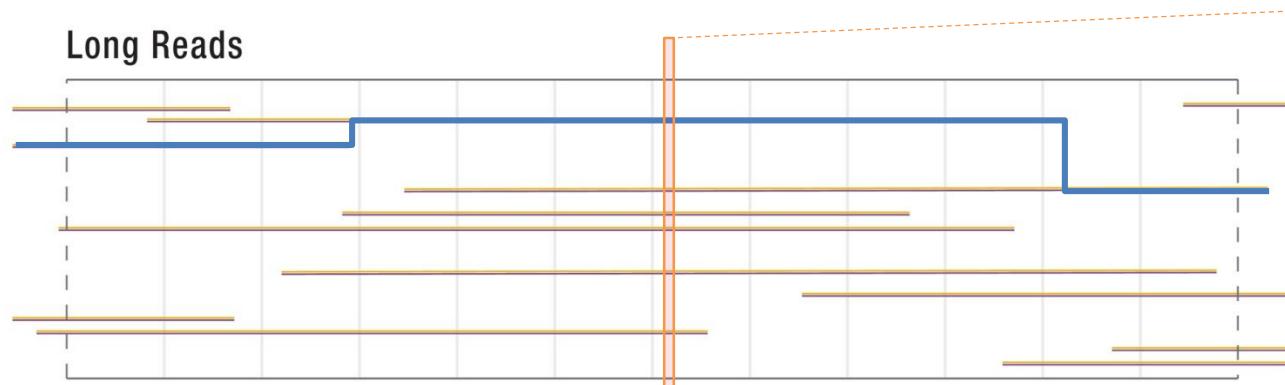
Short Reads



THE JOURNAL OF CLIMATE



Long Reads



- - - - -



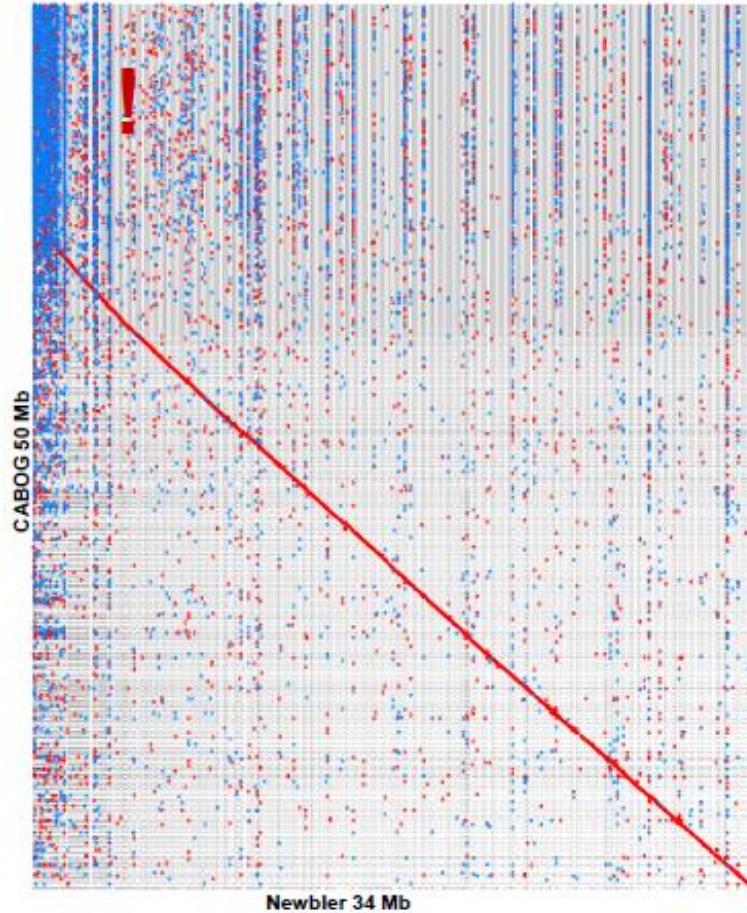
De novo Assembly Process



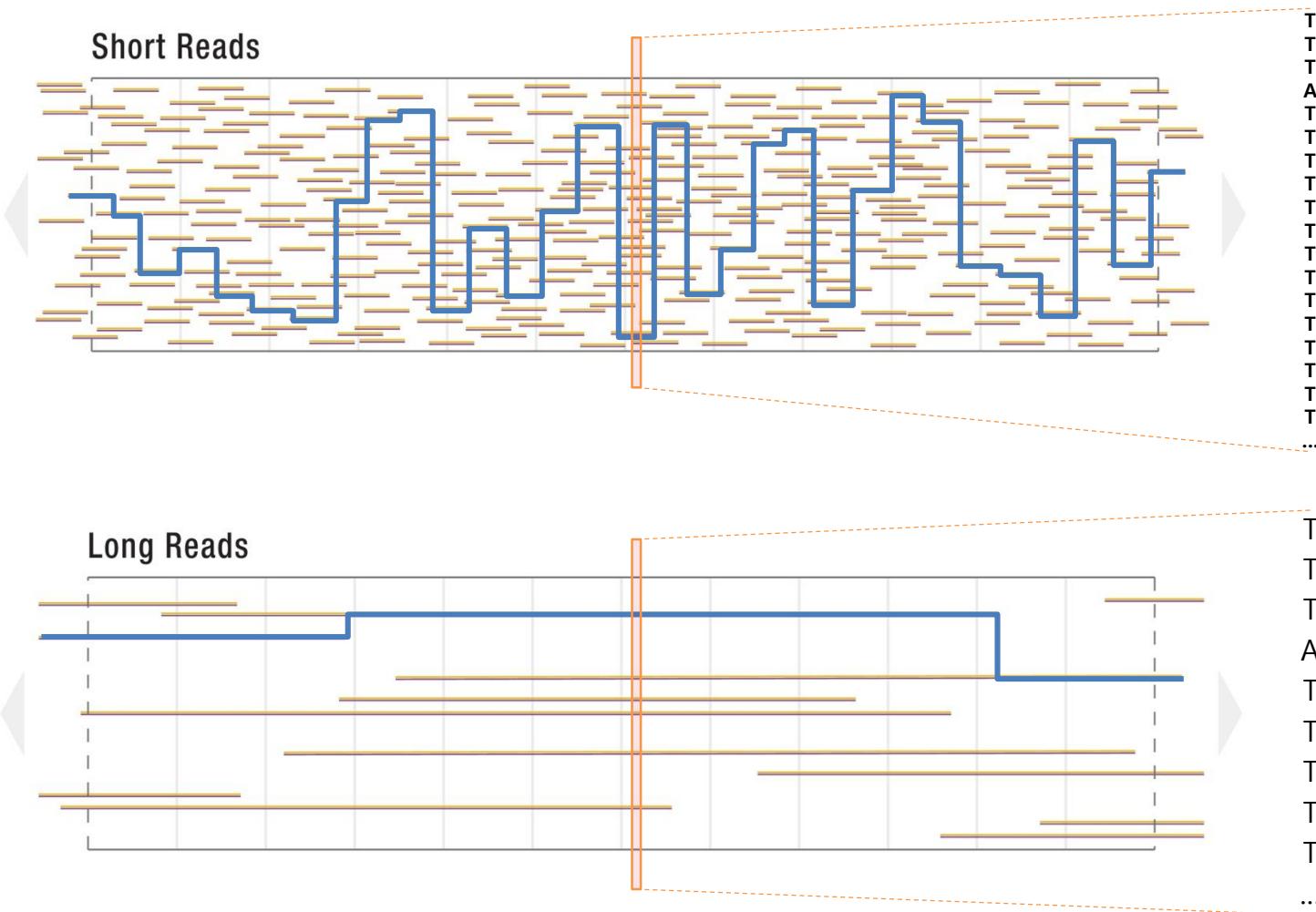
- Overlapping reads assembled to larger continuous pieces : '**contigs**'
 - Assembly '**gaps**' between contigs
 - Contigs ordered to '**scaffolds**'



Mis-assembly / Incompleteness

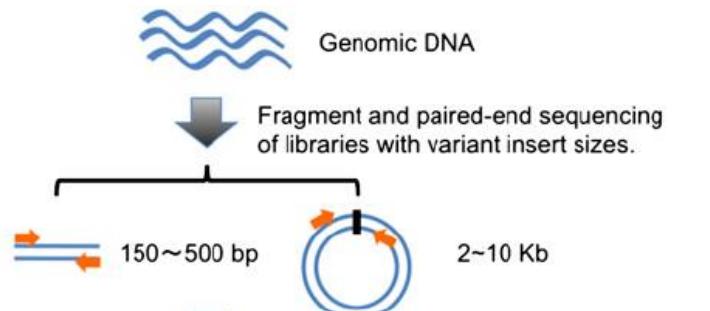


Advantage of Long reads

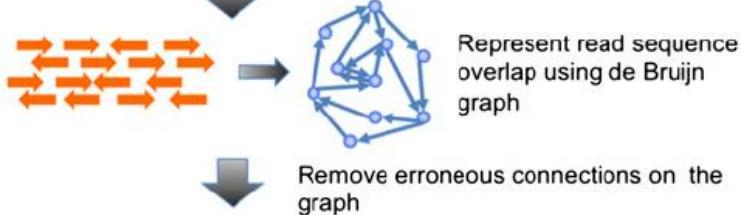


Denovo assembly algorithm from NGS data

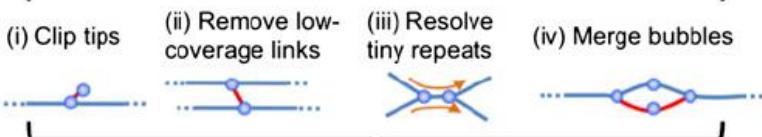
Sequencing (Short, long) A



Overlap (de Bruijn) B



Error correction C



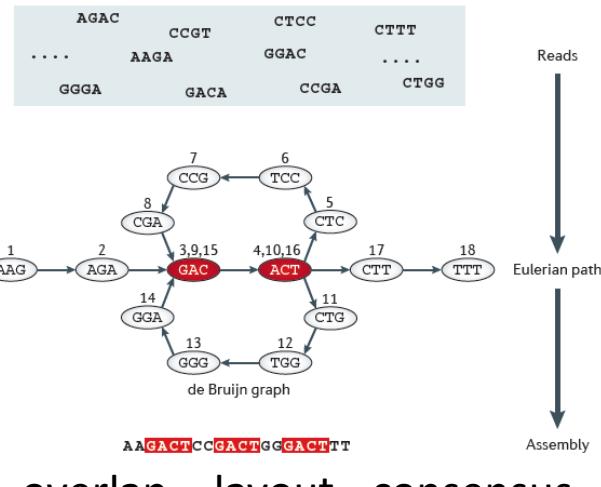
Contigs D



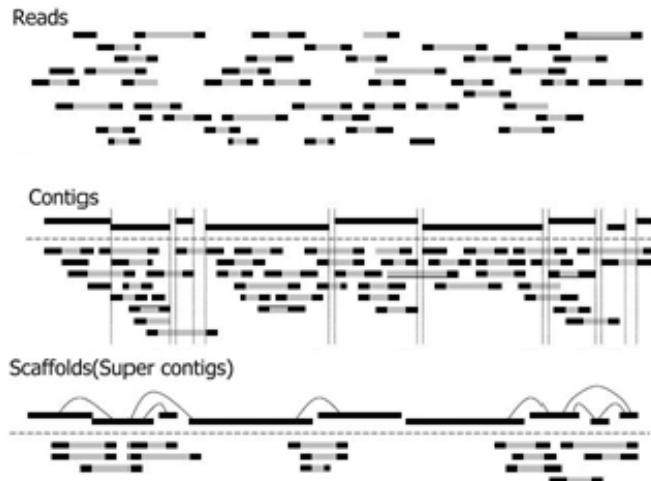
Scaffolds E



Gap closure F



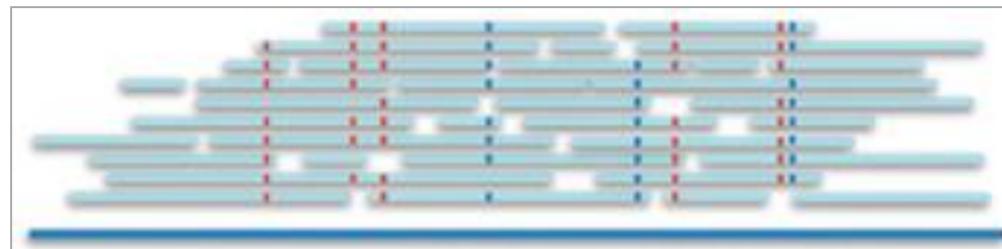
overlap – layout - consensus



(Li et al. 2010; Berger 2013)

Whole-genome *de novo* sequence assembly (PacBio)

SMRT long reads



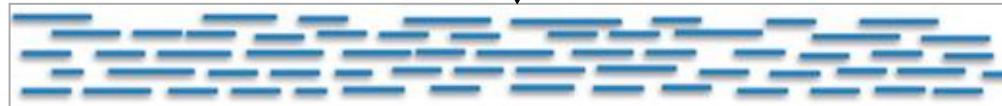
Longest seed reads

Daligner

Construction of longest seed reads



Preassembled reads

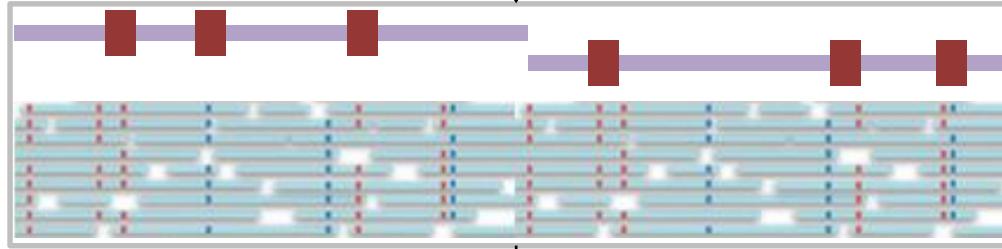


Falcon

Contig assembly



Assembled contig



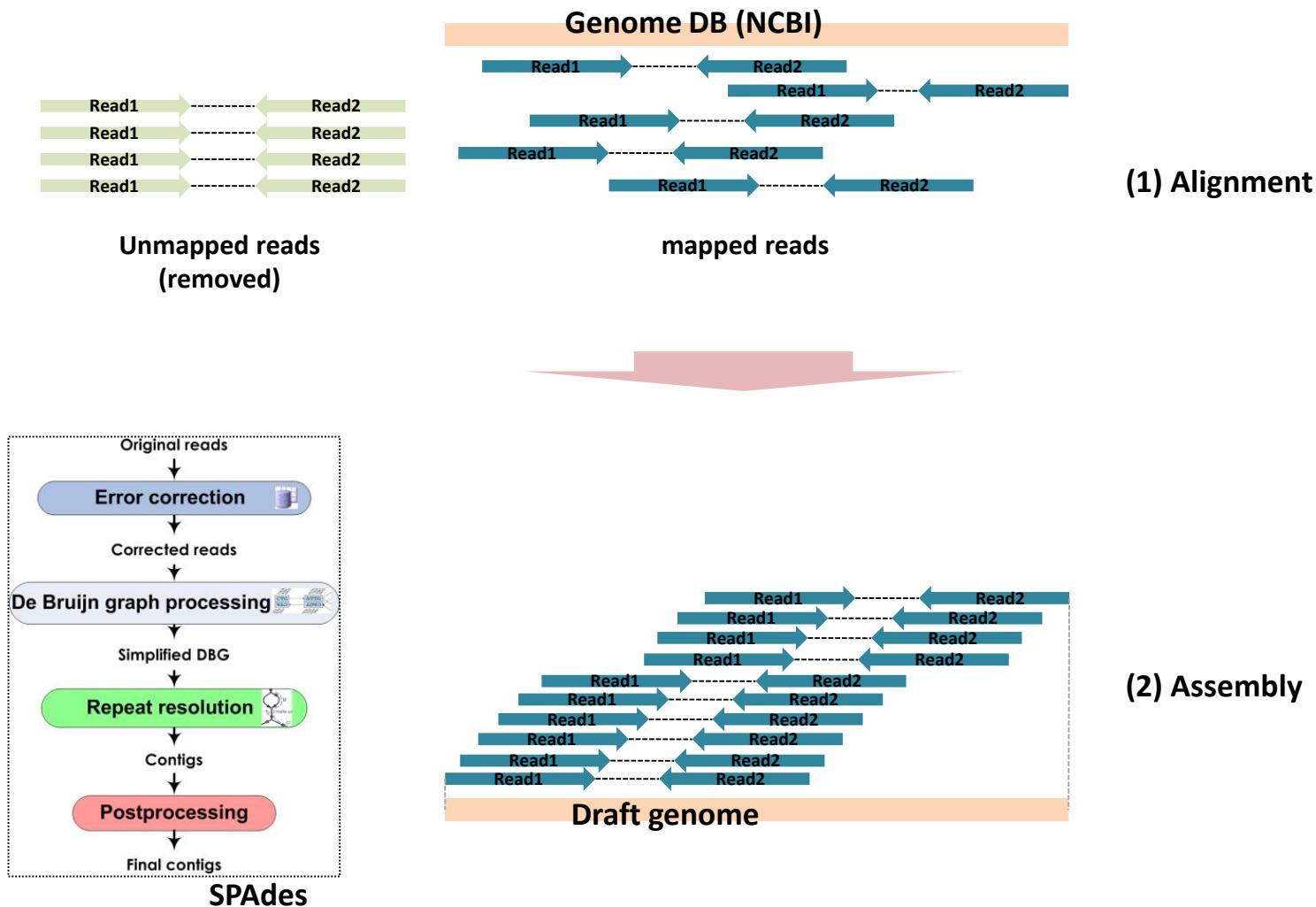
Error correction



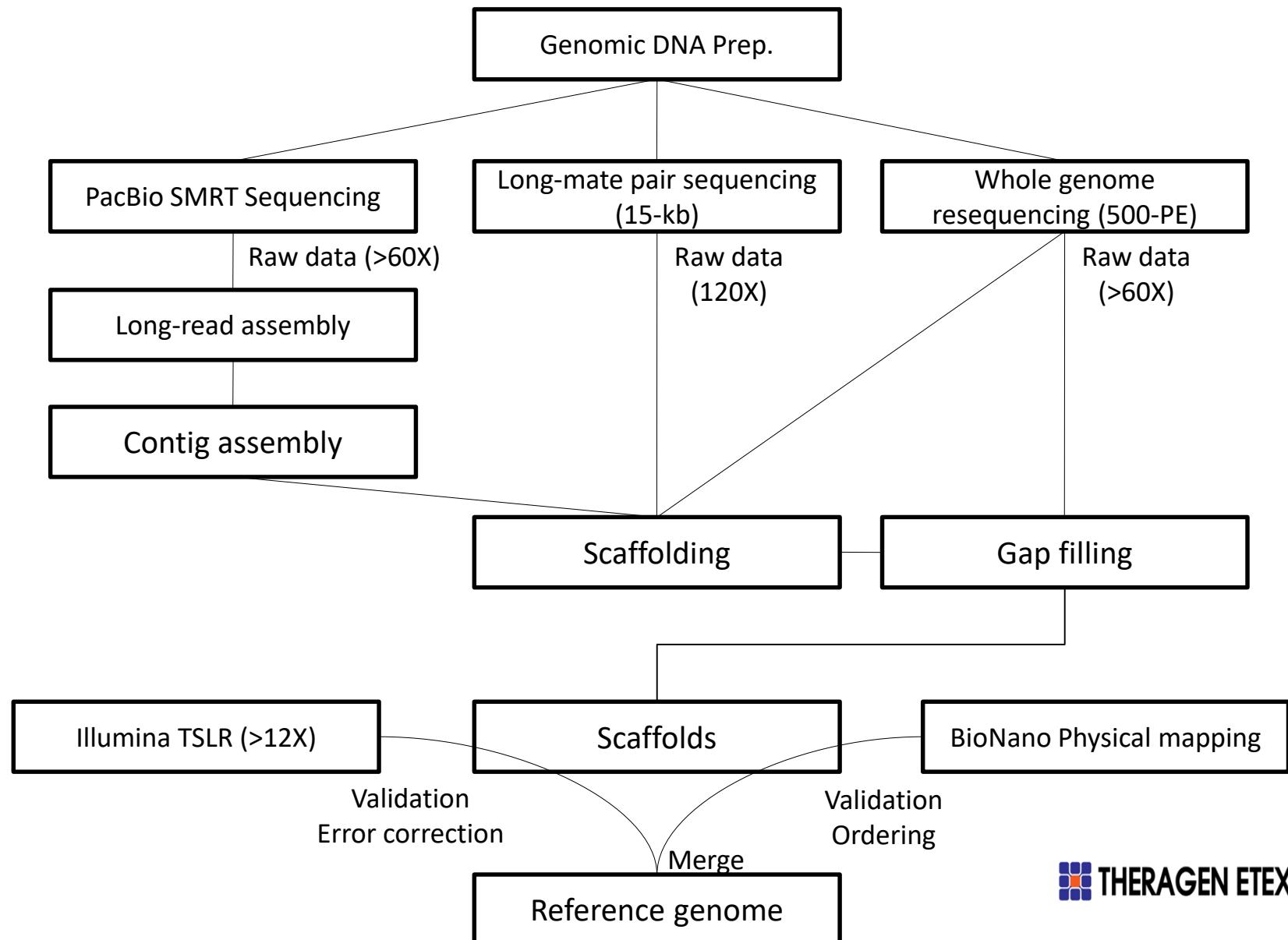
Consensus
corrected contig



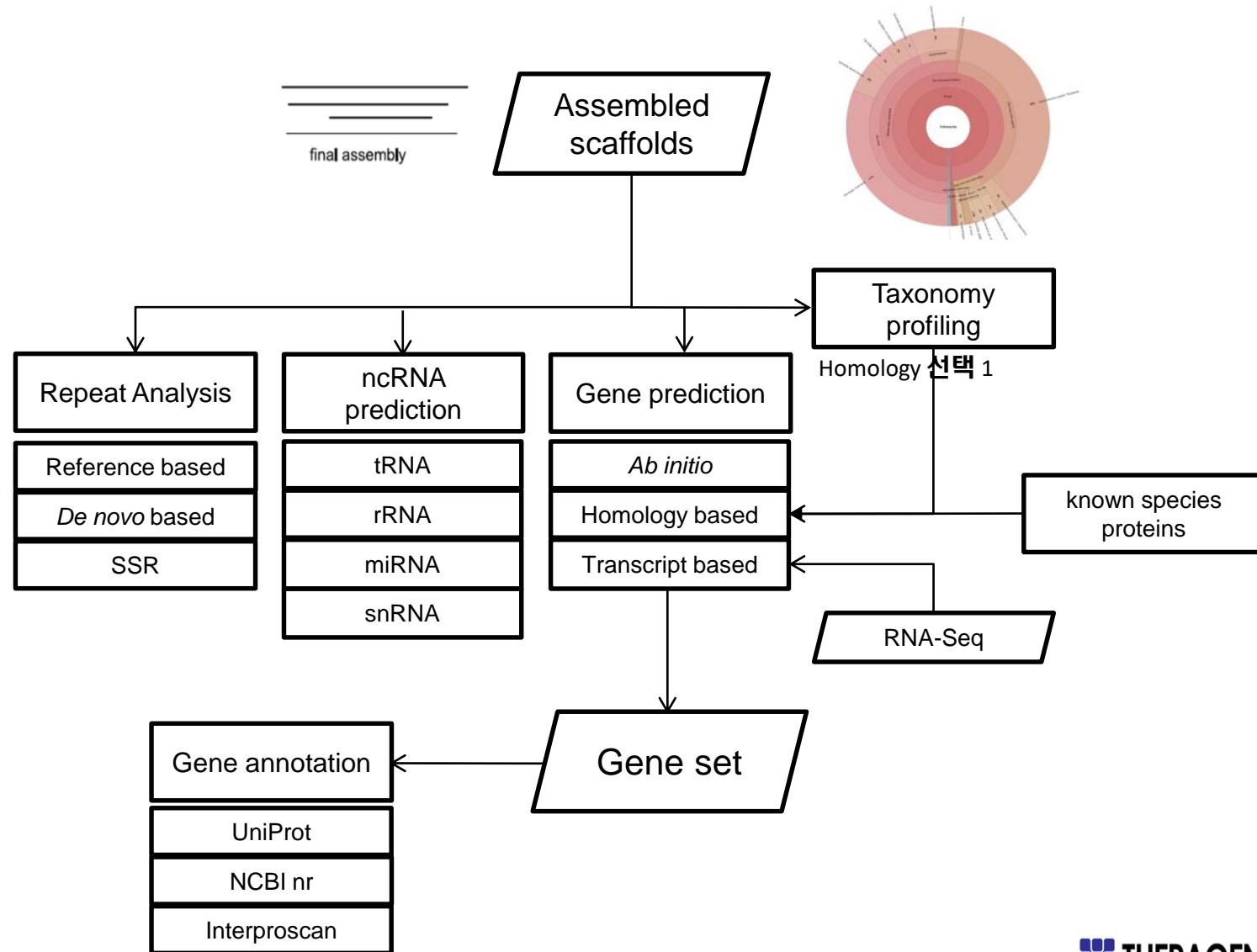
Construction of Genome Guided Assembly



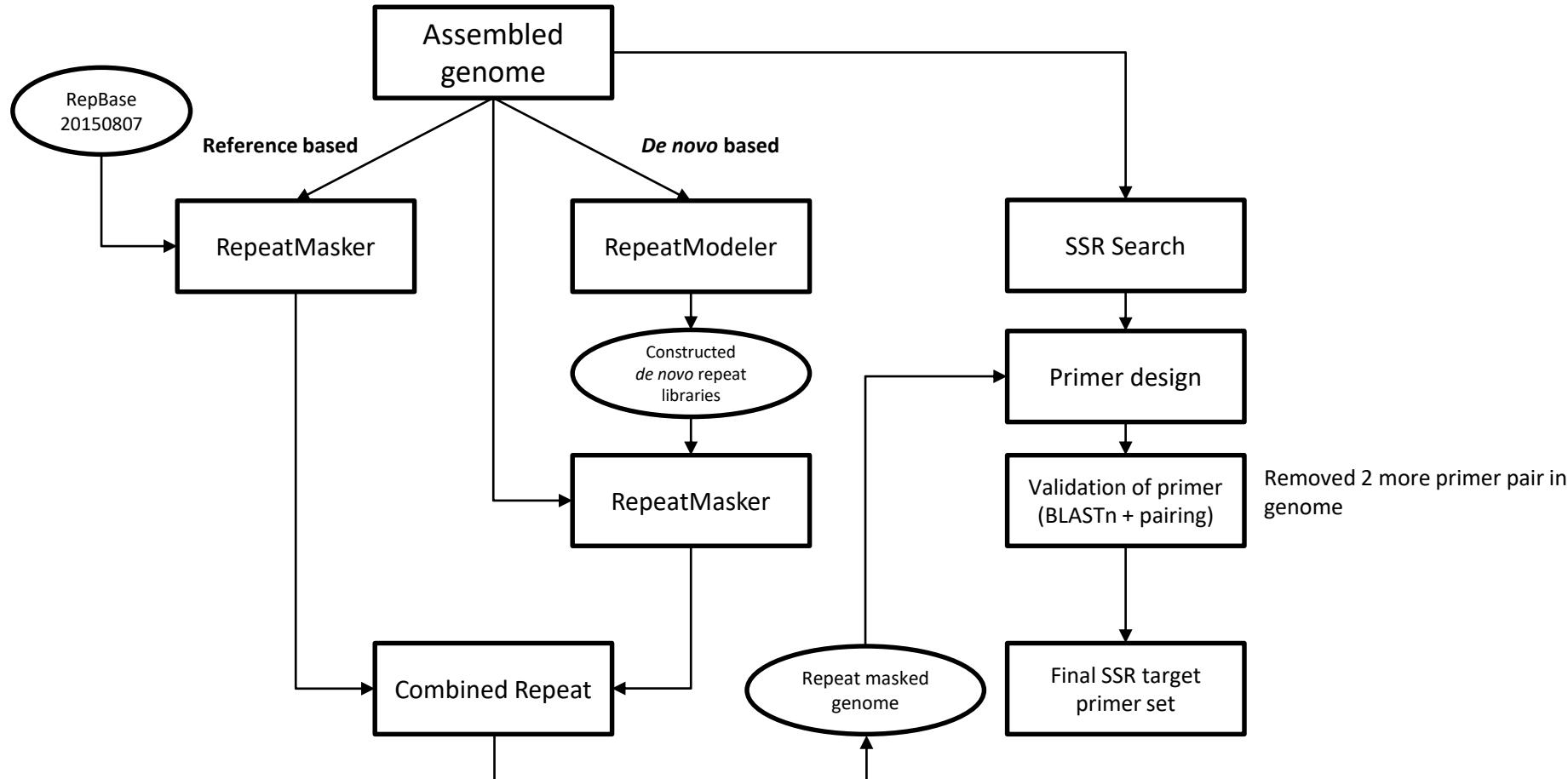
De novo Assembly



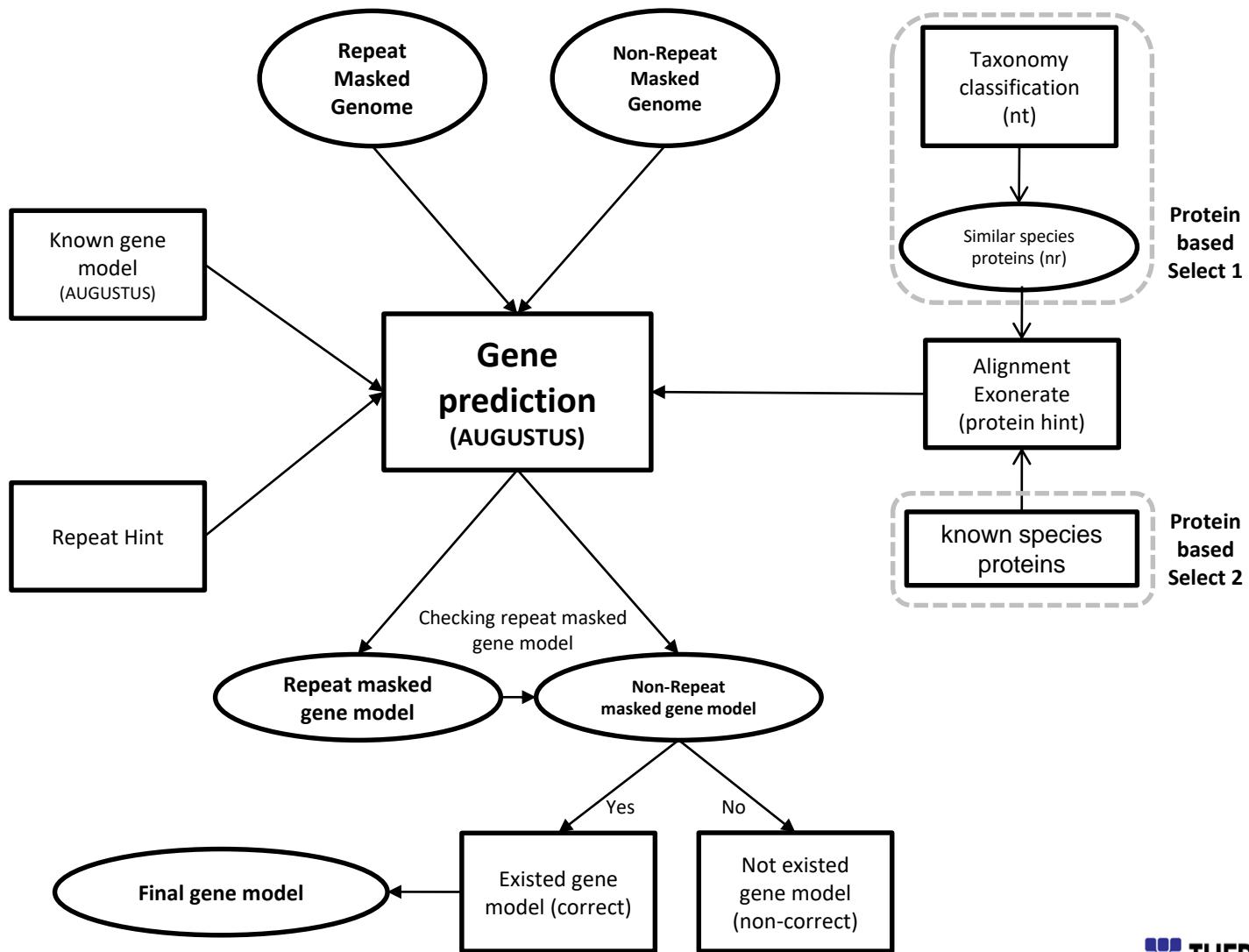
Overview of Genome Annotation Method



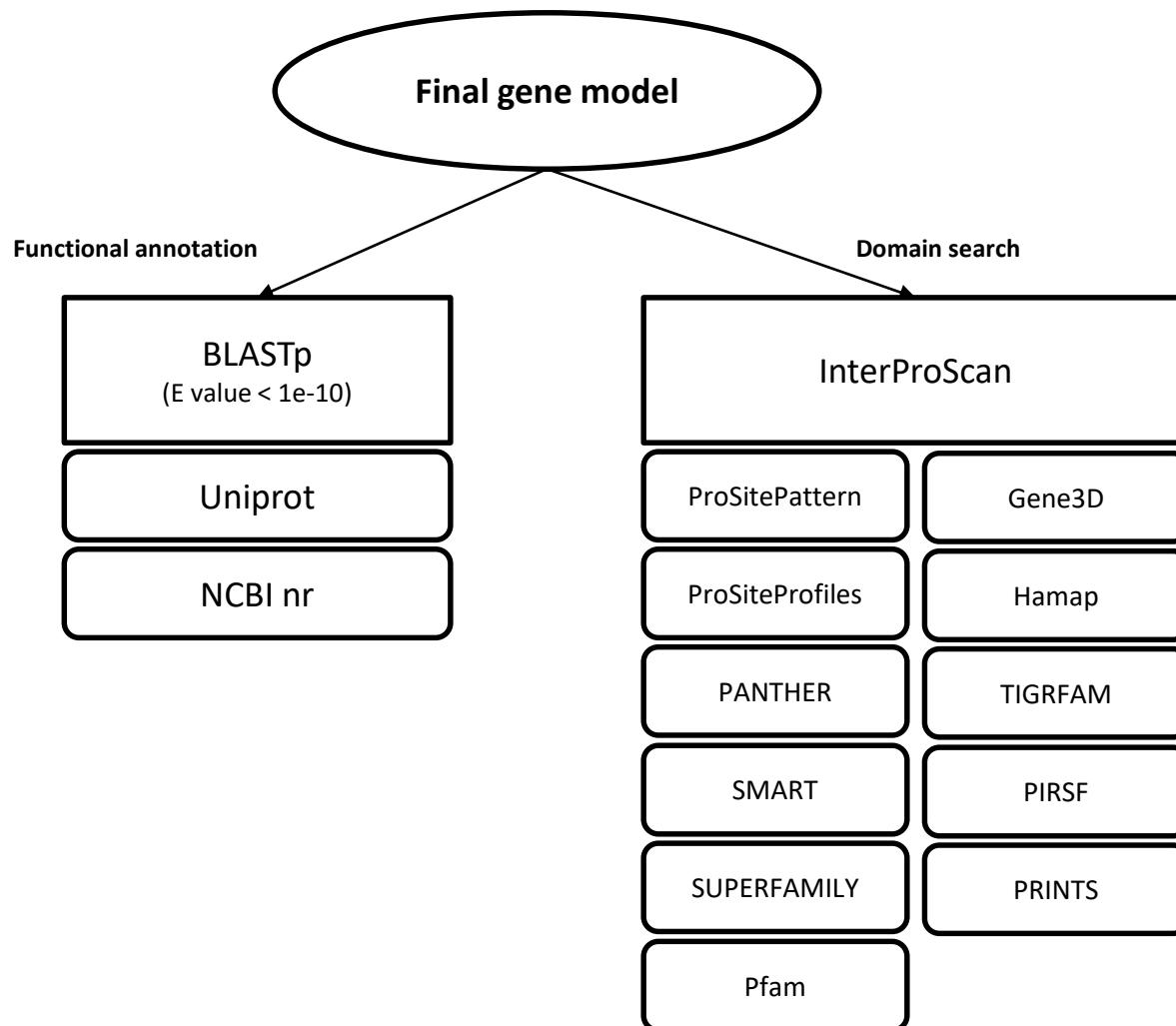
Repeat Analysis



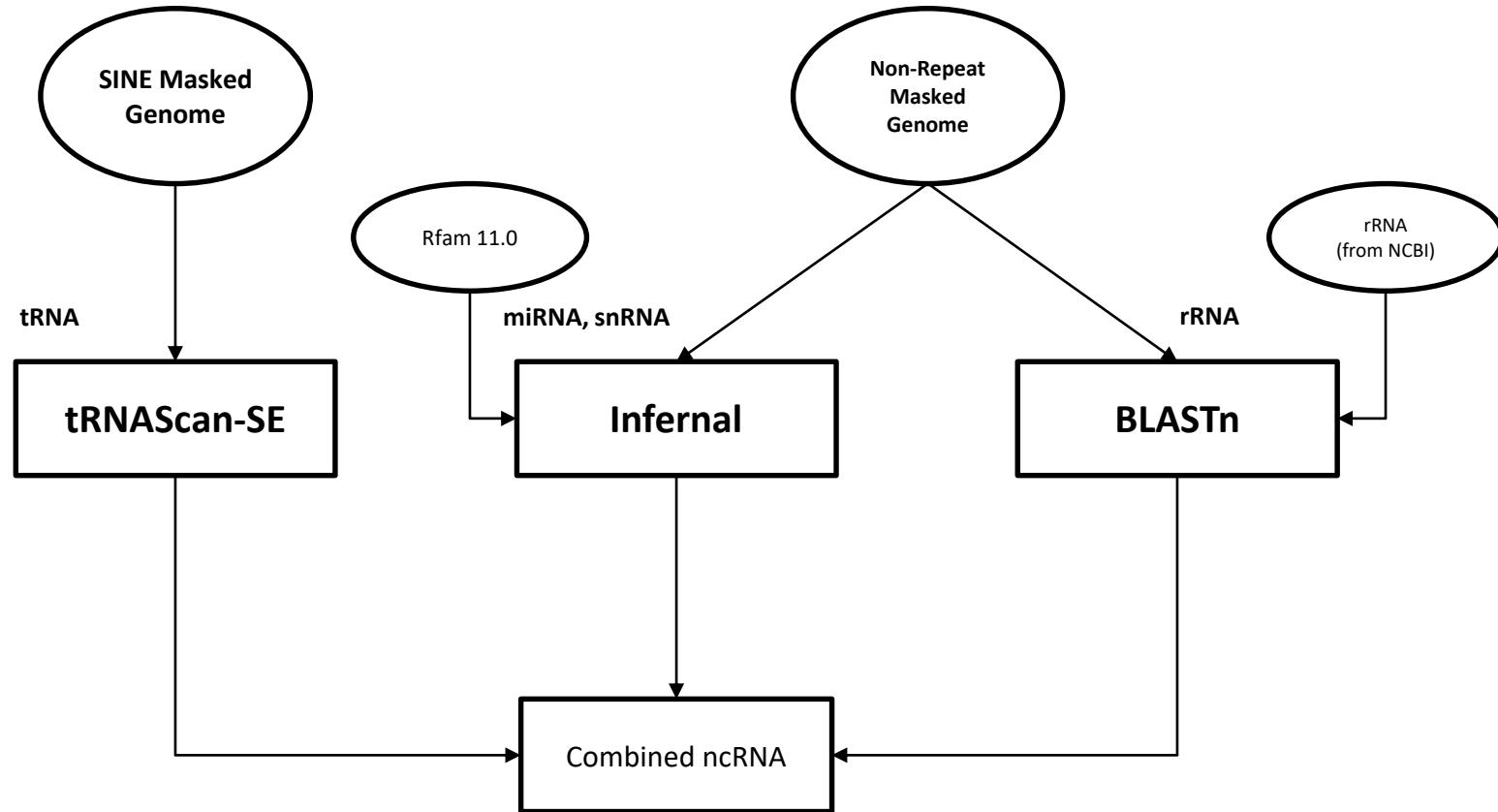
Gene Prediction(Ab initio)



Gene Annotation

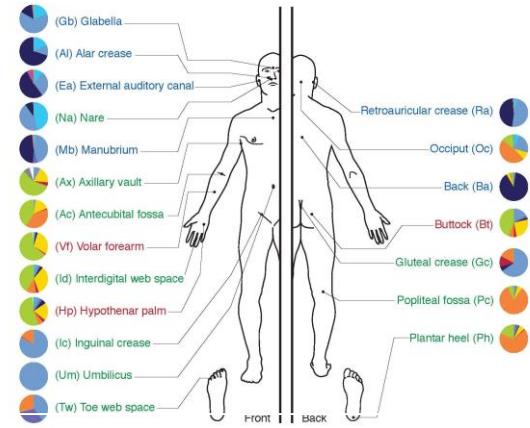
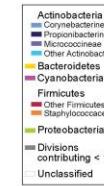
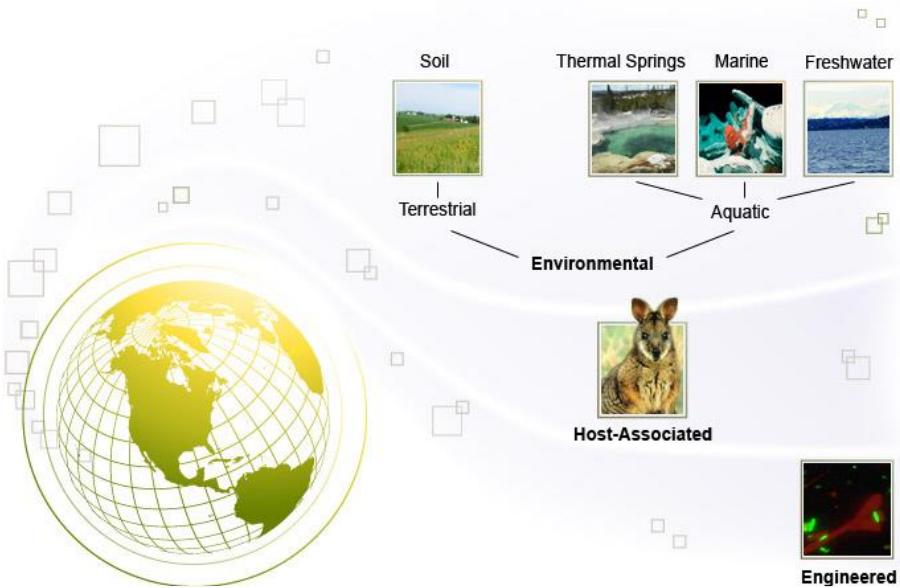


ncRNA Prediction



Metagenomics

- Study of genetic material recovered directly from environmental samples
- The broad field may also be referred to as environmental genomics, eco-genomics or community genomics

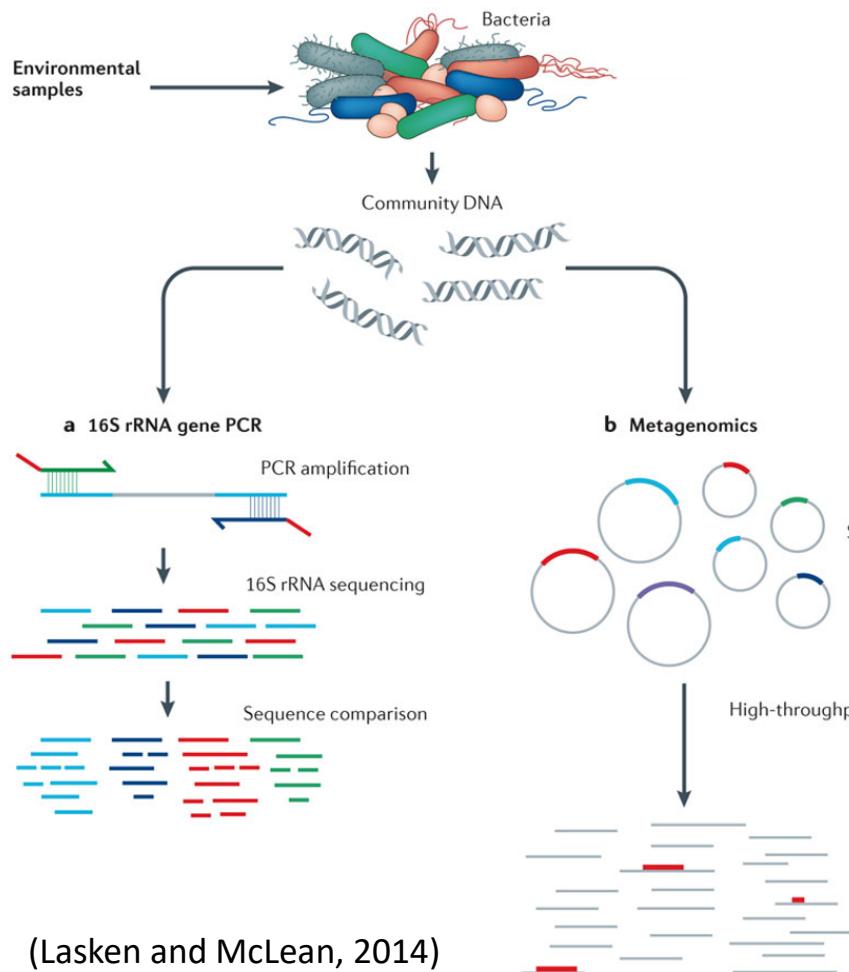


- Objectives

1. Microbial Community Profiling genome from environmental samples
2. To check microbial community differences between environmental samples
3. To check the functional significance of microbial communities

(Burton et al, 2014)

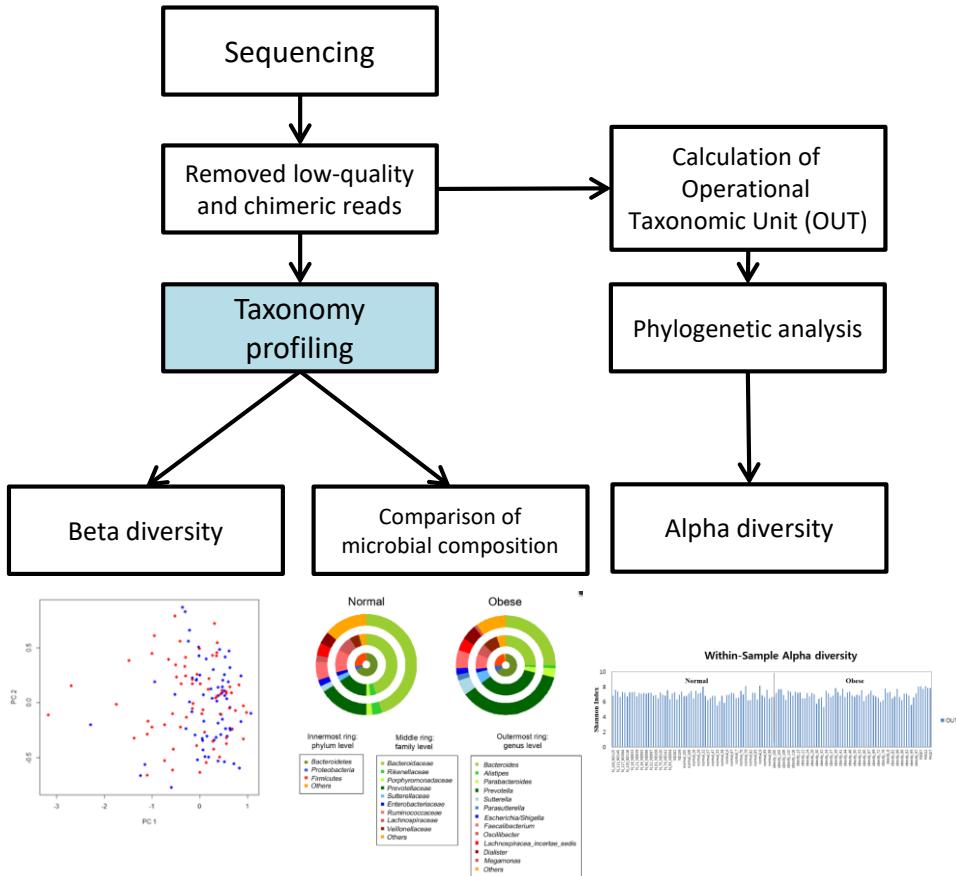
Kind of Metagenomics data by Sequencing



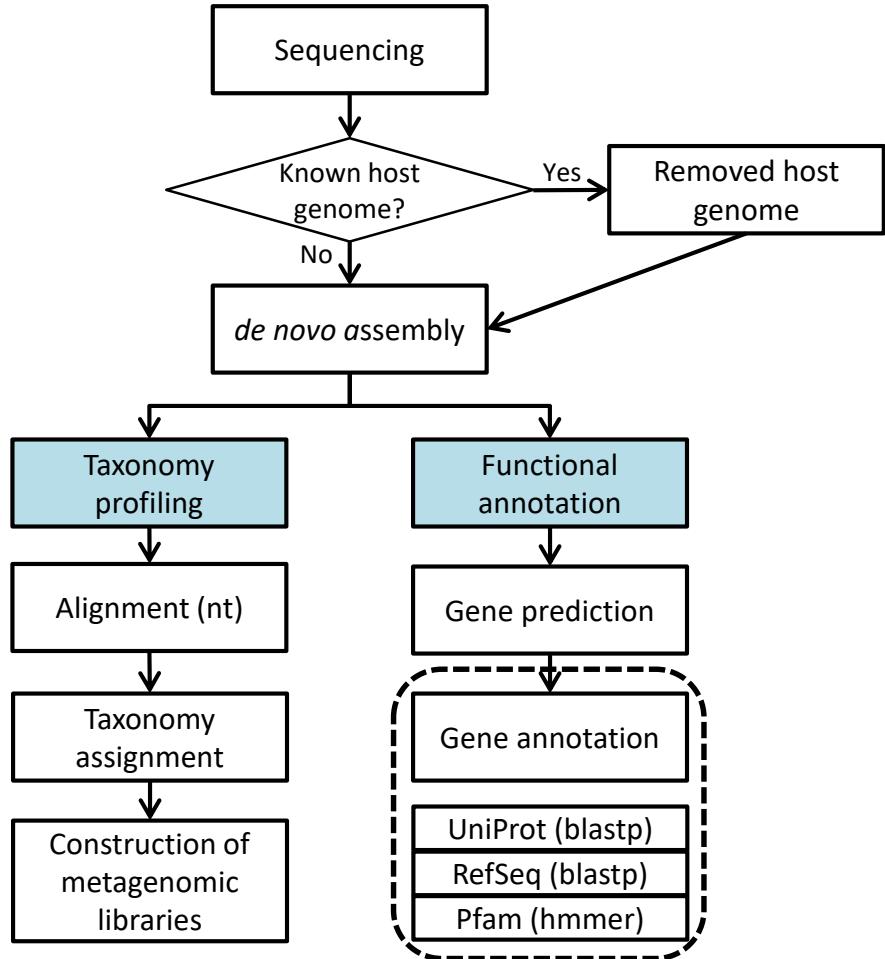
	16 rDNA gene PCR	Whole metagnome
Chimeric DNA	O	X
Cost	low	middle
Covered community	known species	all
Taxonomy profiling	O	O
Functional annotation	X	O
Platform	454, Illumina MiSeq	Illumina HiSeq

Strategy for metagenomics data analysis

(1) 16 rDNA gene PCR

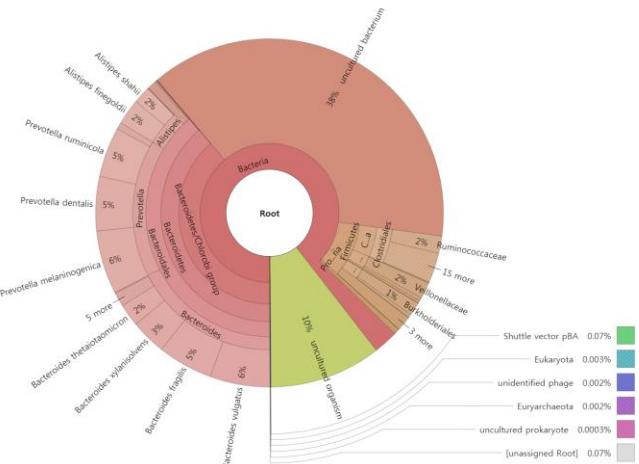
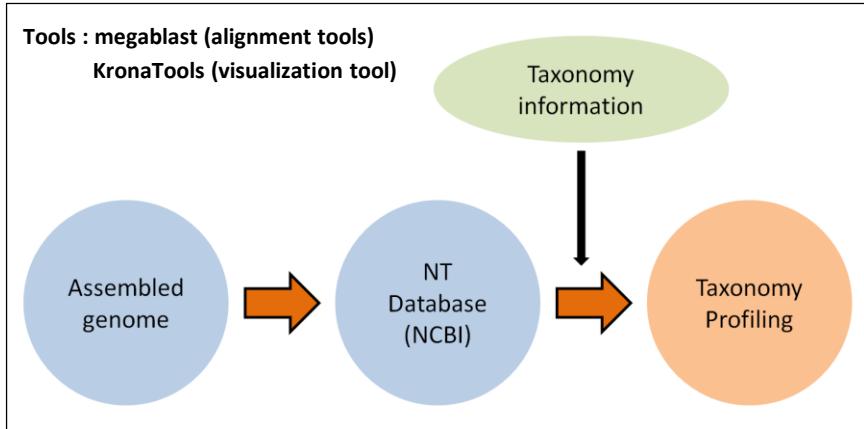


(2) Whole metagenome



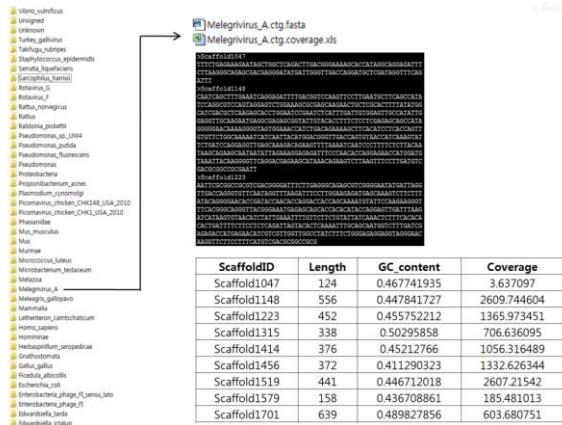
Taxonomy Profiling

- Taxonomy Profiling

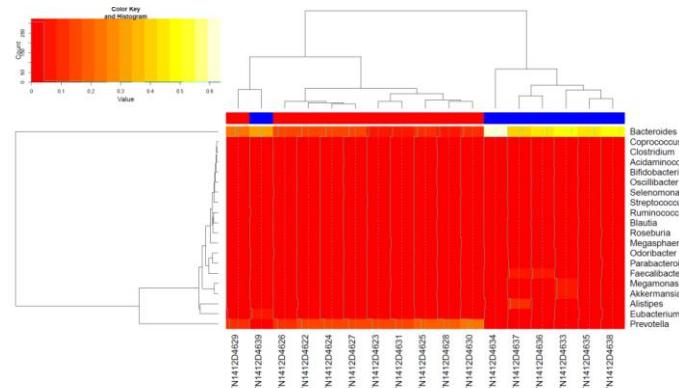


Taxonomy profiling (Krona chart – html format)

- Construction of metagenome libraries



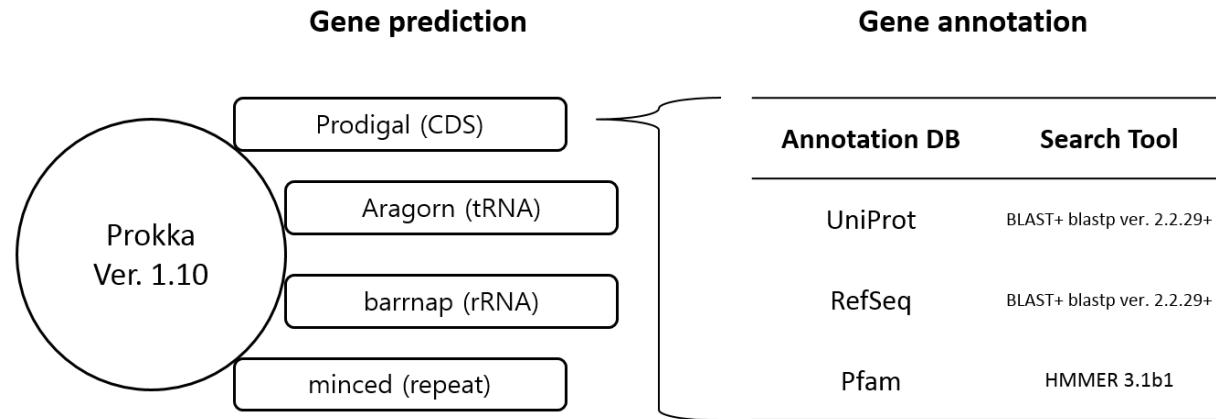
ScaffoldID	Length	GC_content	Coverage
Scaffold1047	124	0.467741935	3.637097
Scaffold1148	556	0.447841727	2609.744604
Scaffold1223	452	0.455752212	1365.973451
Scaffold1315	338	0.50295858	706.636095
Scaffold1414	376	0.45212766	1056.316489
Scaffold1456	372	0.411290323	1332.26344
Scaffold1519	441	0.446712018	2607.21542
Scaffold1579	158	0.436708861	185.481013
Scaffold1701	639	0.489827856	603.680751



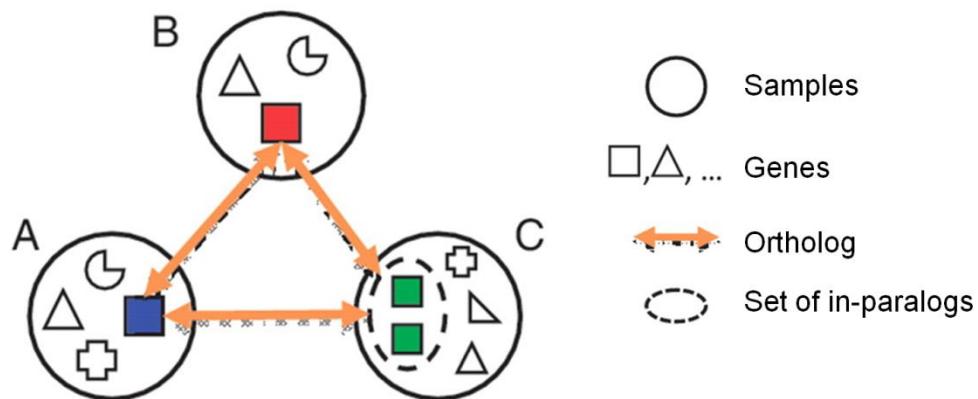
Comparison of microbial composition

Functional Annotation

- Gene prediction and annotation

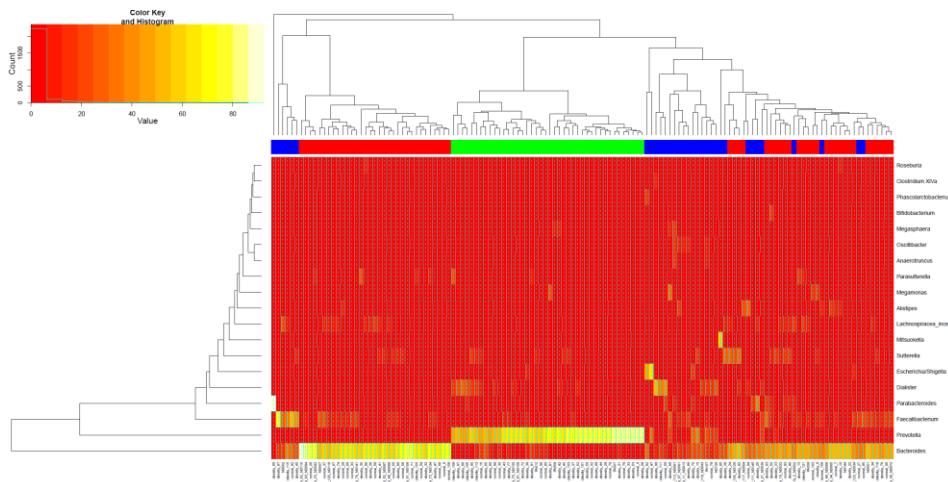


- Gene cluster



Metagenome Profiling(1)

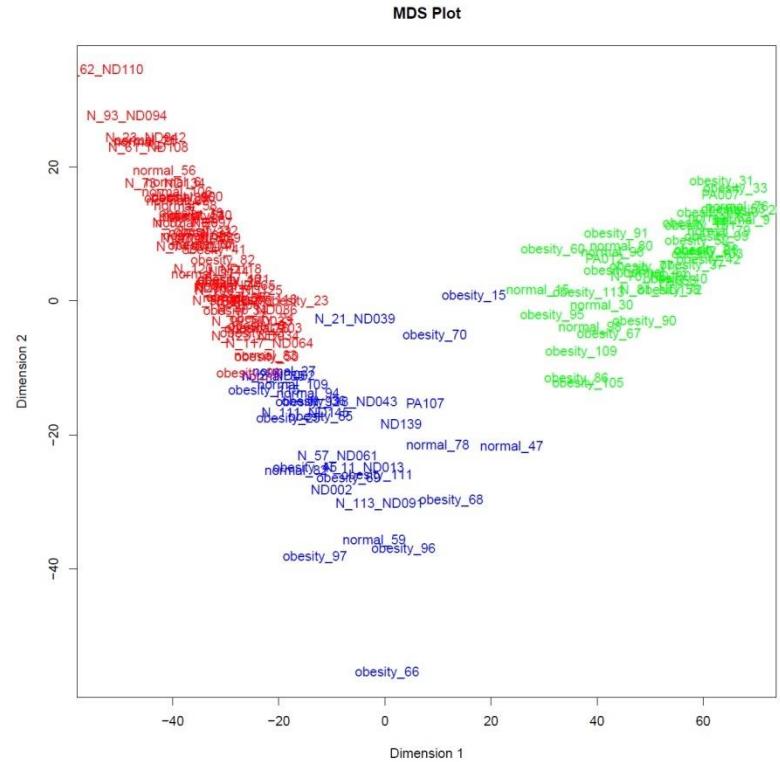
Pattern Analysis



Kmeans Clustering and Heatmap

	Bacteroides Group	Prevotella Group	Other Group
특정형질	23	30	15
normal	38	12	17
Total	61	42	32
Odds Ratio	0.61	2.50	0.88

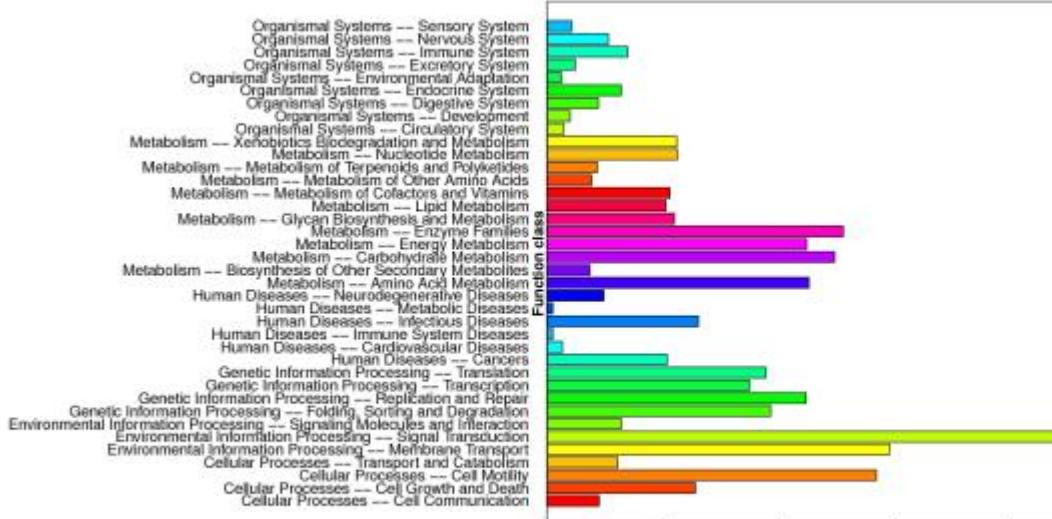
Kmeans Clustering and OddRatio



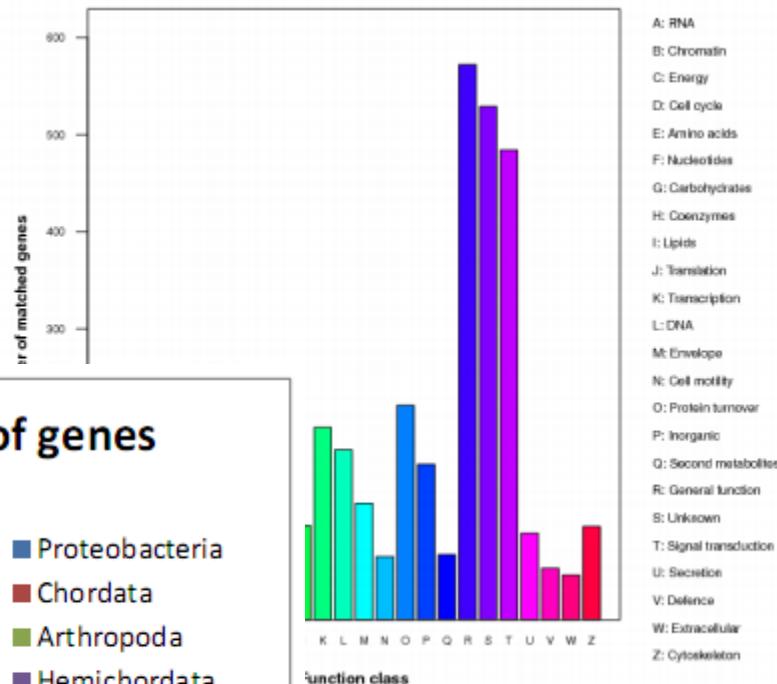
Kmeans Clustering and MDS Plot

Metagenome Profiling(2)

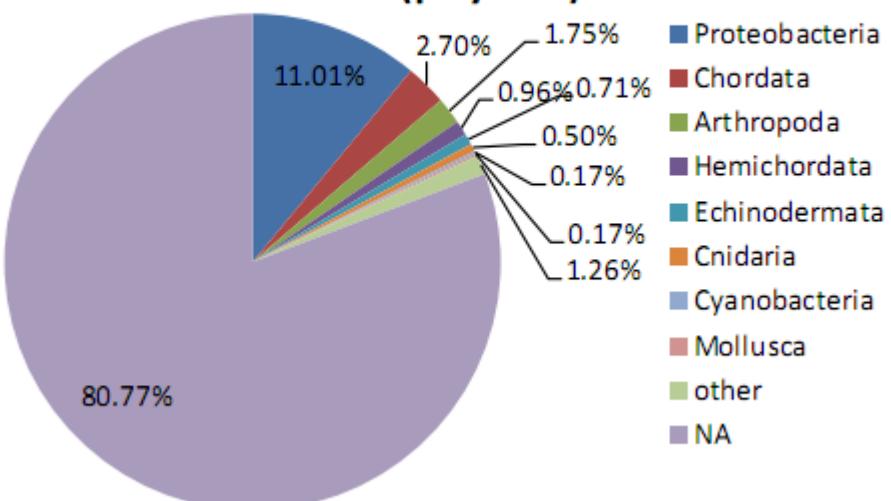
KEGG pathway classification



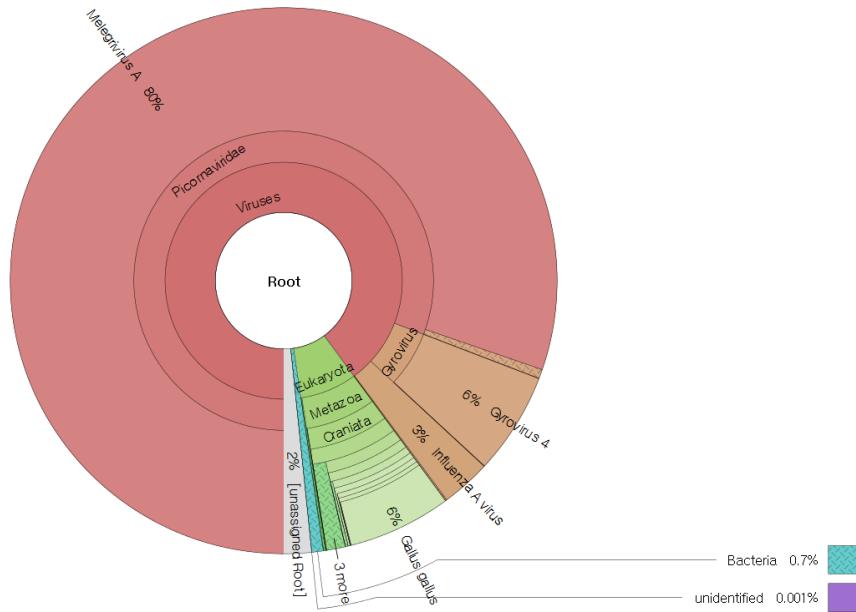
eggNOG function classification



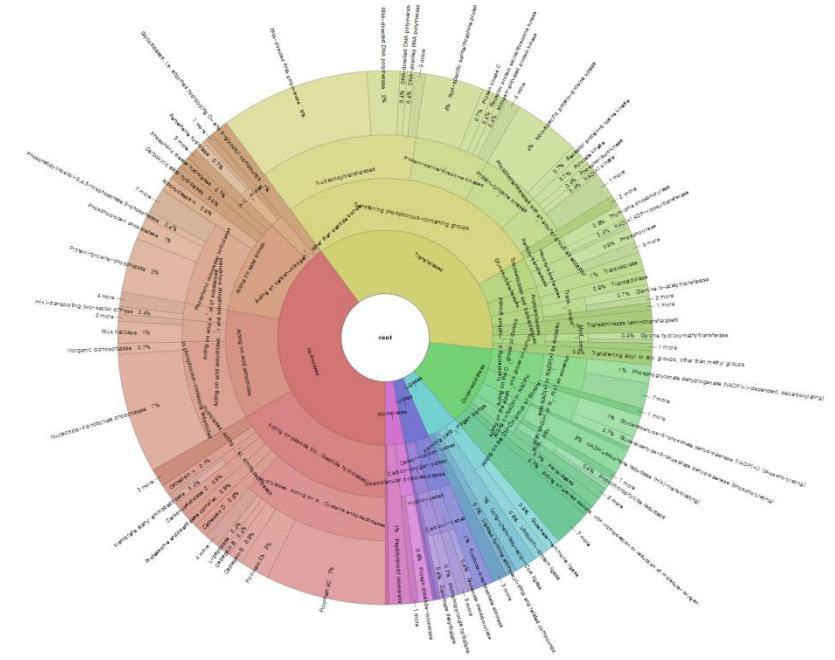
Percentage taxonomic affiliation of genes identified(phylum)



Metagenome Profiling(3)



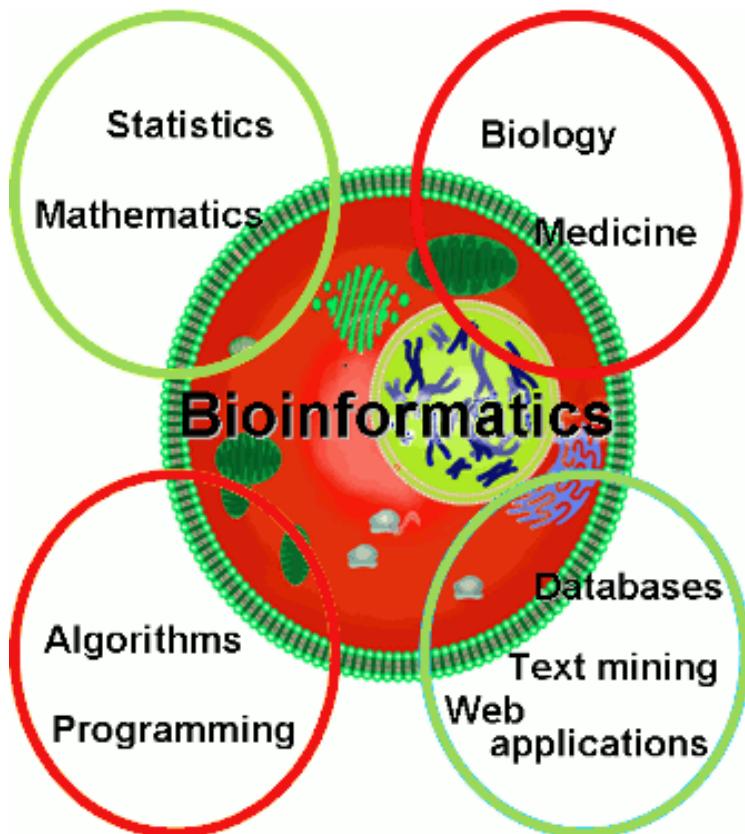
Taxonomy
Classification



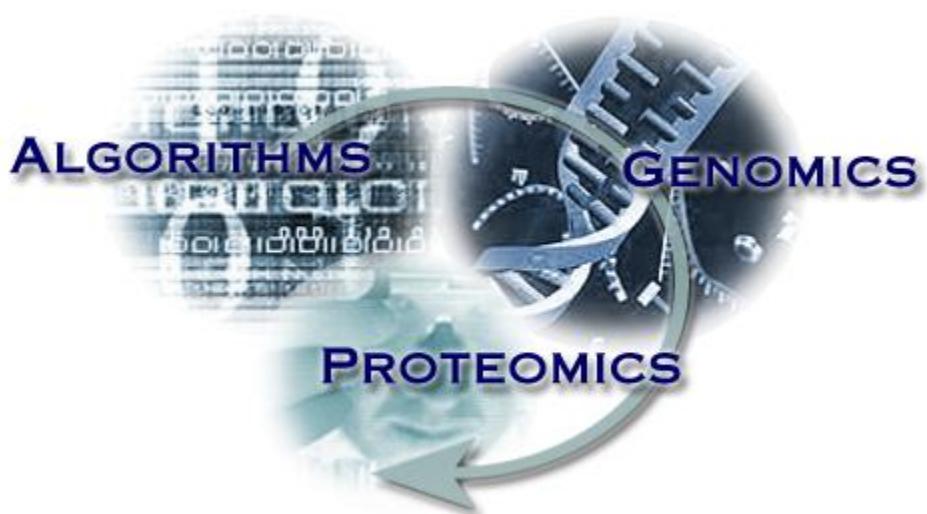
Functional Annotation

Song B, Su X, Xu J, Ning K. **MetaSee: an interactive and extendable visualization toolbox for metagenomic sample analysis and comparison.**
PLoS One. 2012;7(11):e48998. doi: 10.1371/journal.pone.0048998.

Bioinformatics is...



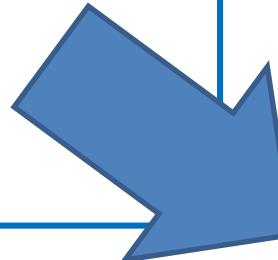
생물학적인 문제를 수학, 정보과학, 통계학, 전산학등의 방법론을 이용하여 정보전달의 차원에서 해석하는 학문 및 연구방법론



Bioinformatics Qualification

Qualification

1. Biology background
 - Molecular Biology, Microbiology, Cell Biology etc.
2. Linux
3. Programming language
 - C-language, JAVA, Script language(perl, python)
4. Web master
 - ASP, JSP, PHP
5. Mathematics
6. Statistics
7. Database
 - MySQL, Oracle



생물학적 이해 및 데이터 마이닝

유전체 분석 결과



HELLO GENE 헬로-진

질병민감성 요약결과

내 질병 위험도 분포

13.3%	13.3%	40%	20%	13.3%
-------	-------	-----	-----	-------

결과보기 To

0.02%	0.00%	0.00%	0.00% (임신)	임명 예측비	높음	매우높음	보통	매우낮음	낮음
-------	-------	-------	------------	--------	----	------	----	------	----

암

	위험도	증상								
유방암	1.05	28쪽	0.91	30쪽	1.05	32쪽	1.03	34쪽	1.13	36쪽
장형 위암	1.12	38쪽	1.31	40쪽	0.85	42쪽	0.62	44쪽	1.09	46쪽
뇌관암	0.77	48쪽	1.04	50쪽	0.87	52쪽	1.22	54쪽	0.96	56쪽

근골격계 질환

	위험도	증상	위험도	증상	위험도	증상		
골관절염	1.27	58쪽	0.88	60쪽	0.76	62쪽	0.82	64쪽
무릎 골관절염	5.501%	4.019%	3.537%	4.019%	1.985%	2.812%	0.645%	0.793%

* 높은 위험도 (Orange Red) : 고지된 위험도 고지되었거나 다른 오류로는 자료들을 나타낸 것입니다.
* 중간 위험도 (Ochre Red) : 각 위험도는 개인의 유전적 전이도를 반영하고 나타낸 것입니다.

※ 유전체는 추가 연구에 의해 변경될 수 있습니다.

Home

Your Results
Red, orange, gray and green indicate the relevance of your results for the tested genetic conditions.

Categories
Browse the tested conditions to see if they impact your carrier status, response to medication, disease risk or lifestyle.

Body
Discover how variations in your DNA influence different body systems.

Chromosomes
Explore your DNA on a chromosome map.

Explore

Track Your Sample
We've shipped your Gentle kit. It should arrive within 2 business days.

MyGenome

Viewing Jay Flatley's Genome Please Download Reference Sequence

Health

Genome Map

Reports

Journey

About

Carrier WiFi 3:38 PM 13321 hCRC Mediterranean Male 14

15 CRC Gene Panel > MLH3 2 RS3A-L3

Sort by: Gene

MLH3

NM_001040108 (MLH3): c.2474A>C (p.N825I) Cn14.753.738.75.533.683

Lady Sanger (14.753.738.75.533.683)

NM_001040108.1(MLH3): c.2531C>A (p.P84L) Cn14.753.738.75.533.683

Small Variations

Genes

유전체 검사 사례

Christina Applegate



- 한쪽 유방에 암이 있었고
- 유전자 검사에서 **BRAC1** 유전자의 돌연변이 발견
- 가족력 (모친이 유방암과 자궁암)

- Preventive Double Mastectomy
- Breast Reconstructive Surgery

유전자 예측 서비스를 통해
향후 일어날지 모르는 위험을
미리 예방

ANGELINA JOLIE



- 유전자 검사에서 **BRAC1** 유전자의 돌연변이 발견
- 가족력 (어머니 난소암으로 사망, 이모 유방암으로 사망)

SLC24A5 (피부 색깔과 관련 있는 유전자)



probably darker-skinned, Asian or African ancestry

rs1426654

나의 코드: GG 98.2% (한국인)



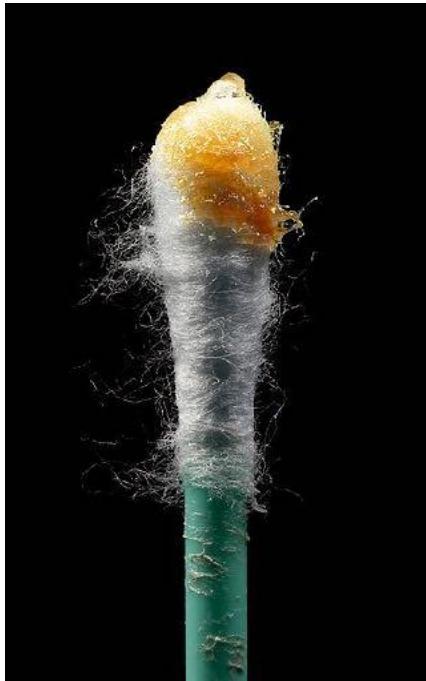
ABCC11 (귀지 타입 관련 유전자)



rs17822931

Dry earwax. No body odour

나의 코드: TT 78.8% (한국인)



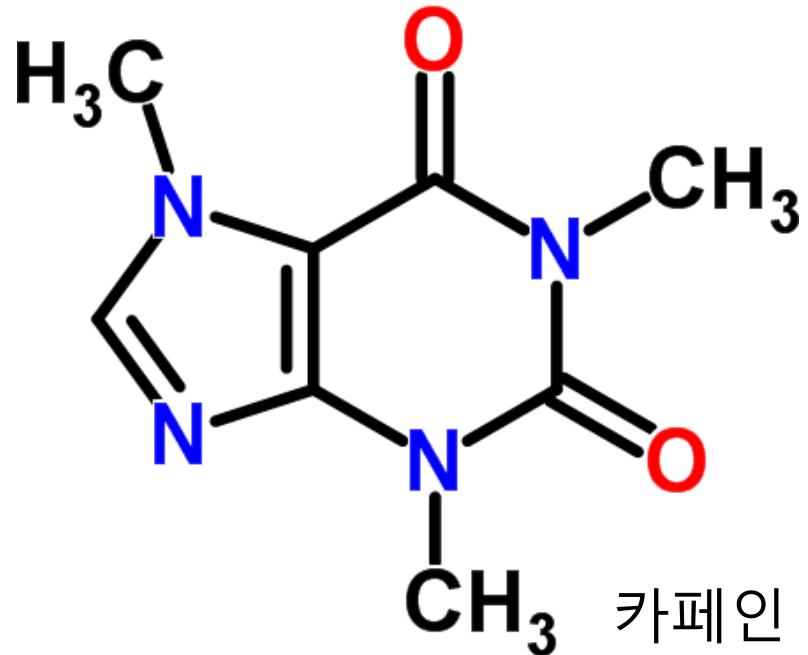
CYP1A2 (카페인 분해 능력 관련 유전자)



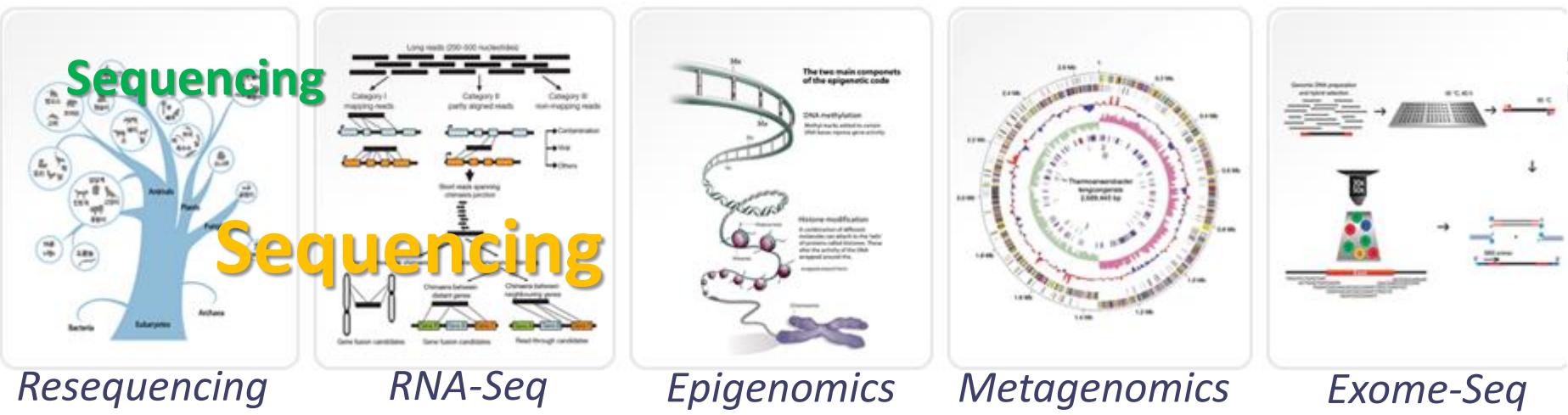
rs762551

Slow Caffeine Metabolizer.

나의 코드: CC 14% (한국인)



Thank you for your attention !



Sequencing

Sequencing

Theragen
bio institute

TotalOomics
solutions

Genome !

Theragen