

Illumina TruPath™ Genome

가장 간편한 전장 유전체
시퀀싱 워크플로우



단 10분의 수동 assay 준비 단계를
통해 구현이 가능한 샘플 준비부터
데이터 생성까지 아우르는 가장
간편한 워크플로우

TAGCA
CGTAG
ATAGC

NovaSeq™ X 시리즈에서 생성된
쇼트 리드 데이터와 근접성 데이터를
통해 매핑이 어려웠던 유전체 영역
분석



향상된 구조적 변이 검출력과
초원거리 페이징으로 원거리 정보
확보

패러다임을 바꾸는 유전체 접근성

그동안 연구자들은 유전체의 많은 부분을 분석할 때 주로 쇼트 리드(short-read) 차세대 시퀀싱(next-generation sequencing, NGS) 방법을 사용해 왔습니다. 그럼에도 극히 반복적이거나, 시퀀스 상동성(sequence homology)이 높거나, 구조적으로 복잡한 일부 영역과 변이 타입은 여전히 그 특징을 파악하는 데에 어려움이 있습니다. 최근 몇몇 연구를 통해 이러한 영역에 인간의 유전 질환에 영향을 줄 수 있는 변이가 있는 것으로 밝혀짐에 따라,^{1,2} 유전 질환의 근본적인 원인을 찾기 위해 이러한 변이 분석의 중요성이 대두되고 있습니다.

롱 리드 시퀀싱(Long-read sequencing)과 같은 대안은 향상된 해상도를 제공하여 매핑이 어려운 영역 및 변이에 대해 보다 깊은 통찰력을 제시합니다. 다만, 롱 리드 시퀀싱은 워크플로우가 다소 복잡한 경향이 있으며, 잘 알려진 정확도 문제로 인해 일관성 없는 결과를 초래할 수 있습니다.

Proximity mapped read technology(근접성 정보로 리드를 매핑하는 기술)를 활용하는 Illumina의 TruPath Genome은 전장 유전체 시퀀싱(whole-genome sequencing, WGS) 워크플로우의 패러다임을 바꿉니다. 이 혁신적인 assay는 단 10분이 소요되는 간편한 수동 assay 준비 단계를 거쳐 긴 DNA 절편을 포획합니다. 시퀀싱에는 기존 NovaSeq X 시리즈가 사용되며, 최첨단 DRAGEN™ Secondary Analysis로 데이터를 분석합니다. 연구자들은 이 워크플로우가 제공하는 정확하고 완전한 유전체를 이용하여 롱 리드 시퀀싱이 일반적으로 생성 가능한 리드의 길이보다 더 긴 길이의 통찰력을 얻을 수 있습니다.

이제 연구자들은 TruPath Genome을 사용하여 쇼트 리드 시퀀싱의 용이성과 정확성을 기반으로 매핑이 어려운 유전체 영역을 분석하고, 구조적 변이(structural variant, SV) 검출력을 높이며, 페이징된 리드(phased read)와 변이 콜(variant call)을 생성할 수 있습니다.

혁신적인 워크플로우 간편성

TruPath Genome은 WGS 절차를 크게 간소화하고 수작업 시간을 단 10분으로 줄여 줍니다(그림 1). Illumina의 신뢰할 수 있는 트랜스포제이즈(transposase) chemistry로 기존의 번거로운 라이브러리 준비 단계를 제거했습니다. 또 진보된 바이오인포매틱스(bioinformatics, 생명정보학) 알고리즘은 일반적인 페어드 엔드 쇼트 리드 데이터*와 나노웰(nanowell)의 근접성 정보를 함께 활용하여 최대 수백만 개의 염기(base) 거리만큼 떨어져 있는 여러 변이 사이의 관계를 분석합니다.

* Standard paired-end, short-read data: 수동 라이브러리 준비 단계와 표준 sequencing by synthesis(SBS) chemistry를 사용하는 WGS 단계를 통해 생성된 데이터

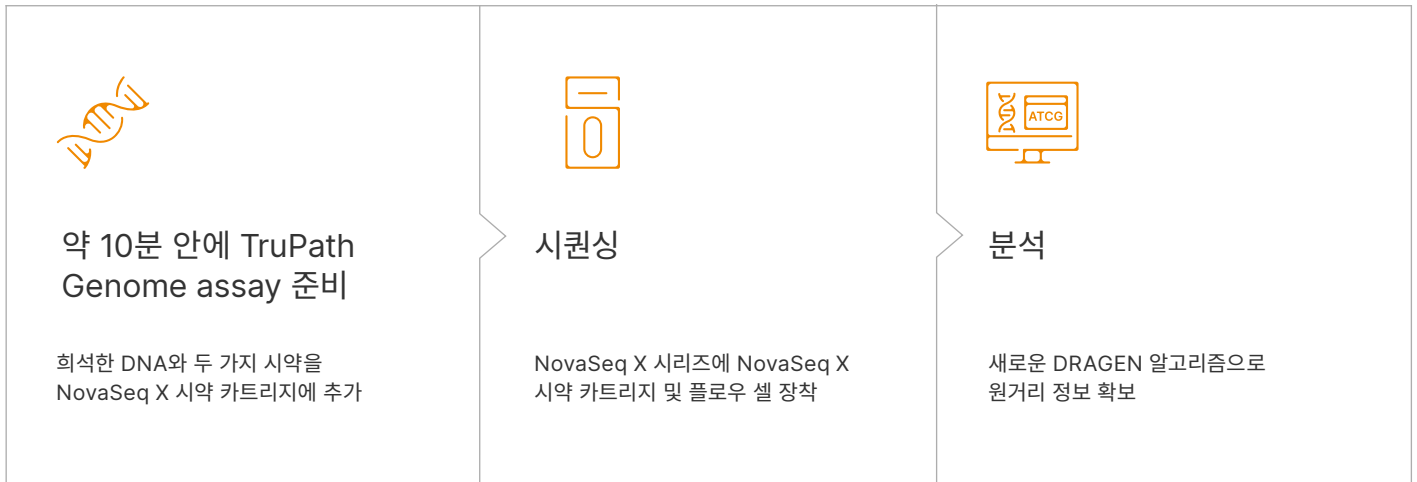


그림 1: TruPath Genome 워크플로우

먼저 DNA와 TruPath Genome 시약 1개를 넣은 라이브러리 튜브 스트립을 NovaSeq X 시약 카트리지에 삽입. 그다음 추가적인 TruPath Genome 시약을 카트리지에 로딩. 마지막으로 카트리지와 플로우 셀을 NovaSeq X 시리즈에 장착한 후 런 시작

샘플 준비

일반적인 상용 키트 또는 고분자량(high molecular weight, HMW) 상용 키트를 사용해 정제된 이중 가닥 DNA(double-stranded DNA, dsDNA)를 추출합니다. 권장 사항에 따라 350 ng의 DNA를 TruPath 태그멘테이션 버퍼(tagmentation buffer)로 희석한 후 라이브러리 튜브 스트립에 넣습니다. 이때 넣는 DNA는 세포주(cell line)나 전혈에서 분리할 수 있습니다. 포르말린 고정, 파라핀 포매(Formalin-fixed, paraffin-embedded, FFPE) 샘플이나 세포유리 DNA(cell-free DNA, cfDNA) 샘플에서 분리한 DNA는 사용할 수 없습니다. 최적의 성능을 위해 DNA 추출 시 큰 절편(60 kb 이상의 길이를 가진 비율이 40%)을 보존하는 방법을 사용하는 것을 권장합니다.³

추가적인 샘플 종류 및 다양한 샘플 품질의 사용과 최적의 샘플 사용량에 대한 지침은 [다양한 샘플 종류 및 품질에 따른 TruPath Genome의 성능 Technical Note](#)를 참조하시기 바랍니다.

시퀀싱

희석한 DNA와 두 가지 TruPath Genome 시약을 넣은 라이브러리 튜브 스트립을 NovaSeq X 시약 카트리지에 삽입한 후 NovaSeq X C2 플로우 셀(인간 유전체 2개 시퀀싱 가능) 또는 C8 플로우 셀(인간 유전체 8개 시퀀싱 가능)과 함께 NovaSeq X 시리즈에 장착합니다.

플로우 셀의 표면에 있는 나노웰은 Illumina의 트랜스포좀[†]으로 프라이밍되어 있습니다. 따라서 DNA를 플로우 셀에 로딩하면 여러 웰에 있는 트랜스포좀이 DNA를 포획하고, 이렇게 DNA가 포획된 나노웰들은 마치 별자리와 같은 패턴을 이룹니다(그림 2A). 트랜스포좀은 태그멘테이션[‡](그림 2B)이라는 과정을 통해 포획된 DNA를 절단한 후 DNA 절편을 플로우 셀의 표면에 결합합니다. 결합되지 않은 DNA와 트랜스포좀은 세척 단계에서 제거됩니다.

나노웰 안에서는 태그멘테이션을 거친 DNA를 사용해 시퀀싱할 클러스터[§]가 생성됩니다. 하나의 긴 DNA 절편에서 유래한 여러 DNA 절편들은 나노웰의 인접한 곳에서 클러스터를 형성합니다. 시퀀싱 단계는 입증된 Illumina의 XLEAP-SBS™ chemistry를 사용하여 NovaSeq X 시리즈에서 2 × 150 bp 런을 수행하는 것으로 마무리됩니다.

분석

시퀀싱이 완료되면, DRAGEN™ Germline 파이프라인이 나노웰의 근접성 정보를 활용하는 새로운 알고리즘을 이용하여 시퀀싱 데이터를 참조 유전체(reference genome)에 정렬(alignment)합니다. 근접성 정보를 기반으로 이전에는 참조 유전체에 매핑되지 않았거나 낮은 신뢰도로만 정렬되었던 리드를 적절하게 배정할 수 있습니다. 이렇듯 일반적인 쇼트 리드를

[†] Transposome: 다이머(dimer)로 존재하는 DNA 트랜스포제이즈 복합체

[‡] Tagmentation: 트랜스포좀을 이용해 DNA 절편을 절단하고 태그(tagging)이라고 불리는 어댑터 시퀀스(adapter sequence)를 추가하는 과정

[§] Cluster: 시퀀싱이 될 플로우 셀에 있는 증폭된 DNA 스폿

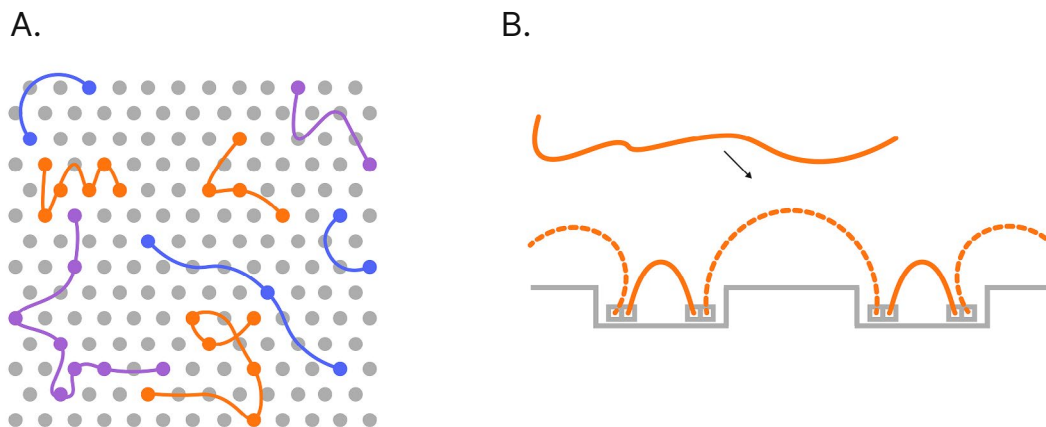


그림 2: Proximity mapped read technology의 원리

(A) DNA가 인접한 여러 나노웰로 흘러들어가 원거리 유전체 정보의 생성에 필요한 근접성 정보의 활용을 가능케 함. (B) DNA가 나노웰(회색 상자) 안의 트랜스포좀에 의해 포획되며, 여기에서 태그멘테이션과 클러스터 생성 과정을 거침

확률적으로 서로 연관시킬 수 있게 되므로 작은 변이(즉, 단일 염기서열 변이(single nucleotide variant, SNV)와 삽입/결실(insertion/deletion, Indel)), 큰 변이(> 50 bp) 그리고 알려진 상동 영역(homologous region)에 대한 포괄적인 변이 검출과 더불어 수백만 염기쌍(base pair) 거리만큼 떨어져 있을 수 있는 작은 변이 쿨의 페이징(phasing)이 가능해집니다.

정확도 높은 쇼트 리드로 완전한 유전체 확보

표준 SBS chemistry는 작은 변이(예: SNV, Indel, 유전자 복제수 변이(copy number variation, CNV))에서의 검출 정확도는 높았으나, 특정 영역에서의 분석과 SV와 같은 변이 타입의 분석에는 어려움을 겪었습니다. TruPath Genome은 쇼트 리드 시퀀싱에 새로운 기능들을 적용하여 보다 완전한 유전체를 제공합니다. 이제 쇼트 리드로 이전에는 분석이 어려웠던 유전체 영역을 보완하고 추가적인 변이 타입에 대한 향상된 검출 정확도를 확보할 수 있습니다.

매핑이 어려운 영역의 커버리지 개선

복잡하고 매우 다형적(polymorphic)인 유전체 영역이나 중복된 유전체 영역은 낮은 커버리지 또는 낮은 매핑 품질(mapping quality, MapQ)로 인해 매핑률(mappability)이 낮아, 조립(assembly)이나 참조 유전체에 대한 정렬이 어려울 수 있습니다. 이전에는 이러한 '암흑 영역(dark region)'⁴에 대한 낮은 커버리지가 낮은 MapQ 점수, 유전체 시퀀스 데이터의 갭, 그리고 전반적으로 제한적인 변이 검출 해상도를 초래했습니다. TruPath Genome은 이러한 갭을 채우기 위해 인접한 클러스터에서 얻은 근접성 정보를 이용하여 올바른 유전체 위치를 배정합니다. 이로써 모호하게 매핑된 리드의 신뢰도 높은 매핑이 가능해집니다(그림 3 및 그림 4).

보다 높은 변이 검출 정확도

TruPath Genome의 작은 변이 검출, SV 검출 및 유전체 페이징(즉, phased sequencing) 성능을 확인하기 위해 TruPath Genome을 HG002 유전체 및 Telomere-to-Telomere(T2T) Consortium의 T2T-Q100 v1.1 v0.019 진리 집합(truth set)과 비교하여 벤치마킹했습니다. Proximity mapped read technology로 높아진 매핑 해상도는 특히

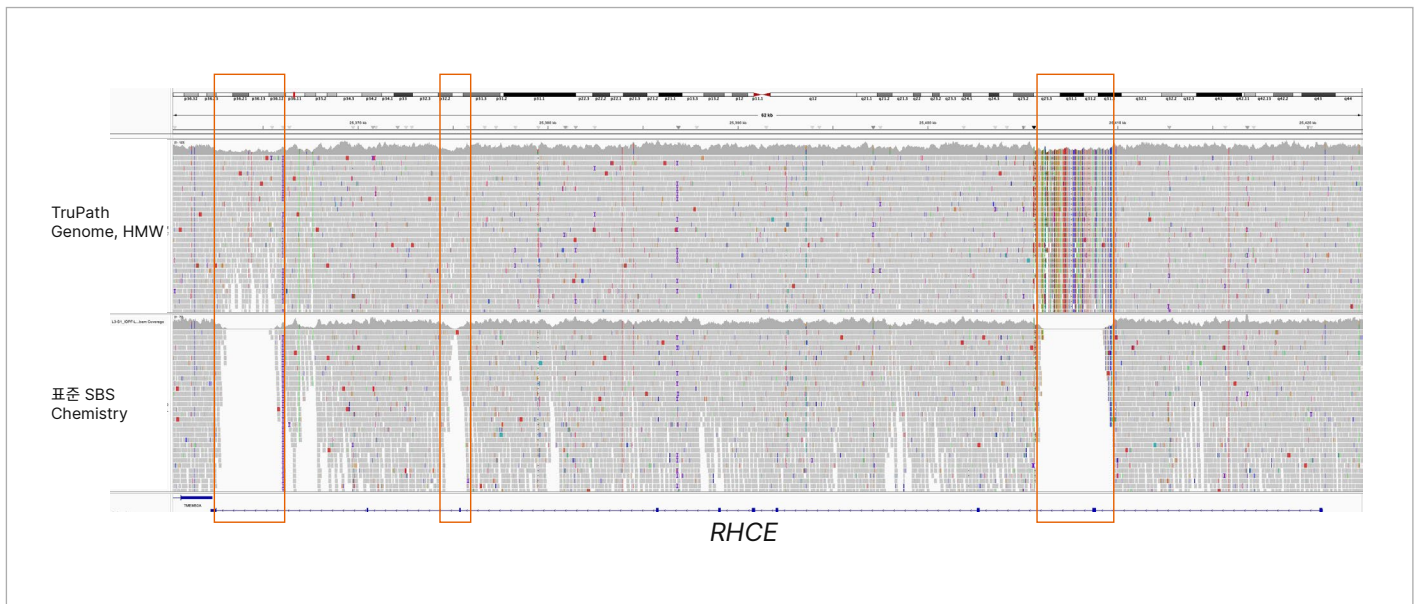


그림 3: 매핑이 어렵고 의학적 관련이 있는 유전자의 효과적인 분석을 지원하는 TruPath Genome

*RHCE*에 인접한 Rh 혈액형군 좌위에 있고 빈번한 유전자 전환(gene conversion) 및 하이브리드 재배열(hybrid rearrangement)과 매우 높은 시퀀스 상동성을 공유하여 일부 위치에서 쇼트 리드 매핑을 모호하게 하는 *RHCE* 유전자에 대한 커버리지가 복구됨. 미스매치(Mismatch)는 용혈성 수혈 반응(hemolytic transfusion reaction) 및 태아/신생아의 용혈성 질환(hemolytic disease of the fetus/newborn, HDFN) 위험을 증가시킬 수 있으므로 정확한 *RHCE* 유전형 분석(genotyping)은 Rh 항원(즉, C/c, E/e) 확인 시 임상적 중요성을 가짐. 표준 SBS chemistry의 데이터 생성에는 Illumina DNA PCR-Free Prep이 사용됨

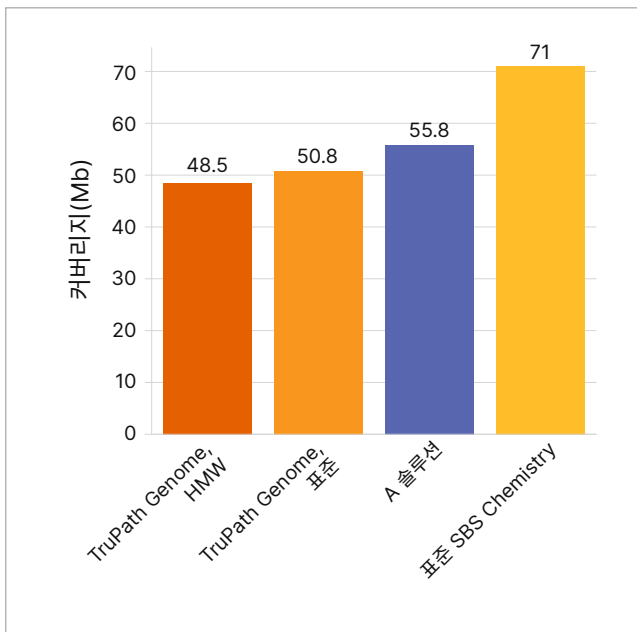


그림 4: 유전체의 '암흑 영역'에 대한 향상된 커버리지를 제공하는 TruPath Genome

TruPath Genome은 Ebbert 외¹의 연구에서 특정 영역을 커버하는 리드 중 90%의 매핑 품질(MAPQ)이 10 미만인 'dark-by-MAPQ'로 정의되어 있는 암흑 영역에 대한 커버리지를 향상시킴. TruPath Genome은 분석이 어려운 유전체 영역에서의 리드 매핑 성능을 향상하여, 암흑 영역 유전체의 크기를 줄임. 표준 SBS chemistry의 데이터 생성에는 Illumina DNA PCR-Free Prep이 사용됨

기존에는 매핑이 어려웠던 유전체의 상동 영역이나 반복 영역에서의 작은 변이 검출 성능을 향상시켰습니다. TruPath Genome은 일반적인 쇼트 리드 시퀀싱이나 변이 검출과 비교했을 때 위음성(false negative, FN) 및 위양성(false positive, FP) 변이 콜의 수를 크게 줄임으로써 높은 정확도를 달성합니다(그림 5).

TruPath Genome은 일반적인 쇼트 리드 시퀀싱 방법보다 SV(> 50 bp) 검출 시 현격히 향상된 성능을 보여줍니다. 예를 들어, SV 재현율(recall)의 경우 표준 SBS chemistry가 86%인 반면 proximity mapped read technology는 94%로 높아진 것이 확인되었습니다(그림 6).

초원거리 페이징

Phased sequencing(유전체 페이징)은 모계 염색체에 있는 대립유전자(allele)와 부계 염색체에 있는 대립유전자를 구분하는 데 사용할 수 있는 하플로타입(haplotype, 일배체형) 정보를 제공합니다.⁸ 이 정보는 연구자가 상염색체 열성 질환(autosomal recessive condition)의 해석에 중요한 역할을 하는 복합 이형접합 변이(compound heterozygous variant)를 분석할 때 유용하게 활용될 수 있습니다.

Proximity mapped read technology는 특히 인간 유전체 페이징에 매우 적합합니다. 플로우 셀에 포획되는 DNA는 수백 킬로베이스(kilobase, kb)에서 수십 메가베이스(megabase, Mb)까지 확장되는 페이징된 블록(phased block)을 생성하여(그림 7 및 그림 8), 하플로타입과 복합 이형접합에 대해 보다 깊은 통찰력을 얻을 수 있도록 해 줍니다.

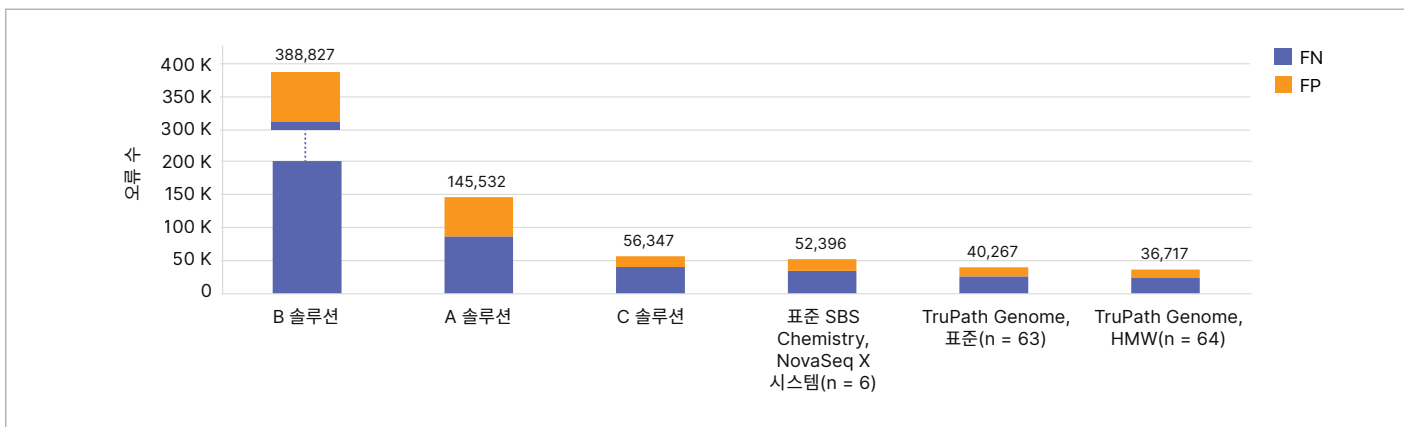


그림 5: 작은 변이 검출 정확도를 높이는 TruPath Genome

다양한 NGS 시스템 및 assay를 사용해 확보한 작은 변이 검출 성능을 Genome in a Bottle Consortium의 NIST T2T-Q100 HG002 SV v1.1 진리 집합과 비교해 벤치마킹함. B 솔루션, A 솔루션 및 C 솔루션에 대한 분석 데이터는 발표된 연구 자료를 통해 확보함.⁵⁻⁷ 표준 SBS chemistry에 대한 데이터는 Illumina DNA PCR-Free Prep 라이브러리를 NovaSeq X 시스템(v1.4 소프트웨어와 10B 플로우 셀 사용)으로 시퀀싱한 후 DRAGEN Germline v4.5.2로 분석해 얻음(6회 반복 실험). TruPath Genome, 표준에 대한 데이터는 일반적인 방법으로 추출한 DNA를 NovaSeq X 시스템(v1.4 소프트웨어와 C8 플로우 셀 사용)으로 시퀀싱한 후 DRAGEN Germline v4.5.2로 분석해 생성함(63회 반복 실험). TruPath Genome, HMW에 대한 데이터는 HMW 방법으로 추출한 DNA를 NovaSeq X 시스템(v1.4 소프트웨어 및 C8 플로우 셀 사용)으로 시퀀싱한 후 DRAGEN Germline v4.5.2로 분석해 생성함(64회 반복 실험). FN = false negative variant call(위음성 변이 콜), FP = false positive variant call(위양성 변이 콜)

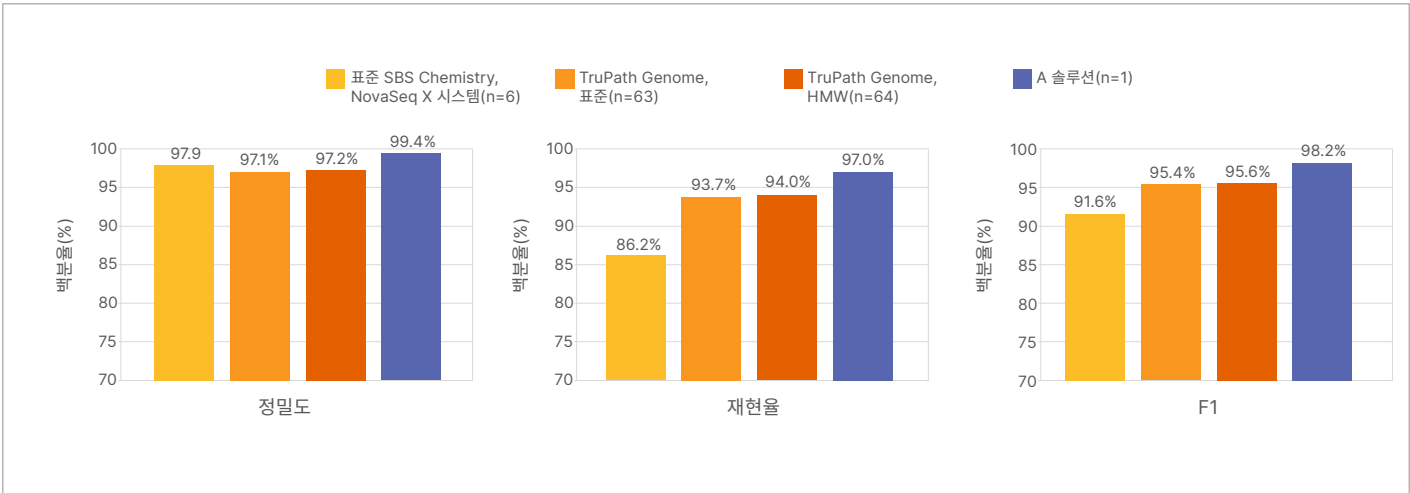


그림 6: SV 검출 성능을 높이는 TruPath Genome

DRAGEN Germline v4.5.2로 확인한 표준 SBS chemistry 대비 TruPath Genome의 SV 검출 성능. 이 분석에는 Genome in a Bottle Consortium의 NIST T2T-Q100 HG002 SV v1.1 진리 집합과 신뢰할 수 있는 SV BED 파일을 사용함. 벤치마킹은 Genome in a Bottle의 SV 벤치마킹 지침에 따라 Truvari v4.2.2에서 "bench" 및 "refine" 명령어를 사용해 진행함. 표준 SBS chemistry에 대한 데이터는 Illumina DNA PCR-Free Prep 라이브러리를 NovaSeq X 시스템(v1.4 소프트웨어 및 10B 플로우 셀 사용)으로 시퀀싱하여 생성함(6회 반복 실험). TruPath Genome, 표준에 대한 데이터는 일반적인 방법으로 추출한 DNA를 NovaSeq X 시스템(v1.4 소프트웨어와 C8 플로우 셀 사용)으로 시퀀싱하여 생성함(63회 반복 실험). TruPath Genome, HMW에 대한 데이터는 HMW 방법으로 추출한 DNA를 NovaSeq X 시스템(v1.4 소프트웨어 및 C8 플로우 셀 사용)으로 시퀀싱하여 생성함(64회 반복 실험). A 솔루션에 대한 데이터는 발표된 연구 자료를 통해 확보함.⁶ 정밀도(Precision)는 진양성(true positive) 비율을 측정함. 재현율(Recall)은 1개의 샘플 내 변이의 수 대비 진음성(true negative)의 %를 나타냄. F1 점수(%)는 진양성 및 진음성 결과를 전체 결과에 대한 비율로 계산한 값을 나타냄. 점수가 높을수록 레퍼런스 데이터를 기준으로 정확도가 향상된 것을 의미함

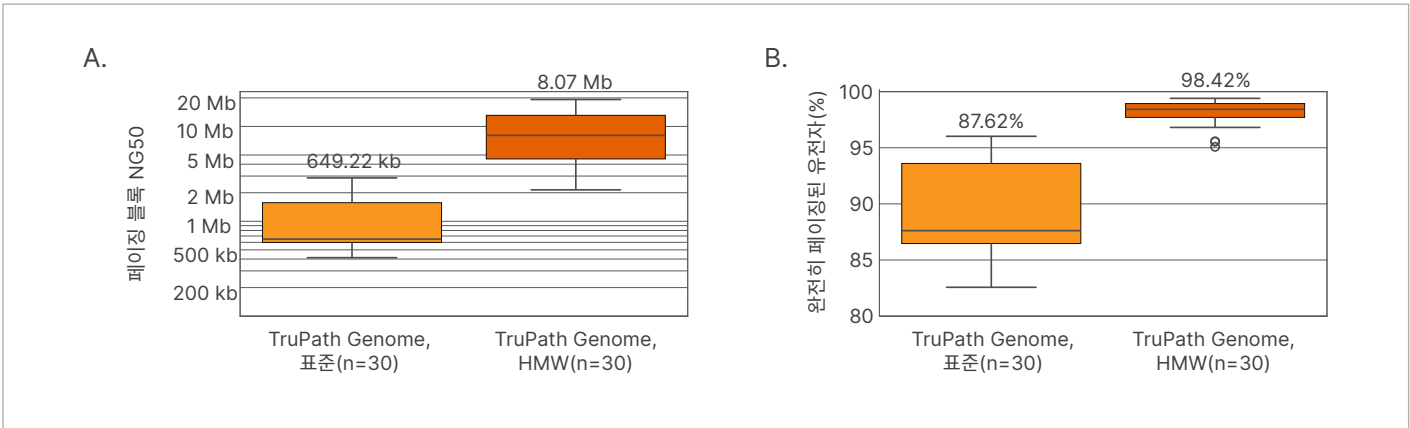
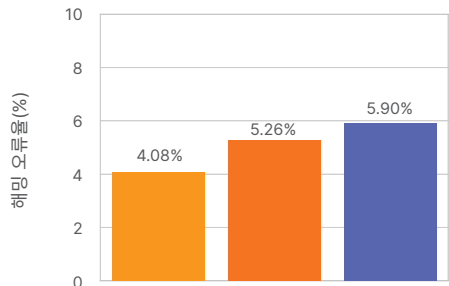
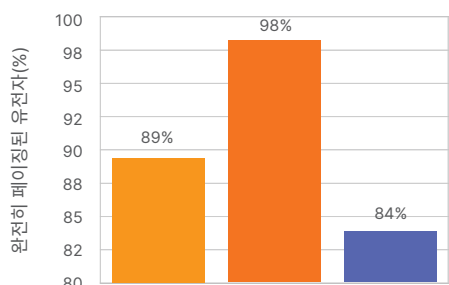
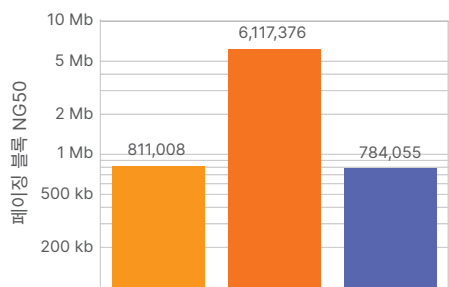


그림 7: 30개의 매칭된 HMW 추출 및 표준 추출 세포주에 대한 TruPath Genome 초원거리 페이징 성능

DRAGEN Germline v4.5.2에서 DRAGEN 페이징 기능을 사용하여 1~22번 염색체에 걸쳐 페이징 블록을 측정함. TruPath Genome의 데이터는 HMW 추출 및 표준 추출 방법을 30개의 다양한 Coriell 세포주 샘플에 적용하여 확보함. '페이징 블록 NG50'이란 타겟 영역(유전체 또는 기타)의 50%가 페이징된 이후의 페이징 블록의 길이를 의미함. 특정 타겟 영역의 50%를 페이징할 수 없는 기술의 NG50은 0 bp로 측정됨. '완전히 페이징된 유전자(%)'란 명시된 유전자 목록(즉, Gencode v44 genes.gtf)에서 1개의 페이징 블록 안에 완전하게 들어 있는 유전자 영역의 백분율을 의미함



TruPath Genome, TruPath Genome HMW, A 솔루션 표준

그림 8: HG002 진리 집합 데이터를 비교해 벤치마킹한 TruPath Genome의 유전체 페이징 성능

DRAGEN Germline v4.5.2에서 DRAGEN 페이징 기능을 사용하여 1~22번 염색체에 걸쳐 페이징 블록을 측정함. A 솔루션 데이터 Phased VCF 파일은 https://downloads.pacbcloud.com/public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2/에서 확보함. TruPath Genome 데이터는 HG002 세포주 샘플과 HMW 추출 및 표준 추출 샘플을 사용한 시퀀싱 반복 실험을 통해 확보함. '페이징 블록 NG50'이란 타겟 영역(유전체 또는 기타)의 50%가 페이징된 이후의 페이징 블록의 길이를 의미함. '완전히 페이징된 유전자(%)'란 명시된 유전자 목록(즉, Gencode v44 genes.gtf)에서 1개의 페이징 블록 안에 완전하게 들어 있는 유전자 영역의 백분율을 의미함. 페이징 해밍 오류율은 https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/analysis/NI-ST_HG002_DraftBenchmark_defrabbv0.020-20250117/GRCh38_HG2-T2TQ100-V1.1_smvar.vcf.gz에서 다운로드한 T2T Q100 진리 집합⁹을 비교하여 벤치마킹함

TruPath Genome은 일반적인 DNA 추출 방법 사용 시 중앙값을 기준으로 모든 유전자의 약 91%, HMW DNA 추출 방법 사용 시에는 유전자의 약 97%를 완전히 페이징합니다. 또한 TruPath Genome은 일반적인 DNA 추출 방법과 HMW DNA 추출 방법 사용 시 모든 이형접합 SNV의 약 98%⁹를 페이징합니다.

HMW DNA 추출 방법 사용 시 DRAGEN 페이징을 적용해 얻은 TruPath Genome 데이터는 해밍 오류**가 약 4%로, *HapCut2*와 같은 공개적으로 사용 가능한 도구가 기록한 약 6%보다 낮습니다.¹⁰

상동성이 높은 영역에서의 변이 검출

패럴로그(Paralog) 영역에서는 높은 시퀀스 상동성으로 인해 일반적인 쇼트 리드를 이용해 정확한 변이 검출이 어려워, 모호한 리드 매핑과 변이 검출 오류가 발생합니다. 이를 해결하기 위해서는 종종 멀티플렉스 라이게이션 의존 프로브 증폭(multiplex ligation-dependent probe amplification, MLPA), 원거리 중합효소 연쇄반응(long-range polymerase chain reaction, LR-PCR)과 같이 많은 비용이 드는 추가 분석이 필요합니다. TruPath Genome은 플로우 셀 근접성 정보와 새로운 DRAGEN MRJD(Multi-Region Joint Detection) 알고리즘을 결합하여 패럴로그 유전자에서 신뢰할 수 있는 하플로타입으로 분리된 유전자 복제수 인식 드 노보 변이 검출(*de novo*, haplotype-resolved, and copy-number-aware variant calling) 결과를 제공합니다(그림 9 및 그림 10).

TruPath Genome은 정확도 높은 매핑을 위해 MRJD를 이용하여 유전체 전체에 걸쳐 연쇄 카피(tandem copy) 또는 비연쇄 카피(non-tandem copy)가 존재하는 것으로 알려진 의학적으로 관련이 있는 15개의 패럴로그 유전자(표 1)를 타겟으로 하여, 하플로타입으로 분리된 작은 변이를 검출합니다(그림 10).

표 1: TruPath Genome이 MRJD 분석 시 타겟으로 하는 유전자

<i>CFHR1</i>	<i>CFHR2</i>	<i>CFHR3</i>
<i>CFHR4</i>	<i>CYP2D6</i>	<i>CYP11B1/2</i>
<i>NCF1</i>	<i>PMS2</i>	<i>RCCX</i> (<i>CYP21A2</i> 및 <i>TNXB</i>)
<i>SMN1/2</i>	<i>STRC</i>	<i>USP18</i>

⁹Percent heterozygous variants phased: 페이징된 이형접합 변이의 %는 페이징된 이형접합 SNV의 백분율을 의미하며, 페이징된 SNV의 수를 이형접합 SNV의 수로 나누어 계산함

**Hamming error: 변이가 잘못된 모계 또는 부계 가닥에 배정될 때 발생함

향상된 짧은 연쇄 반복(STR) 해상도

짧은 연쇄 반복(Short tandem repeat, STR)은 유전체의 반복적인 영역으로, 이 영역이 정상 범위를 넘는 길이로 확장되면 취약 X 증후군(Fragile X syndrome), 근위축성 측삭 경화증(amyotrophic lateral sclerosis, ALS), 헌팅턴병(Huntington's disease) 등 여러 유전 질환과 관련이 있는 돌연변이를 유발할 수 있습니다.¹¹ 그런데 가장 관련성이 높은 확장된 대립유전자는 일반적인 쇼트 리드 시퀀싱 데이터의 리드 길이(read length)를 초과하는 경우가 많기 때문에 이러한 영역의 정확한 크기 추정 및 리드 복원이 쉽지 않았습니다. TruPath Genome은 근접성 정보와 고품질 쇼트 리드 데이터를 결합하여, 정확한 STR 크기 측정에 필요한 리드의 강화된 복원 능력과 페이징된 정보를 활용하는 두 카피의 유전형 분석을 통해 향상된 해상도를 제공합니다(그림 11).

복합적인 SV 검출을 위한 초원거리 정보

TruPath Genome의 거대 구조적 재배열(large structural rearrangement) 분석 능력은 기존의 변이 검출 성능 벤치마크를 넘어서는 새로운 역량을 포함합니다. TruPath Genome은 유전체 영역의 특정 페어 사이 인접해 있는 클러스터^{††}로부터 얻은 리드에 대한 정보를 포착하여, '코로케이션 플롯(colocation plot)'이라고 불리는 고해상도의 유전체 구조 지도를 시각적으로 표현할 수 있도록 해 줍니다(그림 12).

†† Proximal cluster: 플로우 셀에서 물리적으로 서로 인접해 있는 클러스터

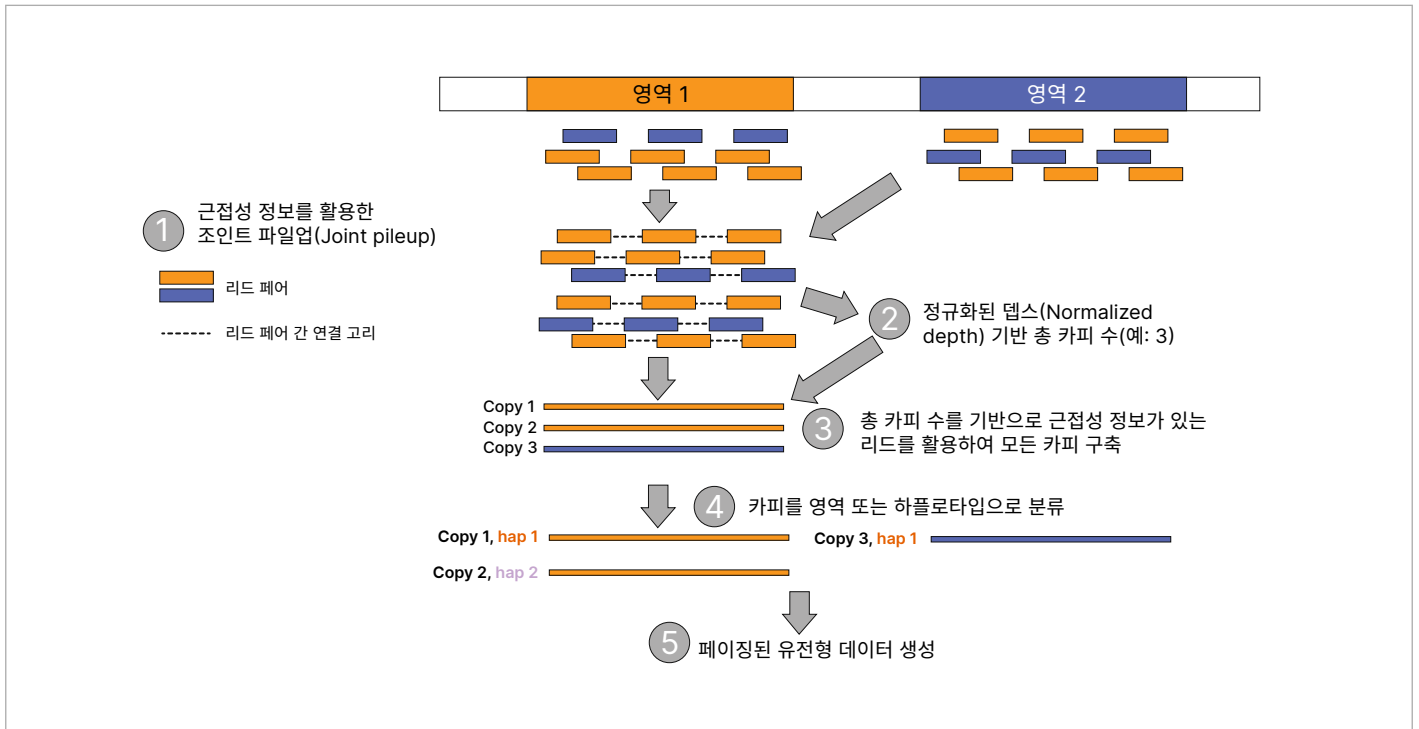
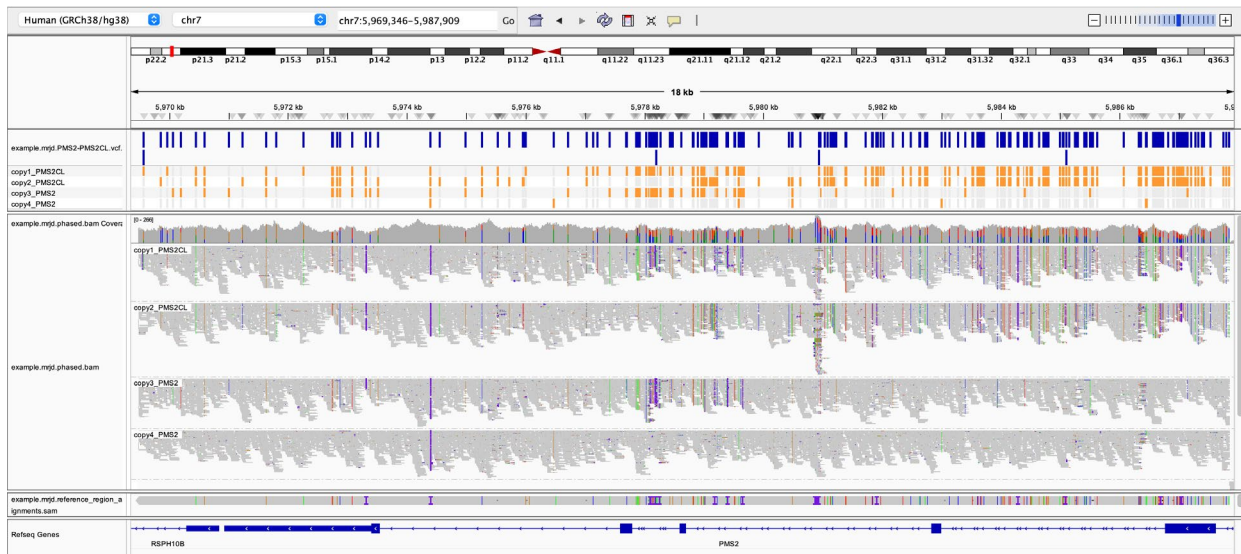


그림 9: 분절 중복 영역 내 변이 검출에 사용되는 MRJD 방법의 도식

1. MRJD는 조인트 분석을 위해 특정 분절 중복 영역(segmental duplication region) 세트에서 해당 영역 세트에 속하는 특정 카피에 매핑된 모든 리드를 추출함. 근접성 정보를 활용해 이러한 리드 간 연결 고리를 만들고 동일한 분자로부터 유래한 리드 세트를 식별함. 2. 1단계에서 추출한 모든 리드의 수는 정규화(normalization) 및 GC 교정(correction) 단계를 거친 후 사용한 샘플 내 분절 중복 시퀀스의 총 카피 수를 추정하는 데 사용됨. 3. 2단계에서 추정한 해당 영역의 카피 수를 근거로 하며 근접성 정보로 연결된 완전한 리드 세트에 적용되는 조인트 지노타이핑(joint genotyping)을 실행하여, 사용한 샘플 내 각각의 분절 중복 시퀀스 카피에 존재하는 모든 변이를 식별하고 페이징함. 4. 각각의 완전히 페이징된 카피는 특정 참조 유전체(reference genome) 위치에 배정되거나(영역이 먼 참조 위치에 있는 경우), 다른 카피로 페이징되어 하플로타입이 됨(분절 중복 영역이 참조 유전체에서 나란히 있거나 인접해 있을 때만 가능). 5. 모든 카피에 걸쳐 완전히 페이징된 변이, 각 카피의 하플로타입 배정, 각 참조 유전체 위치별 카피의 배정 및 각 카피에 배정된 리드에 관한 정보는 MRJD 특정 VCF, JSON 및 BAM 파일에 저장됨

A. *PMS2* 유전자



B. *SMN* 유전자

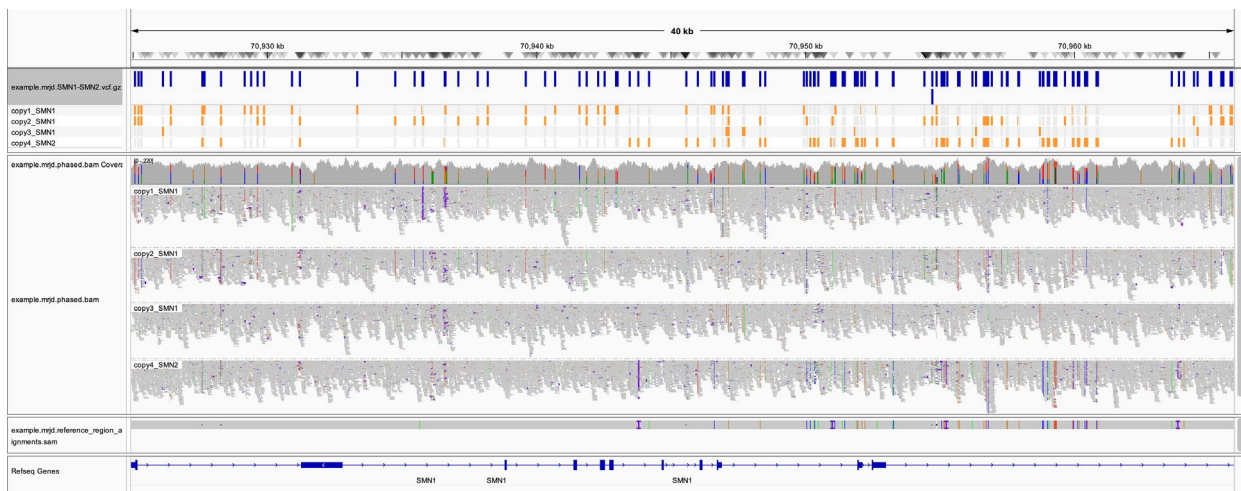
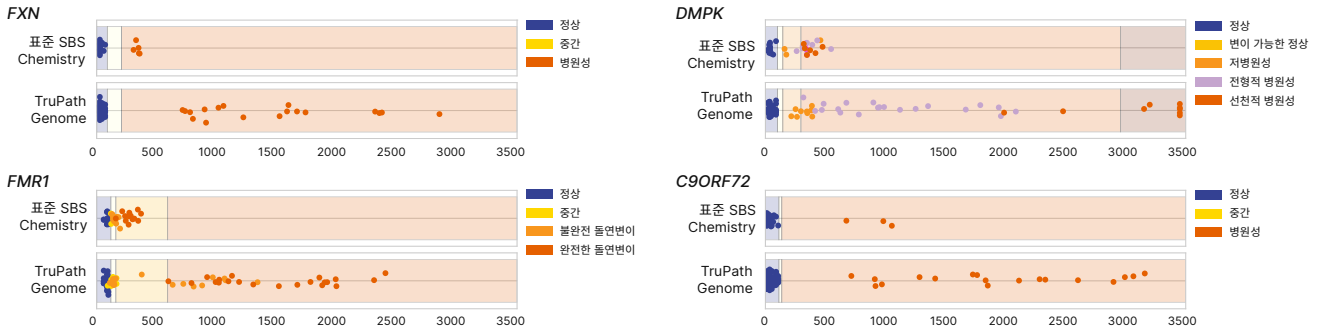


그림 10: TruPath Genome과 MRJD를 사용한 패럴로그 유전자의 변이 검출 및 리드 배정 결과의 예시

(A) *PMS2* 참조 유전자 위치에서 표현된 *PMS2/PMS2CL* 유전자-유사유전자(pseudogene) 페어의 모든 카피에 걸쳐 페이징된 변이 콜 및 관련 뒷받침하는 리드(supporting read). *PMS2/PMS2CL* 플랭크(flank)와의 근접성 연결 고리를 근거로 보았을 때, 4개의 카피 중 2개는 *PMS2CL*(유사유전자) 카피로, 2개는 *PMS2*(활성 유전자) 카피로 확인됨. Copy3(*PMS2*)는 *PMS2CL* 참조 시퀀스와 매칭되는 변이가 존재하는 것으로 보아 13번 엑손과 14번 엑손에 겹친 유전자 전환으로 여겨지는 이벤트를 포함하고 있음. *PMS2CL* 참조 시퀀스의 *PMS2* 참조 시퀀스에 대한 정렬은 위에서 세 번째 패널에서 확인할 수 있음.

(B) *SMN1* 참조 유전자 위치에서 표현된 *SMN1/SMN2* 유전자-유사유전자 페어의 모든 카피에 걸쳐 페이징된 변이 콜 및 관련 뒷받침하는 리드. 각 페이징된 카피의 *SMN1* c.840 포지션에 있는 염기에 따르면(*SMN1* 카피는 C 염기를, *SMN2* 카피는 T 염기를 가지고 있음), 4개의 카피 중 3개는 *SMN1*(활성 유전자) 카피로, 1개는 *SMN2*(비활성 유전자)로 확인됨



긴 대립유전자 크기 추정(bp)

그림 11: 보다 정확한 STR 확장 측정 결과를 제공하는 TruPath Genome

STR의 실제 분류를 각각 다른 색상의 점으로 표현함. 각 유전자좌위에서의 다양한 확장 클래스를 정의하는 데 사용되는 여러 분류 범위는 각각 다른 색상의 줄로 표현함. X축은 TruPath Genome 또는 표준 SBS chemistry(Illumina DNA PCR-Free Prep)를 사용하여 DRAGEN STR로 추정된 STR 크기를 나타냄. TruPath Genome의 분류는 실제 분류와 훨씬 더 높은 일치도를 보이며 더 넓은 범위의 STR 길이를 포함함. 표준 SBS chemistry의 반복 실험: 78개(78명으로부터 수집)의 생체 샘플 사용. TruPath Genome의 반복 실험: 135개(42명으로부터 수집)의 생체 샘플 사용

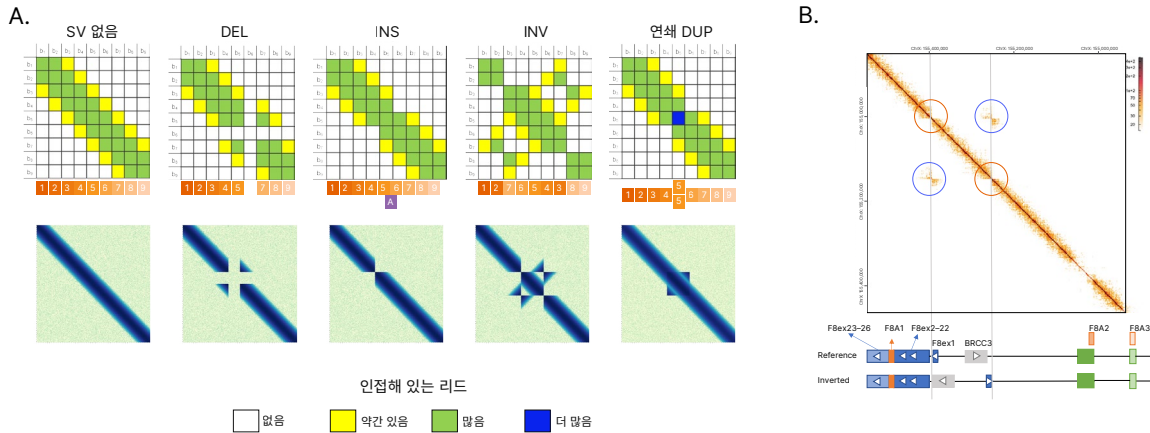


그림 12: TruPath Genome 코로케이션 플롯을 사용해 시각화한 SV 이벤트

(A) 모든 유전체 bin 페어에 매핑되는 플로우 셀 내 서로 인접해 있는 리드의 수를 측정하여 2차원 코로케이션 맵을 생성함. X축과 Y축은 참조 유전체 bin을 나타냄. 2개의 bin 사이에 연결 수가 많은 것은 사용한 샘플 내 유전체 공간에서 이 2개의 bin이 서로 인접해 있음을 의미함. SV가 없는 영역의 경우, 참조 유전체에서 서로 인접해 있는 유전체 bin이 사용한 샘플에서도 서로 인접해 있으므로 코로케이션 플롯의 대각선 근처에서 연결 수가 많은 것으로 나타남. SV가 있는 영역의 경우, 참조 유전체에서 서로 인접해 있는 유전체 bin이 사용한 샘플의 유전체에서는 인접해 있지 않을 수도 있으며, 반대로 참조 유전체에서 서로 멀리 떨어져 있는 유전체 bin이 사용한 샘플에서는 인접해 있을 수도 있음. 이러한 연결 변화는 코로케이션 맵에 반영되어 다양한 형태의 비대각 신호(off-diagonal signal)와 대각선 연결의 부재로 표현됨. (B) 중증 A형 혈우병(hemophilia A)과 연관성이 있는 F8 유전자좌위에서의 1번 인트론 역위(inversion)에 대한 코로케이션 맵의 예시. 비대각 영역에 예상되는 모래시계형 신호(파란색 원)이 있고 이벤트 경계에 대각선 신호(주황색 원)가 없는 것을 확인할 수 있음. 플롯 하단에는 참조 유전체에서의 해당 좌위 및 역위된 유전자 영역에 대한 개략도가 제공됨

가장 간편한 WGS 워크플로우

Proximity mapped read technology를 기반으로 하는 Illumina의 TruPath Genome은 인간 유전체 시퀀싱의 판도를 바꿉니다. 이 혁신적인 워크플로우는 기존의 번거로운 라이브러리 준비 단계를 제거하여 획기적인 간편성을 제공하며, 정확도 높은 일반적인 쇼트 리드 시퀀싱 데이터와 근접성 정보를 활용해 긴 길이의 DNA 가닥 정보를 보존합니다. 그 결과, 쇼트 리드의 우수한 데이터 품질과 정확도 높은 작은 변이 검출력을 기반으로, 향상된 인간 유전체 매핑 커버리지를 제공하고 유전 질환에 대한 이해를 돕는 강력한 assay가 완성되었습니다.

상세 정보

[Illumina TruPath Genome](#)

[Proximity mapped read technology](#)

[NovaSeq X 시리즈](#)

제품 목록	
제품	카탈로그 번호
Illumina TruPath Genome (NovaSeq X C8 flow cell)	20157405
Illumina TruPath Genome (NovaSeq X C2 flow cell)	20157406

참고 문헌

1. Ebbert MTW, Jensen TD, Jansen-West K, et al. [Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight.](#) *Genome Biol.* 2019;20(1):97. Published 2019 May 20. doi:10.1186/s13059-019-1707-2
2. Ryan NM, Corvin A. [Investigating the dark-side of the genome: a barrier to human disease variant discovery?](#) *Biol Res.* 2023;56(1):42. Published 2023 Jul 20. doi:10.1186/s40659-023-00455-0
3. Illumina. TruPath Genome performance with samples of varying type and quality. [illumina.com/content/dam/illumina-marketing/documents/products/technotes/trupath-genome-sample-info-tech-note-m-gl-03932](https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/trupath-genome-sample-info-tech-note-m-gl-03932). Published February 2026. Accessed February 24, 2026.
4. Wadsworth ME, Page ML, Aguzzoli Heberle B, Miller JB, Steely C, Ebbert MTW. [Sequencing the gaps: dark genomic regions persist in CHM13 despite long-read advances.](#) Preprint. *bioRxiv.* 2025;2025.05.23.655776. Published 2025 May 28. doi:10.1101/2025.05.23.655776
5. Ultima Genomics. Genome-in-a-Bottle (GIAB) samples HG001-HG005. <https://cdn.sanity.io/files/l7780ks7/production-2024/0a1b6a62a6da3e3fcafb81cad4c8ff2ffe85dd41.pdf> Published February 2025. Accessed January 29, 2026.
6. PacBio. Index of /public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2. https://downloads.pacbcloud.com/public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2/. Accessed January 29, 2026.
7. Asri M, Chang PC, Mier JC, et al. [Pangenome-aware DeepVariant.](#) Preprint. *bioRxiv.* 2025;2025.06.05.657102. Published 2025 Jun 6. doi:10.1101/2025.06.05.657102
8. Browning SR, Browning BL. [Haplotype phasing: existing methods and new developments.](#) *Nat Rev Genet.* 2011; 12(10):703-714. doi: 10.1038/nrg3054
9. Hansen NF, Dwarshuis N, Ji HJ, et al. [A complete diploid human genome benchmark for personalized genomics.](#) Preprint. *bioRxiv.* 2025;2025.09.21.677443. Published 2025 Sep 21. doi:10.1101/2025.09.21.677443
10. Data on file. Illumina, Inc., 2026.
11. Steely CJ, Watkins WS, Baird L, Jorde LB. [The mutational dynamics of short tandem repeats in large, multigenerational families.](#) *Genome Biol.* 2022;23(1):253. Published 2022 Dec 12. doi:10.1186/s13059-022-02818-4



무료 전화(한국) 080-234-5300
techsupport@illumina.com | www.illumina.com

© 2026 Illumina, Inc. All rights reserved.
모든 상표는 Illumina, Inc. 또는 각 소유주의 자산입니다.
특정 상표 정보는 www.illumina.com/company/legal.html을 참조하십시오.
M-KR-00344 KOR