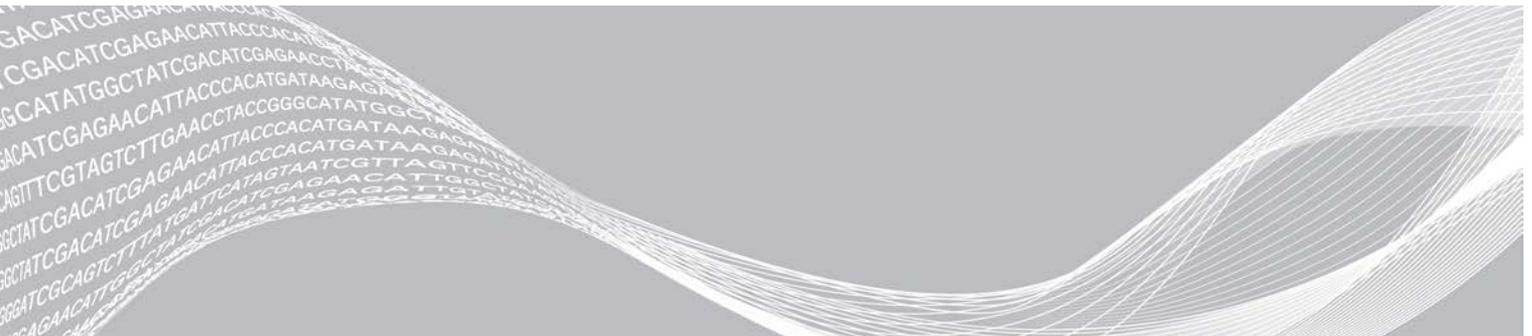


# Local Run Manager Somatic Variant 분석 모듈

## NextSeq 550Dx 워크플로우 가이드

체외 진단용

개요	3
실행 정보 입력	3
분석법	5
실행 및 샘플 데이터 보기	7
분석 보고서	7
분석 출력 파일	8
베이스 콜링 및 인덱스 다양성	15
개정 기록	16
기술 지원	17



이 문서와 이 문서에 수록된 내용은 Illumina, Inc. 및 그 자회사("Illumina")의 재산으로, 여기에 설명된 제품의 사용과 관련하여 전적으로 계약상 보증된 고객만을 위해 사용할 수 있으며 그 외의 목적으로는 사용할 수 없습니다. 이 문서와 이 문서에 수록된 내용은 다른 목적으로 사용되거나 배포될 수 없으며, Illumina의 사전 서면 승인 없이는 어떠한 방식으로도 달리 전달, 공개하거나 복제할 수 없습니다. Illumina는 이 문서를 통해 특허, 상표, 저작권 또는 관습법적 권한이나 유사한 타사 권한에 따라 어떠한 라이선스도 양도하지 않습니다.

이 문서에 설명된 제품의 올바르고 안전한 사용을 위해 이 문서의 지침은 적절한 교육을 받았고 자격을 갖춘 사람에 의해 엄격하고 정확하게 준수되어야 합니다. 해당 제품을 사용하기 전에 이 문서의 모든 내용을 철저히 읽고 숙지해야 합니다.

여기에 설명한 모든 지침을 완전히 읽지 않거나 명확하게 따르지 않을 경우 제품 손상, 사용자나 다른 사람의 신체 부상, 재산상의 손해가 발생할 수 있으며 제품에 적용되는 모든 보증이 무효가 됩니다.

Illumina는 여기에 설명된 제품(그 부품 또는 소프트웨어 포함)을 잘못 사용하여 발생하는 일에 대해 어떠한 책임도 지지 않습니다.

© 2021 Illumina, Inc. All rights reserved.

모든 상표는 Illumina, Inc. 또는 각 소유주의 자산입니다. 구체적인 상표 정보는 [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html) 을 참조하십시오.

## 개요

Local Run Manager Somatic Variant 모듈은 Illumina TruSeq Custom Amplicon Kit Dx 에세이 및 NextSeq 550Dx와 함께 사용합니다. Somatic Variant 모듈과 함께 사용할 경우 이 에세이는 FFPE(Formalin-Fixed Paraffin-Embedded: 포르말린 고정 파라핀 임베딩) 조직에서 얻은 DNA의 시퀀싱에 사용되는 라이브러리 프레임에 사용합니다. 에세이는 낮은 변이 빈도로 Somatic 돌연변이를 검출합니다.

분석 모듈은 증폭된 DNA 또는 변이의 경우 앰프리콘의 짧은 영역을 평가합니다. 앰프리콘에 대한 집중 시퀀싱은 다수의 샘플에서 특정 영역에 대한 높은 커버리지를 지원합니다. 에세이 모듈은 이차 분석으로서 기능하며 정방향 및 역방향 올리고 풀을 포함한 이중가닥 접근법을 사용하여 시퀀싱 실행에서 보고서를 생성합니다. 패키지 설명서 *TruSeq Custom Amplicon Kit Dx(문서 번호 1000000029772)*를 참조하십시오.

Somatic Variant 분석 모듈에는 300 Cycle 시퀀싱 소모품이 필요합니다. 자세한 내용은 *NextSeq 550Dx High Output Reagent Kit v2* 또는 *NextSeq 550Dx High Output Reagent Kit v2.5* 패키지 설명서를 참조하십시오.

## 본 가이드에 대하여

본 가이드에는 Somatic Variant 분석 모듈의 시퀀싱 및 분석 실행 매개변수 설정에 대한 지침이 나와 있습니다. Local Run Manager 대시보드 및 시스템 설정에 대한 정보는 *NextSeq 550Dx 기기 Reference 가이드(문서 번호 1000000009513)*를 참조하십시오.

## Local Run Manager 보기

Local Run Manager 인터페이스는 NextSeq 550Dx 작동 소프트웨어(NOS) 내에서 또는 웹 브라우저를 통해 볼 수 있습니다. 지원되는 웹 브라우저는 Chrome입니다.



### 참고

지원되지 않는 브라우저를 사용하는 경우 "Confirm Unsupported Browser(지원되지 않는 브라우저 확인)" 메시지가 표시되면 지원되는 브라우저를 다운로드합니다. 지원되는 Chrome 버전을 다운로드하려면 "[here\(여기\)](#)"를 선택합니다.

## 기기 모니터에서 보기

- 1 기기 모니터에서 Local Run Manager 인터페이스를 보려면 다음 옵션 중 하나를 선택합니다.
  - ▶ NOS Home(홈) 화면에서 **Local Run Manager**를 선택합니다. 작업이 끝나면 오른쪽 상단 모서리의 X를 클릭하여 NOS로 돌아갑니다.
  - ▶ Minimize NOS(NOS 최소화) 아이콘을 선택하고 기기에 있는 Chrome 웹 브라우저를 연 후 주소 표시줄에 <http://localhost>를 입력합니다. 관리자만 NOS를 최소화할 수 있습니다.

## 네트워크로 연결된 컴퓨터에서 보기

- 1 기기와 동일한 네트워크에 액세스할 수 있는 컴퓨터에서 Chrome 웹 브라우저를 열고, 기기 IP 주소 또는 기기 이름을 사용하여 연결합니다. 예를 들면, <http://myinstrument>입니다.

## 실행 정보 입력

### 매개변수 설정

- 1 Local Run Manager에 로그인합니다.
- 2 **Create Run(실행 생성)**을 선택한 다음 **Somatic Variant**를 선택합니다.

- 3 분석 전반의 시퀀싱에서 실행을 식별할 수 있도록 실행 이름을 입력합니다. 영숫자, 공백, 밑줄 또는 대시를 사용합니다.
- 4 **[선택 사항]** 실행을 식별하는 데 도움이 될 실행 설명을 입력합니다. 영숫자, 공백, 밑줄 또는 대시를 사용합니다.
- 5 드롭다운 목록에서 샘플 개수와 인덱스 세트를 선택합니다. 아래 정보를 참조하여 선택하십시오.
  - ▶ 드롭다운 목록에는 인덱스 세트가 있는 샘플이 여러 개 있습니다. 예를 들어 24-Set 1은 인덱스 세트 1의 인덱스가 있는 테스트 샘플이 24개임을 의미합니다.
  - ▶ 인덱스 세트 숫자는 각기 다른 i5 인덱스 세트를 나타냅니다. Set 1 및 Set 2 모두 인덱스 다양성을 제공합니다. 단일 세트가 크게 모자라는 일이 없도록 인덱스 세트는 2개가 제공됩니다.
  - ▶ 테스트하는 샘플의 개수와 가장 근접한 샘플 개수를 선택합니다. 정확한 샘플 개수가 목록에 없을 경우에는 이와 가장 근접한 개수를 선택하되 테스트하는 개수보다는 적어야 합니다. 예를 들어 18개의 샘플을 테스트하고자 한다면 16을 선택합니다.
  - ▶ 인덱스 다양성 요건을 충족하는 샘플 웰과 인덱스 조합은 녹색으로 강조표시됩니다. 다른 웰과 인덱스 조합을 선택할 경우 실행을 저장할 때 인덱스 다양성 요건이 충족되지 않으면 알림을 받습니다.

## 실행을 위해 메니페스트 파일 가져오기

- 1 액세스 가능한 네트워크 위치 또는 USB 드라이브에서 가져오하고자 하는 메니페스트를 이용할 수 있는지 확인합니다.
- 2 **Import Manifests**(메니페스트 가져오기)를 선택합니다.
- 3 메니페스트 파일을 탐색하고 추가하고자 하는 메니페스트를 선택합니다.



### 참고

Somatic Variant 분석 모듈을 사용하여 모든 실행에 메니페스트 파일을 사용할 수 있는지 확인하기 위해 Module Settings(모듈 설정)를 사용하여 메니페스트를 추가합니다. 이 기능을 사용하려면 관리 사용자 권한이 필요합니다. 자세한 내용은 *NextSeq 550Dx 기기 Reference 가이드(문서 번호 1000000009513)*를 참조하십시오.

## 실행을 위해 샘플 지정

옵션 한 가지와 그에 따른 지침을 사용하여 실행용 샘플을 지정합니다.

- ▶ **샘플을 수동으로 입력** - Create Run(실행 생성) 화면의 빈 표를 사용합니다.
- ▶ **샘플 가져오기** - 심표로 구분된 값(\*.csv) 형식의 외부 파일을 탐색합니다. Create Run(실행 생성) 화면에서 템플릿을 다운로드할 수 있습니다.

샘플 표를 입력한 후에 외부 파일로 샘플 정보를 내보낼 수 있습니다. 라이브러리 프랩 시 또는 다른 실행에 사용할 파일을 가져올 때 이 파일을 Reference로 사용합니다.

## 수동으로 샘플 입력

- 1 Sample Name(샘플 이름) 필드에 고유한 샘플 이름을 입력합니다. 알파벳과 숫자, 대시, 또는 밑줄을 사용합니다. 다른 풀의 해당 웰에 샘플 이름이 자동으로 입력됩니다.
- 2 **[선택 사항]** 양성 또는 음성 대조군 샘플은 오른쪽 클릭 후 대조군 유형을 선택합니다. 한 샘플 웰의 대조군은 동일한 대조군이 있는 다른 풀의 해당 웰에 자동으로 입력됩니다.
- 3 **[선택 사항]** Sample Description(샘플 설명) 필드에 샘플 설명을 입력합니다.

알파벳과 숫자, 대시, 또는 밑줄을 사용합니다.

다른 폴의 해당 웰에 샘플 설명이 자동으로 입력됩니다.

샘플 설명은 샘플 ID와 연결됩니다. 이후 실행에서 동일한 샘플 ID가 사용되면 샘플 설명이 덮어쓰기됩니다.

- 4 Index 1 (i7)(인덱스 1(i7)) 드롭다운 목록에서 Index 1(인덱스 1) 어댑터를 선택합니다.  
제안된 샘플 웰을 사용하면 소프트웨어는 다양성 인덱스 요건에 부합하는 i7 및 i5 인덱스 어댑터를 자동으로 입력합니다. 테스트 중인 샘플 개수와 일치하는 것이 목록에 없으면 추가 웰에 대한 인덱스 어댑터를 반드시 선택해야 합니다. 추가 웰에 대한 인덱스를 선택해야 하거나 권장 인덱스 어댑터 조합을 사용하지 않을 경우에는 인덱스를 선택하기 전에 1페이지의 15페이지의 **베이스 콜링 및 인덱스 다양성**
- 5 Index 2 (i5)(인덱스 2(i5)) 드롭다운 목록에서 Index 2(인덱스 2) 어댑터를 선택합니다.
- 6 Manifest(메니페스트) 드롭다운 목록에서 메니페스트 파일을 선택합니다.  
폴 A의 샘플에 사용되는 메니페스트는 폴 B의 샘플에 사용되는 것과 달라야 합니다.
- 7 라이브러리 프랩에 참고할 수 있도록 플레이트 레이아웃 보기, 인쇄 또는 저장 옵션을 선택하십시오.
  - ▶ 플레이트 레이아웃을 표시하려면  Print(인쇄) 아이콘을 선택합니다. 플레이트 레이아웃을 인쇄하려면 Print(인쇄)를 선택합니다.
  - ▶ 외부 파일로 샘플 정보를 내보내려면 Export(내보내기)를 선택합니다.  
메니페스트 및 샘플 정보가 올바른지 확인합니다. 잘못된 정보는 결과에 영향을 미칠 수 있습니다.
- 8 Save Run(실행 저장)을 선택합니다.

## 샘플 가져오기

- 1 Import Samples(샘플 가져오기)를 선택하고 샘플 정보 파일이 있는 위치를 탐색합니다. 가져올 수 있는 파일 유형은 두 가지입니다.
  - ▶ Create Run(실행 생성) 화면에서 Template(템플릿)을 선택하여 새 플레이트 레이아웃을 만듭니다. 템플릿 파일에는 가져오기에 적합한 올바른 열 제목이 있습니다. 실행 내 샘플에 대한 각 열에 샘플 정보를 입력합니다. 미사용 셀에 있는 예시 정보를 삭제하고 파일을 저장합니다.
  - ▶ Export(내보내기) 기능을 사용하여 Somatic Variant 모듈에서 내보낸 샘플 정보 파일을 사용합니다.
- 2 플레이트 레이아웃을 표시하려면  Print(인쇄) 아이콘을 선택합니다.
- 3 Print(인쇄)를 선택하여 라이브러리 프랩용 Reference로 플레이트 레이아웃을 인쇄합니다.
- 4 **[선택 사항]** 외부 파일로 샘플 정보를 내보내려면 Export(내보내기)를 선택합니다.  
메니페스트 및 샘플 정보가 올바른지 확인합니다. 잘못된 정보는 결과에 영향을 미칠 수 있습니다.
- 5 Save Run(실행 저장)을 선택합니다.

## 실행 편집

시퀀싱 전 실행 정보 편집에 관한 지침은 *NextSeq 550Dx 기기 Reference 가이드(문서 번호 1000000009513)*를 참조하십시오.

## 분석법

Somatic Variant 분석 모듈은 다음 분석 단계를 수행한 후 분석 출력 파일을 Alignment 폴더에 작성합니다.

- ▶ 인덱스 리드 디멀티플렉싱
- ▶ FASTQ 파일 생성
- ▶ Reference에 배열
- ▶ 변이 식별

## 디멀티플렉싱

디멀티플렉싱은 각 인덱스 리드 시퀀스와 실행에 대해 지정된 인덱스 시퀀스를 비교합니다. 이 단계에서 퀄리티 값은 고려되지 않습니다.

인덱스 리드는 아래 단계를 통해 식별됩니다.

- ▶ 샘플은 나열된 실행 순서를 바탕으로 1부터 번호가 지정됩니다.
- ▶ 샘플 번호 0은 샘플에 할당되지 않은 클러스터용으로 예약됩니다.
- ▶ 클러스터는 인덱스 시퀀스가 정확히 일치하거나 인덱스 리드당 최대 1개의 불일치가 있을 경우 샘플에 할당됩니다.

## FASTQ 파일 생성

디멀티플렉싱 후 소프트웨어는 시퀀스를 나타내는 데 사용되는 텍스트 형식인 FASTQ 형식으로 중간 분석 파일을 생성합니다. FASTQ 파일에는 각 샘플에 대한 리드와 연결된 퀄리티 스코어가 있습니다. 필터를 통과하지 못한 클러스터는 제외됩니다.

각 FASTQ 파일에는 샘플 1개에 대한 리드만 있으며, 해당 샘플의 이름이 FASTQ 파일 이름에 포함됩니다. FASTQ 파일은 배열에 대한 일차 입력입니다. 올리고 풀 샘플당 8개, 리드 1에서 4개, 리드 2에서 4개의 FASTQ 파일이 생성되어 샘플당 총 16개의 FASTQ 파일이 생성됩니다.

## 배열

배열 단계에서는 밴디드 스미스-워터맨(Banded Smith-Waterman) 알고리즘이 각 샘플의 클러스터를 메니페스트 파일에 지정된 앰프리콘 시퀀스와 비교해 배열합니다.

밴디드 스미스-워터맨 알고리즘은 세미글로벌 시퀀스 배열을 수행하여 두 시퀀스 간의 유사한 영역을 결정합니다. 스미스-워터맨 알고리즘은 전체 시퀀스를 비교하는 것이 아니라 모든 가능한 길이의 세그먼트를 비교합니다.

각 페어드 엔드 리드는 해당 리드의 관련 프로브 시퀀스에 대한 배열과 관련하여 평가됩니다.

- ▶ 리드 1은 DLSO(Downstream Locus-Specific Oligos: 다운스트림 좌위 특이적인 올리고)의 역방향 보체와 비교하여 평가됩니다.
- ▶ 리드 2는 ULSO(Upstream Locus-Specific Oligos: 업스트림 좌위 특이적인 올리고)와 비교하여 평가됩니다.
- ▶ 리드 시작이 프로브 시퀀스와 일치하고 차이(불일치 또는 리딩 Ins/Del로 인한 이동)가 3개 이하인 경우 리드의 전체 길이가 해당 시퀀스의 앰프리콘 타겟에 배열됩니다.
- ▶ DLSO 및 ULSO 내의 Ins/Del은 해당 assay chemistry에서 관찰되지 않습니다.

앰프리콘 길이에 따라 관심 영역 전체 또는 전체 앰프리콘의 불일치율을 바탕으로 배열 결과에서 배열이 필터링됩니다. 필터링된 배열은 미배열로 배열 파일에 작성되며 변이 콜링에 사용되지 않습니다.

## 변이 콜링

Illumina에서 개발한 Pisces 변이 콜러는 DNA 샘플에서 낮은 빈도로 존재하는 변이를 식별합니다.

Pisces 변이 콜러는 SNV, MNV 및 작은 Ins/Del을 3단계로 식별합니다.

- ▶ Reference 게놈 내 각 위치를 개별적으로 고려합니다.
- ▶ 위치와 중첩되는 배열된 리드에 대해 해당 위치에서 염기를 카운트합니다.
- ▶ Poisson 모델을 사용하여 콜의 퀄리티를 측정하는 변이 스코어를 계산합니다. 퀄리티 스코어가 Q30 미만인 변이는 제외됩니다.

변이는 먼저 각 풀에 대해 개별적으로 콜링됩니다. 이후 각 풀의 변이가 비교되고 하나의 출력 파일로 통합됩니다. 두 풀에 모두 변이가 존재하고 1페이지의 11페이지의 VCF 파일 주석 필터를 통과할 경우 변이는 변이 콜 파일(VCF)에서 PASS로 표시됩니다.

## 실행 및 샘플 데이터 보기

- 1 Local Run Manager 대시보드에서 실행 이름을 클릭합니다.
- 2 Run Overview(실행 개요) 탭에서 시퀀싱 실행 메트릭을 검토합니다.
- 3 **[선택사항]** Copy to Clipboard(클립보드로 복사)  아이콘을 클릭하여 출력 실행 폴더 경로를 복사합니다.
- 4 Sequencing Information(시퀀싱 정보) 탭을 클릭하여 실행 매개변수와 소모품 정보를 검토합니다.
- 5 Samples and Results(샘플 및 결과) 탭을 클릭하여 분석 보고서 위치를 확인합니다.
  - ▶ 분석이 반복되었다면 Select Analysis(분석 선택) 드롭다운 목록을 확장하여 해당 분석을 선택합니다.
- 6 Copy to Clipboard(클립보드로 복사)  아이콘을 선택하여 Analysis 폴더 경로를 복사합니다.

Run Overview(실행 개요) 및 Sequencing Information(시퀀싱 정보) 탭과 분석 requeue 방법에 대한 자세한 내용은 *NextSeq 550Dx 기기 Reference 가이드(문서 번호 1000000009513)*를 참조하십시오.

## 분석 보고서

분석 결과는 Samples(샘플) 및 Results(결과) 탭에 요약되며 Alignment 폴더에 종합 보고서로 요약됩니다. 각 샘플에 대한 보고서는 또한 샘플별 PDF 파일 형식으로도 이용할 수 있습니다.

### 샘플 및 결과 탭 정보

- 1 샘플 보고서를 확인하려면 목록에서 샘플을 클릭합니다.

표 1 실행 및 샘플 정보

열 제목	설명
Run Status(실행 상태)	시퀀싱 실행 통과 또는 실패 여부를 명시합니다.
Total Yield (GB)(총 수율(GB))	시퀀싱 실행 중 콜링된 염기의 개수입니다. 통과 임계값, 통과 또는 실패 상태를 보여줍니다.
% ≥ Q30	시퀀싱 실행에서 퀄리티 스코어 30(Q30) 이상인 리드의 백분율입니다. 통과 임계값, 통과 또는 실패 상태를 보여줍니다.
Sample Name(샘플 이름)	실행이 생성될 때 제공된 샘플 이름입니다.
Total PF Reads(총 PF 리드)	필터를 통과하는 리드의 총 개수입니다.
Read 1% ≥ Q30(리드 1% ≥ Q30)	샘플에 대해 퀄리티 스코어 30(Q30) 이상인 리드 1의 리드 백분율입니다.
Read 2% ≥ Q30(리드 2% ≥ Q30)	샘플에 대해 퀄리티 스코어 30(Q30) 이상인 리드 2의 리드 백분율입니다.
Autosome Call Rate(상염색체 콜 비율)	사전 정의된 신뢰 수준 임계값에 부합하는 상염색체(염색체 1부터 22까지) 전체의 게놈 위치 개수를, 정보를 확인한 상염색체 게놈 위치 총 개수로 나눈 값입니다. 콜 비율은 샘플 기준으로 기술되며 {1 - (불완전 콜이 있는 상염색체 위치 개수를 시퀀싱된 상염색체 위치 총 개수로 나눈 값)}을 계산하여 백분율로 보고합니다.

표 2 샘플 보고서 정보

열 제목	설명
Sample(샘플)	실행이 생성될 때 제공된 샘플 이름입니다.
Report Date(보고서 날짜)	보고서가 생성된 날짜입니다.
샘플 정보	실행이 생성될 때 제공된 샘플 ID, 샘플에서 필터를 통과한 총 리드, 샘플의 퀄리티 스코어 30(Q30) 이상인 리드 백분율, 상염색체 콜 비율입니다.
Amplicon Summary(엠프리콘 요약)	시퀀싱된 엠프리콘 영역 총 개수, 풀 A 및 풀 B의 샘플에 대해 타겟 영역에서 시퀀싱된 엠프리콘의 염기쌍 총 길이, 각 풀에 사용된 메니페스트 파일입니다. 메니페스트 파일은 배열 단계에 사용되는 Reference 게놈 및 타겟 Reference 영역을 지정합니다.
Read Level Statistics(리드 레벨 통계)	풀 A 및 풀 B의 리드 1 및 리드 2에 대해 Reference 내 각 위치를 포함하는 샘플의 리드 개수 및 백분율입니다.
Variants Summary(변이 요약)	퀄리티 결과가 허용 범위에 속하는지 결정하기 위해 제안된 값을 통과한 샘플에 대해 결정된 SNV, Insertion, Deletion의 개수입니다.
Coverage Summary(커버리지 요약)	배열된 염기의 총 개수를 타겟 영역 크기로 나눈 값 및 풀 A 및 풀 B의 샘플에 대해 낮은 커버리지 임계값인 $0.2 * \text{엠프리콘 평균 커버리지}$ 보다 커버리지 값이 큰 엠프리콘 영역의 백분율입니다.
Coverage Plots(커버리지 그래프)	엠프리콘 영역별 커버리지 그래프는 샘플에 대한 엠프리콘 영역 전반의 커버리지를 보여줍니다. 커버리지 값이 커버리지 임계값보다 낮은 영역은 빨간색으로 강조 표시됩니다. 모든 값의 평균은 주황색 선으로 표시됩니다. 풀 A 및 풀 B의 커버리지에 대해 하나의 그래프가 제공됩니다.
Software Versions(소프트웨어 버전)	샘플 시퀀싱 시의 소프트웨어 버전입니다. NextSeq 550Dx 작동 소프트웨어(NOS), Local Run Manager 소프트웨어, RTA 소프트웨어, Somatic Variant 모듈 버전을 포함합니다.

## 분석 출력 파일

Somatic Variant 분석 모듈에 대해 아래 분석 출력 파일이 생성되며, 이러한 파일은 배열 및 변이 콜링에 대한 분석 결과를 제공합니다. 분석 출력 파일은 Alignment 폴더에 있습니다.

파일 이름	설명
디멀티플렉싱(*.txt)	디멀티플렉싱 요약 결과가 들어 있는 중간 파일.
FASTQ(*.fastq.gz)	퀄리티 스코어가 매겨진 베이스 콜이 들어 있는 중간 파일. FASTQ 파일은 배열 단계에 대한 일차 입력입니다.
BAM 형식(*.bam)의 배열 파일	해당 샘플에 대해 배열된 리드가 들어 있습니다.
VCF 형식(*.vcf)의 풀당 변이 콜 파일	정방향 풀 또는 역방향 풀의 각 위치에서 콜링된 변이가 들어 있습니다.
게놈 VCF 형식(*.genome.vcf.gz)의 변이 콜 파일	변이로 콜링되었거나 Reference로 콜링된 각 위치의 유전자형이 들어 있습니다.
VCF 형식(*.vcf.gz)의 일치 변이 콜 파일	두 풀의 각 위치에 콜링된 변이가 들어 있습니다.
AmpliconCoverage_M1.tsv	제공된 각 메니페스트에 대한 각 샘플의 엠프리콘당 커버리지에 대한 정보가 들어 있습니다. M#는 메니페스트 번호를 나타냅니다.

## 디멀티플렉싱 파일 형식

디멀티플렉싱 프로세스는 각 클러스터에 부착된 인덱스 시퀀스를 판독하여 클러스터가 기원한 샘플을 결정합니다. 클러스터와 샘플 번호 간 매핑은 플로우 셀의 각 타일에 대한 디멀티플렉싱(\*.demux) 파일에 작성됩니다.

디멀티플렉싱 파일 이름 형식은 s\_1\_X.demux이며, 여기서 X는 타일 번호입니다.

디멀티플렉싱 파일은 다음과 같은 머리글로 시작합니다.

- ▶ 버전(4바이트 정수), 현재 1
- ▶ 클러스터 카운트(4바이트 정수)

파일 알림은 타일의 각 클러스터에 대한 샘플 번호로 구성됩니다.

디멀티플렉싱 단계가 완료되면 소프트웨어는 DemultiplexSummaryF1L1.txt라는 이름으로 디멀티플렉싱 파일을 생성합니다.

- ▶ 파일 이름에서 F1은 플로우 셀 번호를 나타냅니다.
- ▶ 파일 이름에서 L1은 레인 번호를 나타냅니다.
- ▶ 타일당 행 1개, 샘플당 열 1개(샘플 0 포함)로 구성된 디멀티플렉싱 결과 표.
- ▶ 인덱스 리드에서 가장 흔히 발생하는 시퀀스.

## FASTQ 파일 형식

FASTQ는 리드별 베이스 콜과 퀄리티 값을 포함하고 있는 텍스트 기반 파일 형식입니다. 각 기록에는 4줄이 있습니다.

- ▶ 식별자
- ▶ 시퀀스
- ▶ 플러스 부호(+)
- ▶ ASCII + 33 인코딩 형식의 Phred 퀄리티 스코어

식별자 형식은 다음과 같습니다.

**@기기:실행ID:플로우셀ID:레인:타일:X:Y 리드번호:필터플래그:0:샘플번호**

예:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAA9#:<#<;<<<????#=#
```

## BAM 파일 형식

BAM 파일(\*.bam)은 최대 128Mb의 배열된 시퀀스를 나타내는 데 사용되는 SAM 파일의 압축된 이진 버전입니다. SAM 및 BAM 형식은 [samtools.github.io/hts-specs/SAMv1.pdf](https://samtools.github.io/hts-specs/SAMv1.pdf)에 자세히 설명되어 있습니다.

BAM 파일은 SampleName\_S#.bam과 같은 형식의 파일 이름을 사용하며, 여기에서 #은 샘플 실행 순서에 의해 결정되는 샘플 번호입니다.

BAM 파일에는 머리글 섹션과 배열 섹션이 있습니다.

- ▶ **머리글** - 샘플 이름, 샘플 길이, 배열 방법과 같은 전체 파일에 대한 정보를 포함합니다. 배열 섹션의 배열은 머리글 섹션의 특정 정보와 연결되어 있습니다.
- ▶ **배열** - 리드 이름, 리드 시퀀스, 리드 퀄리티, 배열 정보, 사용자지정 태그를 포함합니다. 리드 이름에는 염색체, 시작 좌표, 배열 퀄리티, 일치 설명자 스트링이 포함됩니다.

배열 섹션에는 각 리드 또는 리드 쌍에 대한 다음 정보가 포함됩니다.

- ▶ **AS:** 페어드 엔드 배열 퀄리티
- ▶ **BC:** 리드와 연결된 디멀티플렉싱된 샘플 ID를 나타내는 바코드 태그.
- ▶ **SM:** 싱글 엔드 배열 퀄리티.
- ▶ **XC:** 일치 설명자 스트링

- ▶ **XN:** 리드와 연결된 엠프리콘 ID를 기록하는 엠프리콘 이름 태그  
BAM 인덱스 파일(\*.bam.bai)은 해당 BAM 파일의 인덱스를 제공합니다.

## VCF 파일 형식

변이 콜 형식(VCF)은 유전체학 커뮤니티에서 개발되었으며 일반적으로 사용되는 파일 형식입니다. 이 파일에는 Reference 게놈의 특정 위치에서 발견된 변이에 대한 정보가 들어 있습니다. VCF 파일의 확장명은 .vcf입니다.

VCF 파일 머리글에는 VCF 파일 형식 버전과 변이 콜러 버전이 포함되어 있으며 파일 알림에 사용되는 주석이 기재되어 있습니다. VCF 머리글에는 Reference 게놈 파일과 BAM 파일도 포함되어 있습니다. 머리글의 마지막 라인에는 데이터 라인의 열 제목이 있습니다. 각 VCF 파일 데이터 라인에는 하나의 변이에 대한 정보가 있습니다.

## VCF 파일 제목

제목	설명
CHROM	Reference 게놈의 염색체입니다. 염색체는 Reference FASTQ 파일 내 순서와 동일한 순서로 표시됩니다.
POS	Reference 염색체 내 변이의 단일 염기 위치입니다. SNP의 경우 이 위치는 변이가 있는 Reference 염기이며 Ins/Del 또는 Deletion의 경우 이 위치는 변이 직전의 Reference 염기입니다.
ID	dbSNP.txt에서 얻은 변이의 rs 번호입니다(해당하는 경우). 이 위치에 rs 번호가 여러 개 있을 경우 목록은 세미콜론으로 구분됩니다. 이 위치에 dbSNP 항목이 없을 경우 누락값 표시자('.')가 사용됩니다.
REF	Reference 유전자형입니다. 예를 들면 단일 T의 Deletion은 Reference TT 및 Alternate T로 나타냅니다. A부터 T까지의 단일 뉴클레오티드 변이는 Reference A 및 Alternate T로 나타냅니다.
ALT	Reference 리드와 다른 대립유전자입니다. 예를 들어 단일 T의 Insertion은 Reference A 및 Alternate AT로 나타냅니다. A부터 T까지의 단일 뉴클레오티드 변이는 Reference A 및 Alternate T로 나타냅니다.
QUAL	변이 콜러에 할당된 Phred-scale 퀄리티 스코어입니다. 스코어가 높을수록 변이 신뢰 수준이 높고 오류 확률이 낮습니다. 퀄리티 스코어 Q는 추정된 오류 확률이 $10^{-Q/10}$ 입니다. 예를 들어 Q30 콜 세트의 오류율은 0.1%입니다. 다수의 변이 콜러가 통계 모델을 바탕으로 퀄리티 스코어를 할당하며, 이는 관찰 오류율에 비해 상대적으로 높습니다.

## VCF 파일 주석

제목	설명
FILTER(필터)	<p>모든 필터를 통과하면, 필터 열에 <b>PASS</b>(통과)가 작성됩니다.</p> <ul style="list-style-type: none"> <li>• <b>LowDP</b> - 각 풀에서 450x 미만의 커버리지 깊이로 부위에 적용됩니다. 정방향 리드와 역방향 리드 모두 적용되는 앰프리콘 위치의 경우 900x 단일 리드 커버리지와 동일합니다.</li> <li>• <b>LowGQ</b> - 유전형 분석 퀄리티(GQ)가 컷오프 미만입니다.</li> <li>• <b>q30</b> - 퀄리티 스코어가 30 미만입니다.</li> <li>• <b>LowVariantFreq</b> - 변이 빈도가 해당 임계값 미만입니다.</li> <li>• <b>PB</b> - 프로브 풀 편향입니다. 변이가 발견되지 않았거나 1~2개의 프로브 풀에서 낮은 빈도로 발견되었습니다.</li> <li>• <b>R3x6</b> - 변이 콜에 인접한 반복서열(1~3bp 길이)의 개수가 6개 이상입니다.</li> <li>• <b>SB</b> - 가닥 편향이 해당 임계값보다 큼니다.</li> </ul>
INFO(정보)	<p>INFO(정보) 열에 포함될 수 있는 항목은 다음과 같습니다.</p> <ul style="list-style-type: none"> <li>• <b>AC</b> - 각 ALT 대립유전자에 대한 유전자형 내 대립유전자 카운트이며 열거된 순서와 동일합니다.</li> <li>• <b>AF</b> - 각 ALT 대립유전자에 대한 대립유전자 빈도이며 열거된 순서와 동일합니다.</li> <li>• <b>AN</b> - 콜링된 유전자형 내 대립유전자의 총 개수입니다.</li> <li>• <b>CD</b> - 1개 이상의 RefGene 항목 코딩 영역 내에 SNP가 발생했음을 나타내는 플래그입니다.</li> <li>• <b>DP</b> - 깊이(위치에 배열되고 변이 콜링에 사용된 베이스 콜의 개수)입니다.</li> <li>• <b>Exon</b> - 침표로 구분된 RefGene 엑손 영역 리드 목록입니다.</li> <li>• <b>FC</b> - 기능적 결과입니다.</li> <li>• <b>GI</b> - 침표로 구분된 RefGene 유전자 ID 리드 목록입니다.</li> <li>• <b>QD</b> - 깊이별 변이 신뢰 수준/퀄리티입니다.</li> <li>• <b>TI</b> - 침표로 구분된 RefGene 전사 ID 리드 목록입니다.</li> </ul>
FORMAT(형식)	<p>형식 열에는 콜론으로 구분된 필드가 나열됩니다. 예를 들면 GT:GQ와 같습니다. 제공되는 필드 목록은 사용된 변이 콜러에 따라 다릅니다. 사용 가능한 필드는 다음과 같습니다.</p> <ul style="list-style-type: none"> <li>• <b>AD</b> - X, Y의 형태로 입력하며, 이때 X는 Reference 콜의 개수이고 Y는 Alternate 콜의 개수입니다.</li> <li>• <b>DP</b> - 대략적인 리드 깊이로, MQ=255이거나 메이트가 불량인 리드는 필터링됩니다.</li> <li>• <b>GQ</b> - 유전자형 퀄리티입니다.</li> <li>• <b>GQX</b> - 유전자형 퀄리티입니다. GQX는 GQ 값 및 QUAL 열의 최소값입니다. 일반적으로 두 값은 유사합니다. 단, GQX는 최소값을 취하기 때문에 유전자형 퀄리티 측정이 더 보수적이라고 할 수 있습니다.</li> <li>• <b>GT</b> - 유전자형입니다. 예를 들어, 0은 Reference 염기를, 1은 ALT 열의 첫 번째 항목을 나타냅니다. 슬래시(/)는 이용 가능한 위상화 정보가 없음을 의미합니다.</li> <li>• <b>NL</b> - 잡음 수준입니다. 이 위치에서의 베이스 콜링 잡음 추정치입니다.</li> <li>• <b>PB</b> - 프로브 풀 편향입니다. 값이 0에 가까울수록 한 프로브 풀에 대한 편향이 커지고 변이 콜의 신뢰 수준이 낮아집니다.</li> <li>• <b>SB</b> - 이 위치에서의 가닥 편향입니다. 음수 값이 클수록 편향이 적고, 값이 0에 가까울수록 편향이 커집니다.</li> <li>• <b>VF</b> - 변이 빈도입니다. Alternate 대립유전자를 뒷받침하는 리드 백분율입니다.</li> </ul>
SAMPLE(샘플)	샘플 열에는 FORMAT(형식) 열에 지정된 값이 제공됩니다.

## 게놈 VCF 파일

게놈 VCF(gVCF) 파일은 일련의 규칙에 따라, 합리적인 압축 형식으로 게놈 내 모든 부위를 나타내는 VCF v4.1 파일입니다. gVCF(\*.genome.vcf.gz) 파일에는 각 샘플의 단일 파일에 있는 관심 영역 내 모든 부위가 포함됩니다.

모든 필터를 통과하지 못하는 위치에서는 콜 없음(no-call)이 gVCF 파일에 표시됩니다. 유전자형(GT) 태그 **./.**은(는) 콜 없음을 나타냅니다.

자세한 내용은 [sites.google.com/site/gvcftools/home/about-gvcf](https://sites.google.com/site/gvcftools/home/about-gvcf)를 참조하십시오.

## 폴별 및 일치 VCF 파일

Somatic Variant 워크플로우는 변이 콜 파일 2세트를 생성합니다.

- ▶ **폴별 VCF 파일** - 정방향 폴 또는 역방향 폴에서 콜링된 변이를 포함합니다. 폴별 파일은 VariantCallingLogs 폴더에 작성됩니다.
  - ▶ **일치 VCF 파일** - 두 폴에서 콜링된 변이를 포함합니다. 일치 파일은 Alignment 폴더에 작성됩니다.
- 폴별 및 일치 VCF 파일에는 VCF(\*.vcf) 파일과 gVCF(\*.genome.vcf) 파일이 모두 포함되며 아래의 명명 규칙을 사용합니다. 여기서 S#는 샘플 실행 순서를 나타냅니다.
- ▶ **모든 부위 보고서** - 샘플이름\_S#.genome.vcf
  - ▶ **변이 보고서** - 샘플이름\_S#.vcf

소프트웨어는 폴별 VCF 파일을 비교하고 각 위치에서 데이터를 조합하여 샘플에 대한 일치 VCF 파일을 생성합니다.

각 폴의 변이 콜은 아래 기준을 사용하여 일치 VCF 파일로 병합됩니다.

기준	결과
각 폴의 Reference 콜	Reference 콜
한 폴의 Reference 콜 및 다른 폴의 변이 콜	필터링된 변이 콜
각 폴에서 유사한 빈도로 변이 콜을 매칭	변이 콜
각 폴에서 매우 다른 빈도로 변이 콜을 매칭	필터링된 변이 콜
각 폴에서 매칭되지 않은 변이 콜	필터링된 변이 콜

각 폴의 메트릭은 다음 값을 사용하여 병합됩니다.

메트릭	값
깊이	두 폴 깊이의 합
변이 빈도	총 변이 카운트를 총 커버리지 깊이로 나눈 값
Q-Score	두 폴의 최소값

## 엠프리콘 커버리지 파일

각 메니페스트 파일에 대해 엠프리콘 커버리지 파일이 생성됩니다. 파일 이름의 M#는 메니페스트 번호를 나타냅니다.

각 파일에는 해당 메니페스트와 연결된 샘플 ID가 포함된 머리글 행이 있습니다. 파일에는 다음 정보가 있습니다.

- ▶ 메니페스트에 기재된 타겟 ID.
- ▶ 필터를 통과하는 리드의 커버리지 깊이.

## 보충 출력 파일

다음 출력 파일은 보충 정보를 제공하거나 실행 결과 및 분석 오류를 요약합니다. 이러한 파일들은 분석 결과를 평가하는 데 필요하지는 않지만 문제 해결 목적으로 사용할 수 있습니다. 모든 파일은 달리 명시되지 않는 한 Alignment 폴더에 있습니다.

파일 이름	설명
AnalysisLog.txt	현재 실행 폴더를 분석하는 동안 발생한 모든 단계를 설명하는 처리 로그입니다. 이 파일에는 오류 메시지가 포함되지 않습니다. Alignment 폴더에 있습니다.
AnalysisError.txt	분석 중 발생한 모든 오류를 열거하는 처리 로그입니다. 발생한 오류가 없는 경우에는 이 파일에 아무 내용도 기록되지 않습니다. Alignment 폴더에 있습니다.
DemultiplexSummaryF1L1#.txt	타일당 행이 1개고 샘플당 열이 1개인 표에 디멀티플렉싱 결과를 보고합니다. 여기서 #는 플로우 셀의 레인 1, 2, 3, 4를 나타냅니다. Alignment 폴더에 있습니다.
AmpliconRunStatistics.xml	실행별 요약 통계가 있습니다. Alignment 폴더에 있습니다.

## Analysis 폴더

분석 폴더에는 Local Run Manager 소프트웨어에서 생성된 파일이 보관됩니다.

출력 폴더와 분석 폴더 간 관계는 다음과 같이 요약됩니다.

- ▶ 시퀀싱 중에 실시간 분석(RTA)이 이미지 분석, 베이스 콜링, 퀄리티 스코어 평가 중 생성된 파일을 출력 폴더에 채웁니다.
- ▶ RTA는 파일을 실시간으로 분석 폴더로 복사합니다. RTA가 각 Cycle에 대한 각 염기에 퀄리티 스코어를 할당하고 나면 소프트웨어가 두 폴더에 RTAComplete.txt 파일을 작성합니다.
- ▶ RTAComplete.txt 파일이 존재하면 분석이 시작됩니다.
- ▶ 분석이 계속되는 중에 Local Run Manager는 분석 폴더에 출력 파일을 작성한 후 해당 파일을 다시 출력 폴더로 복사합니다.

## Alignment 폴더

분석이 requeue될 때마다 Local Run Manager는 **Alignment\_N**이라는 이름의 Alignment 폴더를 생성하며, 여기서 N은 일련 번호입니다.

## 폴더 구조

**Alignment** - \*.bam, \*.vcf, FASTQ 파일 및 분석 모듈에 특정한 파일이 있습니다.

**Date and Time Stamp** - YYYYMMDD\_HHMMSS와 같은 분석의 날짜\_시간 스탬프

- [-] AnalysisError.txt
- [-] AnalysisLog.txt
- [-] aggregate.report.html
- [-] aggregate.report.pdf
- [-] aggregate.summary.csv
- [-] AmpliconCoverage\_M#.tsv
- [-] AmpliconRunStatistics.xml
- [-] Sample1.genome.vcf.gz
- [-] Sample1.coverage.csv
- [-] Sample1.report.pdf
- [-] Sample1.summary.csv
- [-] Sample1.vcf.gz
- [-] Sample1.bam
- FASTQ**
  - Sample1**
    - [-] Sample1\_L001\_R1\_001\_fastq.gz
  - Stats**
    - [-] DemuxSummaryF1L1.txt
    - [-] FastqSummaryF1L1.txt

**Data**

**Intensities**

**BaseCalls**

**L001** - Contains \*.bcl 파일.

**L001** - Contains \*.locs 파일.

**RTA Logs** - RTA 소프트웨어 분석의 로그 파일이 있습니다.

**InterOp** - 시퀀싱 실행 메트릭을 보고하는 데 사용되는 이진 파일이 있습니다.

**Logs** - 시퀀싱 중 수행되는 단계를 설명하는 로그 파일이 있습니다.

- [-] RTAComplete.txt
- [-] RunInfo.xml
- [-] RunParameters.xml

## 베이스 콜링 및 인덱스 다양성

NextSeq 550Dx 기기에서 샘플을 시퀀싱할 때 베이스 콜링은 해당 타일 또는 플로우 셀의 이미지 생성 영역의 모든 클러스터에 대해 특정 Cycle에서 염기(A, C, G 또는 T)를 결정합니다. NextSeq 550Dx 기기에서는 이미지 2개(빨간색 채널에서 1개, 녹색 채널에서 1개)만으로 DNA 염기 4개에 대한 데이터를 인코딩하는 2채널 시퀀싱을 사용합니다.

베이스 콜링 인덱스 리드에 대한 처리 방법은 다른 리드 중의 베이스 콜링과 다릅니다.

인덱스 리드는 처음 2개의 Cycle 중 1개에 G 이외의 염기를 1개 이상 사용하여 시작해야 합니다. 인덱스 리드가 G 베이스 콜 2개로 시작되면 신호 인텐시티가 생성되지 않습니다. 디멀티플렉싱 성능을 보장하려면 처음 2개의 Cycle 중 1개에 신호가 있어야 합니다.

실행 생성 중 인덱스를 선택하는 경우 인덱스가 다양성 요건에 부합하지 않으면 다양성 낮음 경고가 표시됩니다. 다양성 낮음 경고가 발생하지 않도록 하려면 모든 Cycle에서 두 채널 모두에 신호를 제공하는 인덱스 시퀀스를 선택합니다.

- ▶ 빨간색 채널 - A 또는 C
- ▶ 녹색 채널 - A 또는 T

이 베이스 콜링 프로세스는 로우플렉스 샘플 분석 시 정확도를 보장합니다. 인덱스의 시퀀스에 대한 자세한 내용은 *TruSeq Custom Amplicon Kit Dx*(문서 번호 1000000029772)의 패키지 설명서를 참조하십시오.

Local Run Manager에서 실행을 생성할 때 검사할 샘플 개수를 선택하게 됩니다. 인덱스 다양성 요건에 부합하는 제안된 인덱스 조합은 소프트웨어가 자동으로 입력합니다. 제안된 인덱스 조합을 반드시 사용해야 하는 것은 아니지만 사용이 권장됩니다.

## 개정 기록

문서	날짜	변경사항 설명
문서 번호 1000000030330 v04	2021 년 8 월	업데이트된 EU 수권 담당자 주소입니다.
문서 번호 1000000030330 v03	2020 년 4 월	업데이트된 EU 수권 담당자 주소입니다. 업데이트된 호주 의뢰사 주소입니다.
문서 번호 1000000030330 v02	2019년 1월	v2.5 시약 키트 관련 정보 추가.
문서 번호 1000000030330 v01	2018년 8월	규제 표시 업데이트.
문서 번호 1000000030330 v00	2017년 11월	초판.

## 기술 지원

기술 지원을 받으려면 Illumina 기술 지원 부서에 문의하십시오.

웹사이트: [www.illumina.com](http://www.illumina.com)  
이메일: [techsupport@illumina.com](mailto:techsupport@illumina.com)

### Illumina 고객 지원 전화 번호

지역	무료 통화	지역
북미	+1-800-809-4566	
네덜란드	+31-8000222493	+31-207132960
노르웨이	+47-800-16836	+47-21939693
뉴질랜드	0800-451-650	
대만	00806651752	
덴마크	+45-80820183	+45-89871156
독일	+49-8001014940	+49-8938035677
벨기에	+32-80077160	+32-34002973
스웨덴	+46-850619671	+46-200883979
스위스	+41-565800000	+41-800200442
스페인	+34-911899417	+34-800300143
싱가포르	+1-800-579-2745	
아일랜드	+353-1800936608	+353 016950506
영국	+44-8000126019	+44-2073057197
오스트리아	+43-800006249	+43-19286540
이탈리아	+39-800985513	+39-236003759
일본	0800-111-5011	
중국	400-066-5835	
프랑스	+33-805102193	+33-170770446
핀란드	+358-800918363	+358-974790110
호주	+1-800-775-688	
홍콩	800960230	
기타 국가	+44-1799-534000	

SDS(안전보건자료) - Illumina 웹사이트([support.illumina.com/sds.html](http://support.illumina.com/sds.html))에서 확인할 수 있습니다.

제품 설명서 - Illumina 웹사이트에서 PDF로 다운로드할 수 있습니다. [support.illumina.com](http://support.illumina.com)에서 제품을 선택한 다음, Documentation & Literature(설명서 및 문헌)를 선택하십시오.



Illumina  
5200 Illumina Way  
San Diego, California 92122 U.S.A.  
+ 1-800-809.ILMN(4566)  
+ 1-858-202-4566(북미 외)  
techsupport@illumina.com  
www.illumina.com



Illumina Netherlands B.V.  
Steenoven 19  
5626 DK Eindhoven  
The Netherlands

**호주 의뢰사**

Illumina Australia Pty Ltd  
Nursing Association Building  
Level 3, 535 Elizabeth Street  
Melbourne, VIC 3000  
Australia

체외 진단용

© 2021 Illumina, Inc. All rights reserved.

