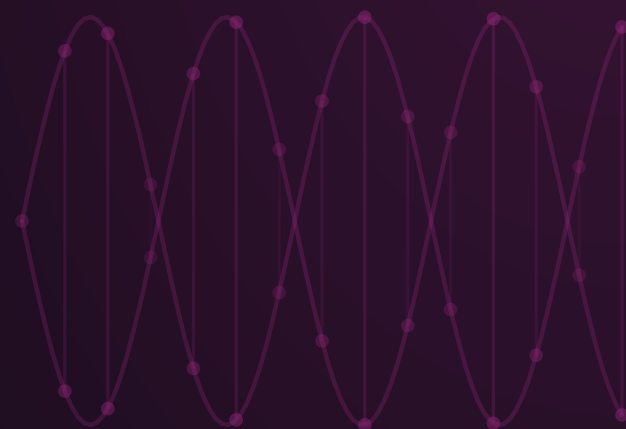


레슨: 유전체 파일 읽는 법, 한 줄씩



생명과학

만 13세+

입문

13세 이상을 위한, 바로 쓸 수 있는 입문용 유전학 레슨입니다. 학교·칼리지·대학교 어디서나 무료로 쓸 수 있습니다. 유전체 파일의 네 칸이 각각 무엇인지, 마지막 칸에 글자가 왜 두 개인지, 어느 판의 지도로 댄 것인지 모르면 위치 숫자가 왜 아무 뜻이 없는지 — 그리고 이런 파일이 결코 알려줄 수 없는 한 가지. 학생의 데이터가 아니라 우리 예제 파일을 씁니다.

자유롭게 읽고, 쓰고, 고쳐 쓰실 수 있습니다. 계정도, 구매도, 수집되는 정보도 없습니다.

전체 레슨과 그림, 갱신본: <https://www.mygenelog.com/ko/updates/lesson-reading-a-genome-file>

이 자료는, 바로 쓸 수 있는 유전학 레슨입니다. 학교·칼리지·대학교 어디서나 자유롭게 쓰고 고칠 수 있습니다. 우리에게서 살 것도, 계정도, 누구에게서 수집하는 데이터도 없습니다.

- 수준: 입문용. 약 13세부터 쓸 수 있습니다. 사전 유전학 지식이 필요 없습니다 — 나오는 용어를 여기서 모두 설명합니다. 설치할 것이 없고, 수학은 나눗셈 하나뿐입니다.
- 시간: 40-50분. 읽기 활동에 20분쯤 걸립니다.
- 준비물: 학생당 또는 두 명당 브라우저 하나, 그리고 우리 예제 파일: `example-genotypes.txt` (15줄, 1 KB). 인쇄해도 되고 아무 텍스트 편집기로 열어도 됩니다. 유전체도, 검체도, 어떤 검사도 필요 없습니다.
- 다루는 내용: 유전체 파일의 네 칸이 각각 무엇인지, 마지막 칸에 글자가 왜 두 개인지, 어느 지도로 댄 것인지 모르면 위치 숫자가 왜 아무 뜻이 없는지, 그리고 이런 파일이 알려줄 수 없는 가장 중요한 한 가지.

전체를 다섯 줄로

누군가 통에 침을 뱉어 보내면, 나중에 "원시 데이터" 또는 "내 데이터 내려받기" 같은 메뉴 아래로 돌아오는 것은 평범한 텍스트 파일입니다. 그림도 아니고 보고서도 아닙니다. 텍스트 파일, 그것도 꽤 심심한 파일입니다.

- 각 줄은 실험실이 들여다본 유전체 상의 자리 하나입니다.
- 각 줄에는 네 칸이 있습니다: 그 자리의 이름, 몇 번 염색체인지, 그 염색체에서 얼마나 들어간 자리인지, 그리고 거기서 발견된 글자.
- 마지막 칸에는 글자가 두 개 있습니다. 사람은 모든 염색체를 두 벌씩 — 부모에게서 한 벌씩 — 가지고 있기 때문입니다.
- 파일에는 의미가 들어 있지 않습니다. "유당 내성"이나 "위험 높음" 같은 말은 어디에도 없습니다. 좌표와

글자뿐입니다. 의미는 전부 그 자리를 다른 곳에서 찾아봐야 나옵니다 — 이 사이트가 있는 이유가 그것입니다.

- 그리고 이 파일은 대부분이 침묵입니다. 약 30억 개 자리 가운데 수십만 자리만 담고 있으며, 그 침묵은 정보가 아닙니다.

레슨 전체가 이것입니다. 아래는 모두 저 다섯 줄 가운데 하나를 좀 더 들여다본 것이거나, 학생들이 그것으로 실제로 해보는 일입니다.

시작하기 전에 — 단 하나의 규칙. 누구도 자기 유전체 파일을 이 레슨에 가져오지 않으며, 검사를 받아본 적이 있는지 묻지 않습니다. 교실 안 누구에 대한 결과도 만들지 않고, 누가 무엇을 가졌는지 적은 명단도 만들지 않습니다. 우리가 주는 파일은 누구도 아닙니다. 이 사이트의 공개 목록에서 생성한 것으로, 만드는 데 어떤 유전체도 읽지 않았고, 거기서 어떤 사람에 대한 것도 복원할 수 없습니다. 학생이 자기 DNA에 대해 물으면, 정직한 답은 교실은 그것을 알아볼 자리가 아니고 의사나 유전상담사가 맞는 자리라는 것입니다.

한 줄을 뜯어보기

파일에 실제로 있는 그대로의 한 줄입니다. 칸은 탭 문자로 나뉘어 있어서, 편집기에 따라 간격이 어색해 보이기도 하고 가지런해 보이기도 합니다.

파일의 한 줄을 크게 본 것

1. rsid	2. chromosome	3. position	4. genotype
rs1815739	11	66560624	CT
이 자리에 붙은 영구적인 이름	몇 번 염색체인지 1-22, X, Y, MT	염색체 시작점에서 몇 글자째인지	

이 줄이 말해주지 않는 것

- 무슨 뜻인지
- 그 판독이 맞는지
- 다른 자리에 대한 그 무엇도

마지막 칸에 글자가 두 개인 이유

C

T

사람은 모든 염색체를 두 벌씩
가지므로, 모든 자리에 글자가
하나가 아니라 둘 있습니다

모식도. 예제 파일의 한 줄을, 탭으로 나뉜 네 칸으로 뜯어놓은 것입니다. 마지막 칸의 두 글자는 두 벌 — 부모에게서 한 벌씩 물려받은 것 — 입니다. 어느 글자가 어느 쪽 부모에게서 왔는지는 파일에 기록되지 않으며, 이 줄의 어디에도 그것이 무슨 뜻인지는 적혀 있지 않습니다.

칸	예	무엇인가
rsid	rs1815739	자리 하나에 붙은 영구적인 목록 번호입니다. 지구 위 누구나 같은 자리를 가리키도록 부여합니다. 자리의 이름이며, 거기서 누가 어떤 글자를 가졌는지는 결코 뜻하지 않습니다.

칸	예	무엇인가
chromosome	11	23쌍 가운데 어느 것에 그 자리가 있는지. 1부터 22까지, 그다음 X, Y, 그리고 미토콘드리아 안의 작고 별개인 DNA 고리를 뜻하는 MT입니다.
position	66560624	그 염색체에서 몇 글자째에 그 자리가 있는지 — 여기서는 약 6천6백만 글자쯤 들어간 곳입니다. 이 숫자가 함정을 가진 숫자이며, 함정에는 아래에 따로 절을 두었습니다.
genotype	CT	거기서 실제로 발견된 글자입니다. 한쪽 부모에게서 온 벌에 하나, 다른 쪽 부모에게서 온 벌에 하나. CT와 TC는 같은 뜻입니다. 어느 쪽에서 왔는지를 파일이 기록하지 않기 때문입니다.

같은 글자 둘 — CC, GG — 인 경우를 그 자리에서 동형접합이라고 합니다. 서로 다른 글자 하나씩 — CT — 인 경우는 이형접합입니다. 이 레슨에 나오는 전문 용어는 이 둘뿐이고, 아래 활동은 사실상 그 둘을 손에 잡히게 만들기 위한 구실입니다.

해봅시다: 파일 읽기

여기 나오는 모든 것은 예제 파일과 이 사이트의 무료 페이지만 씁니다. 제출하는 것도, 올리는 것도 없고, 계정도 필요 없습니다.

1. 파일을 열기 전에 먼저 예측

학생들에게 이것만 알려줍니다: 파일은 15줄이고, 각 줄의 마지막 칸에는 글자가 두 개 있습니다. 그런 다음, 아무도 아무것도 열기 전에 학급이 합의해 칠판에 적을 숫자 하나를 받습니다:

15줄 가운데 같은 글자 두 개가 나오는 줄은 몇 줄일까요?

대부분의 학급은 7이나 8 근처를 답합니다. 추론은 좋고 답은 틀렸는데, 이것이 쓸모 있는 조합입니다. 예측을 적어 두십시오. 먼저 못을 박아야만 작동합니다.

2. 세어 보기

파일을 열고 세어 봅니다. #로 시작하는 줄은 주석이며 데이터가 아닙니다 — 짚고 넘어갈 만한 관행입니다. 학생이 앞으로 열어볼 거의 모든 과학 데이터 파일에서 같은 관행을 만나게 되기 때문입니다.

마지막 칸이 보여주는 것	부르는 이름	줄 수
같은 글자 둘, 예: CC	동형접합	10
서로 다른 글자 하나씩, 예: CT	이형접합	5
합계	15	

같은 글자 둘이 흔한 쪽이고, 이유는 이렇습니다. 대부분의 자리에서 가능한 두 글자는 똑같이 흔하지 않습니다 — 하나는 전체 벌의 90퍼센트에서 나오고 다른 하나는 10퍼센트에서 나올 수 있습니다. 그런

집단에서 두 벌을 무작위로 뽑으면 보통 흔한 쪽 둘을 얻게 됩니다. 이형접합은 두 글자의 빈도가 서로 가까울 때 나오며, 그쪽이 더 드문 상황입니다. 그러니 이 칸의 모양은 한 사람에 대해서만이 아니라 집단 전체에 대해 무언가를 말해주고 있는 셈입니다.

3. 세 줄을 찾아보기

두 명씩 짝을 지어 아무 줄이나 셋을 고릅니다. 각 줄마다 이 사이트에서 rsID를 검색해 세 가지를 적습니다:

1. 그 자리가 어느 유전자 안에 있는지, 있다면.
2. 페이지가 바로 그 유전형에 대해 무엇과 연관된다고 말하는지 — 그 자리 일반이 아니라, 파일에 적힌 두 글자에 대해.
3. 그것에 대해 페이지가 어느 논문을 인용하는지.

15개 rsID에는 모두 여기 페이지가 있고, 파일의 모든 유전형은 그 페이지들이 실제로 설명하는 것이라, 어느 짝도 막다른 길에 부딪히지 않습니다. 셋 가운데 세 번째가 가장 중요합니다. 유전형에 대한 주장은 그 뒤에 있는 연구만큼만 값어치가 있으며, 학생들은 무엇을 팔 수 있는 나이가 되기 전에 그것이 어느 연구였는지 먼저 묻는 습관을 들여야 합니다.

4. 대조: 파일에 없는 줄을 찾아보기

이제 모든 짝에게 같은 rsID를 하나 줍니다. 파일에 없는 것입니다: rs4244285. 여기 페이지는 있습니다. CYP2C19라는 유전자 안에 있고, 심근경색이나 스텐트 시술 뒤 혈전을 막으려고 주는 약인 클로피도그렐을 몸이 얼마나 잘 활성화하는지를 바꿉니다. 신기한 이야깃거리가 아니라, 심장내과 의사가 알고 싶어 하는 종류입니다.

학급에 묻습니다: 우리 파일은 이것에 대해 뭐라고 말합니까?

아무 말도 하지 않습니다. 줄이 없습니다. 그리고 이제 이 레슨 전체가 다다르려고 만들어진 질문입니다. 그 침묵은 그 자리에 특이한 것이 없다는 뜻일까요?

아닙니다. 그럴 수가 없습니다. 이런 파일에는 실험실이 들여다보기로 고른 자리들이 들어 있습니다 — 미리 정해진 고정 목록이고, 모든 고객에게 똑같습니다. 애초에 목록에 없던 자리는 그 사람이 거기서 무엇을 가졌든 줄을 만들지 않습니다. 파일에 없다는 것은 그 사람에 대한 사실이 아니라 그 검사에 대한 사실입니다.

이것이 이런 파일에서 가장 많이 오독되는 성질이고, 양쪽 방향으로 다 오독됩니다. "아무것도 안 나왔으니 나는 괜찮다"는 위의 이유로 틀렸고, "있다고 나왔으니 나는 있다"도 틀렸습니다 — 그 이유는 다음다음 절에 있습니다.

셋째 칸의 함정

위치 숫자는 그 줄에서 가장 단단해 보이는 것입니다. 실은 가장 무릅니다.

위치는 염색체의 한쪽 끝에서부터 센 값이며, 따라서 어느 판(版)의 염색체를 따라 세었느냐에 전적으로 달려 있습니다. 참조 인간 유전체는 빈틈이 메워지고 오류가 고쳐지면서 주기적으로 다시 만들어지고, 그때마다 번호가 밀립니다. 지금 두 판이 함께 널리 쓰이고 있습니다. 2009년에 나온 GRCh37과 2013년에 나온 GRCh38입니다.

rsID는 두 판에서 그대로입니다. 숫자는 그렇지 않습니다.

같은 자리, 두 개의 지도 판



232,529 글자 차이 — 그리고 그 차이는 자리마다 크기도 방향도 다릅니다

모식도. 참조 유전체의 두 판에서 본 같은 자리입니다. 좌표는 2026년 9월 9일 Ensembl에서 읽었습니다. rsID는 설계상 고정이고, 숫자는 그렇지 않습니다.

그리고 그 밀림은 상수가 아니라서 계산으로 바로잡을 수 없습니다. 우리 파일에 있는 자리 셋을 두 판 모두에서 Ensembl로 읽은 값입니다:

rsID	GRCh37 위치	GRCh38 위치	차이
rs1815739	chr11 66,328,095	chr11 66,560,624	GRCh38이 232,529만큼 큼
rs4988235	chr2 136,608,646	chr2 135,851,076	GRCh38이 757,570만큼 작음
rs6025	chr1 169,519,049	chr1 169,549,811	GRCh38이 30,762만큼 큼

크기도 다르고, 방향조차 같지 않습니다. 한 판을 다른 판으로 바꿔주는 산술은 없습니다. 유일하게 옳은 방법은 원하는 판에서 그 자리를 다시 찾아보는 것입니다.

가르쳐야 할 습관. 유전체 좌표가 든 파일을 받으면 가장 먼저 할 일은 어느 판을 기준으로 켜 것인지 알아내는 것이고, 볼 곳은 머리말 — 맨 위의 주석 줄들입니다. 우리 예제 파일은 머리말에 그렇게 적어 두었습니다: "Positions are GRCh38, read from Ensembl rather than typed." 받은 파일에 그 말이 없다면, 알아내기 전까지 그 위치 숫자는 무엇과도 안전하게 비교할 수 없습니다. 초보자용 주의사항이 아닙니다. 판이 어긋나는 일은 전문가의 작업 에서도 지금 살아 있는 오류 원인입니다.

이 파일에 들어 있지 않은 것

이 나뭇섬은 학급이 소리 내어 해보게 하십시오. 이런 검사에 돈을 내본 사람일수록 답에 놀랍니다.

사람의 유전체는 약 30억 글자 길이입니다. 소비자용 유전형 파일에는 수십만 줄이 들어 있습니다. 하나를 다른 하나로 나눠 봅니다. 이 파일이 담는 것은 유전체 안 자리의 대략 1퍼센트의 100분의 1 수준이며, 그 자리들은 하나하나 실험실이 미리 골라둔 것입니다. 사람들이 흔히 다른 자리라고 이미 알려져 있었기

때문입니다.

여기서 세 가지가 따라 나오며, 이 제품들의 광고를 만나게 될 교실에는 분명하게 말해줄 가치가 있습니다:

- 드문 것은 대개 보이지 않습니다. 목록에 오른 자리는 흔한 자리들입니다. 의학에서 가장 중요한 변화 가운데 상당수는 드문 것이고, 흔한 자리를 보도록 만든 검사는 그런 것을 그대로 지나칩니다.
- 목록은 누가 연구되었는지를 반영합니다. 고른 자리는 앞선 연구가 찾아낸 자리인데, 그 연구는 유럽계 조상을 가진 사람들에게 치우쳐 수행되었습니다. 다른 집단에서 더 중요한 자리는 빠져 있을 가능성이 큼 — 존재하지 않아서가 아니라, 찾아본 연구가 적어 서입니다.
- 판독 자체가 틀릴 수 있습니다. 이런 칩은 실수를 하고, 그 실수는 하필 아픈 곳에 몰립니다. 드문 변이 — 걱정하는 독자가 가장 찾아보려 하는 바로 그것입니다.

마지막 항목은 측정된 적이 있습니다. 미국의 한 임상 검사실이, 사람들이 소비자용 원시 데이터에서 찾아낸 변이를 확인해 달라며 보내온 환자 검체 49건을 검토했고, 그 변이의 40퍼센트가 위양성이었다고 보고했습니다 — 파일에 적힌 글자가 그 사람에게 있는 글자가 아니었던 것입니다. 그 밖에 몇 건은 실재했지만 "위험 증가"로 표시되어 있었고, 그 검사실과 다른 여러 곳은 그것을 양성(benign)으로 분류했습니다 (Tandy-Connor S 외, Genetics in Medicine 2018, PMID:29565420).

그러니 원시 데이터 파일을 정직하게 요약하면 이렇습니다. 당신 대신 골라진, 흔한 자리들의 불완전한 목록이고, 대체로 맞지만 늘 맞지는 않은 장비로 켜 것이며, 해석은 전혀 들어 있지 않습니다. 정말로 흥미롭습니다. 진단은 아니며, 그 안의 어떤 줄도 임상 검사실이 먼저 확인해 주지 않은 채 의학적으로 따라 행동해서는 안 됩니다.

아직 열려 있는 논쟁

그렇다면 회사가 고객에게 원시 데이터 파일을 아예 넘겨줘야 할까요?

한쪽 입장: 넘겨주지 말아야 하거나, 적어도 임상의를 사이에 두지 않고는 안 됩니다. 이 파일은 일상적으로 오독되고, 제3자 서비스는 말아야 할 의무 없이 돈을 받고 해석해 주며, 위의 위양성 비율은 실제 사람들이 자기 유전체에 있지도 않은 글자 때문에 겁을 먹어 왔다는 뜻입니다.

다른 쪽: 그 데이터는 그 사람에 대한 것이고, 그 사람이 돈을 냈으며, 주지 않는 것은 온정적 간섭입니다. 사람들이 읽고 따져 묻고 의사에게 가져갈 수 있는 파일이, 믿고 받아들이는 수밖에 없는 잘 만든 보고서보다 낫습니다 — 그리고 오독을 고치는 방법은 읽는 법을 가르치는 것이며, 이 한 시간이 바로 그것입니다.

둘 다 진지하게 주장되고 있고, 어느 쪽도 이기지 않았습니다. 학생들은 이 분야가 가진 것과 같은 근거를 그대로 질 수 있으며, 위의 숫자들을 결론이 아니라 낱것으로 준 이유가 그것입니다.

질문

몸풀기

1. 유전체 파일의 네 칸 이름을 대고, 각각 무엇이 들어 있는지 한 문장으로 말해 보세요.
2. 마지막 칸에는 왜 글자가 하나가 아니라 둘일까요?
3. 어떤 줄은 GA이고 다른 줄은 AG입니다. 둘 사이에 차이가 있습니까? 설명해 보세요.

4. 파일 맨 위의 #로 시작하는 줄들은 무엇을 합니까? 파일을 읽는 프로그램이 왜 그것을 알아야 합니까?
5. 파일에는 동형접합 줄이 더 많습니까, 이형접합 줄이 더 많습니까? 왜 그쪽입니까?

본문제

1. rs4244285는 이 사이트에 페이지가 있고 우리 파일에는 줄이 없습니다. 그것이 왜 그 사람에게 거기 특이한 것이 없다는 뜻이 아닌지, 친구에게 설명하는 한 문장을 써 보세요.
2. rs1815739는 참조 유전체의 한 판에서 66,328,095이고 다른 판에서 66,560,624입니다. 한 자리가 어떻게 두 개의 주소를 가질 수 있는지 설명하고, 둘 중 어느 숫자가 "옳은지" 말해 보세요.
3. 두 검사실이 같은 환자의 결과를 주고받습니다. 한 곳은 GRCh37로, 다른 곳은 GRCh38로 일하는데, 어느 쪽도 그 말을 하지 않았습니다. 이것이 잘못되는 구체적인 경로를 하나 서술하고, 어느 쪽이든 그것을 막기 위해 할 수 있었던 일 하나를 말해 보세요.
4. 파일은 유전체의 대략 1퍼센트의 100분의 1을 덮고, 그 자리들은 사람들이 흔히 다르기 때문에 골라진 것입니다. 그 설계가 왜 어떤 일에는 좋고 다른 일에는 나쁜지, 두 가지 일을 모두 이름 붙여 설명해 보세요.
5. 확인을 위해 보내진 변이의 40퍼센트가 위양성이었습니다. 그것은 이 검사가 전체적으로 40퍼센트 틀렸다는 뜻입니까? 그 숫자가 무엇을 재고 무엇을 재지 않는지 조심스럽게 설명해 보세요.

더 나아가기

1. 파일은 각 글자가 어느 쪽 부모에게서 왔는지 결코 기록하지 않습니다. 그것을 알아낼 수 있게 해줄 실험이나 추가 정보를 서술하고, 파일만으로는 내릴 수 없는 어떤 결론을 그것이 가능하게 하는지 말해 보세요.
2. 이 레슨에 나온 오류들을 애초에 저지를 수 없게 만드는 유전체 데이터 파일 머리말을 설계해 보세요. 반드시 넣을 항목을 모두 나열하고, 각각을 한 문장으로 정당화해 보세요.
3. 칩은 이중나선의 특정 가닥 하나에 있는 글자를 보고하고, 두 가닥은 서로 상보적인 글자를 가집니다 — A는 T와, C는 G와 짝을 이룹니다. 두 파일이 같은 사람의 유전형에 대해 서로 다르게 말하면서도 둘 다 옳을 수 있는 경우를 설명하고, 그것을 어떻게 알아낼지 말해 보세요.
4. 빠진 자리들이 중요한지 알고 싶다고 합시다. 소비자용 유전형 파일이 애초에 거냥하지 않은 집단에서 얼마나 많이 놓치는지를 재는 연구를 설계하고, 어떤 결과가 나오면 생각을 바꾸겠는지 말해 보세요.

교사용 메모 — 각 질문이 향하는 곳

1-4는 파일을 훑기만 한 것이 아니라 읽었는지 확인합니다. 3번에는 끌어낼 만한 미묘함이 있습니다. GA와 AG는 데이터로서 같으며, 그 이유는 어느 쪽 부모가 어느 글자를 주었는지를 파일이 애초에 기록하지 않았기 때문입니다. "차이 없다"고 답한 학생은 맞았고, 왜인지 말할 수 있는 학생은 이 파일이 무엇을 버렸는지를 이해한 것입니다.

5는 예측 질문이고, 보람은 집단 수준에 있습니다. 15줄 가운데 10줄이 동형접합인 것은 대부분의 자리에서 한 글자가 다른 글자보다 훨씬 흔하기 때문입니다. 어떤 학급은 대부분이 빨간 구슬인 주머니에 빗대어 여기에 도달하는데, 정확히 옳은 직관이며 그렇게 이름을 붙여줄 만합니다. 대립유전자 빈도를 쓸모 있는 양으로 만드는 것이 바로 그 추론입니다.

6은 안전에 관한 질문이고, 묻지 않아도 가장 먼저 나올 질문입니다. 흔히 "제 검사에서는 깨끗하다고 나왔는데요"라는 형태로 나옵니다. 단단히 짚으십시오. 검사하는 자리 목록은 미리 고정되어 있고 모든 고객에게 동일하므로, 파일에 없다는 것은 그 검사의 성질입니다. 이 대화가 교실 안 누군가의 실제

결과를 해석하는 쪽으로 흘러가지 않게 하십시오.

7에는 학생들이 흔히 저항하는 답이 있습니다. 어느 쪽도 그 자체로는 옳지 않습니다. 좌표는 그것이 속한 판 없이는 아무 뜻이 없기 때문입니다. rsID가 안정적인 식별자이고 좌표는 거기서 파생된 값입니다 — rsID가 존재하는 이유가 바로 그것입니다. 좋은 답은 이것이 주소에 도시 이름이 필요한 것과 같은 이유임을 알아챱니다.

8은 7번의 전문가판이고 구체성에 점수를 줍니다. 구체적인 실패라면 무엇이든 받아주십시오 — 엉뚱한 위치에서 변이를 찾아보고 없다고 보고하는 것, 두 데이터셋을 합치면서 소리 없이 어긋나는 것, 좌표가 아예 다른 유전자 안에 떨어지는 것. 예방책은 필수 머리말 항목이며, 그것이 12번으로 이어집니다.

9는 양쪽을 다 원합니다. 잘하는 일: 아주 많은 사람에게서 흔한 차이를 싸게 재는 것 — 대규모 집단 연구를 애초에 가능하게 만든 것이 그것입니다. 못하는 일: 한 개인이 드물고 심각한 변화를 가졌는지 알려주는 것 — 정작 대부분의 고객이 마음에 두고 있는 바로 그 용도입니다.

10은 유전학 레슨 안에 숨어 있는 통계 질문이고, 여기서 가장 어렵습니다. 40퍼센트는 누군가 임상 검사실에 보낼 만하다고 판단한 변이들 사이의 비율입니다 — 드물고 놀라운 결과 쪽으로 심하게 치우친, 선택된 집합이고, 그쪽이 바로 이런 칩이 가장 약한 곳입니다. 파일 전체를 서술하는 숫자가 아닙니다. "분모가 생각하는 그것이 아니다"에 도달한 학생은 이 과목 바깥에서도 오래 쓸 것을 배운 것입니다.

11-14는 연구 문제를 원하는 학생을 위한 것입니다. 11번은 부모 구분 (phasing)을 가리키며, 일상적인 답 — 부모를 검사한다 — 도 좋은 답입니다. 13번은 가닥 문제이고, 실재하며, 전문가도 걸리고, 알아낼 수 있습니다. A/G 자리가 C/T로 보고되었다면 그것은 단순히 다른 것이 아니라 상보적이기 때문입니다. 정말로 모호한 경우는 A/T와 C/G 로, 가닥을 뒤집어도 그럴듯해 보이는 답이 그대로 나와서 구별할 방법이 없습니다. 그것을 스스로 찾아낸 학생은 이 문제를 어렵게 만드는 핵심을 찾아낸 것입니다.

파일 자체에 대한 메모. 이 파일은 그럴듯한 사람이 아니며, 그러려고 만든 것도 아닙니다. 유전형은 활동에 경우를 하나씩 다 넣으려고 고른 것이지 어떤 집단에서 뽑은 것이 아니어서, 실제로는 함께 나타나지 않을 드문 조합 몇 개가 한자리에 있습니다. 누군가 알아채면 학급에 말해줄 만합니다 — 그리고 알아챈 학생은 앞 절을 이 활동이 요구한 것보다 더 잘 이해한 것입니다.

더 나아가려면

여기 나온 모든 자리에는 이 사이트에 페이지가 있고, 무료이며, 출처가 적혀 있습니다. 이어가기 좋은 곳:

- 유전자와 rsID, 그리고 목록이 실제로 재는 것 — 짝이 되는 레슨입니다. 이 이름들이 어디서 오는지, 그리고 그것을 세는 일이 왜 생물학이 아니라 연구 관심도를 재는지.
- rs1815739 — 위 그림에서 뜯어본 그 줄. 근육 단백질 하나를 바꾸는 글자 하나입니다.
- rs4244285 — 대조군. 우리 파일이 침묵하는 자리 이고, 용량이 거기에 달린 약이 있습니다.
- 유전자 지식 그래프 — 이 목록이 오늘 담고 있는 자리·유전자·연결이 몇 개인지, 실시간으로 셉니다.
- 이 사이트가 딛고 선 논문들 — 이 페이지들 뒤에 있는 모든 연구를, 모든 저자의 이름과 함께.

자유롭게 쓰세요

어느 교실에서든, 어느 수준에서든, 상업적 이용을 포함해 자유롭게 쓰고 복제하고 고칠 수 있습니다. 출처 표기는 환영하지만 필수가 아닙니다. 이 사이트는 학생에게서 아무것도 수집하지 않습니다. 계정도, 업로드도, 유전 정보도, 결코.