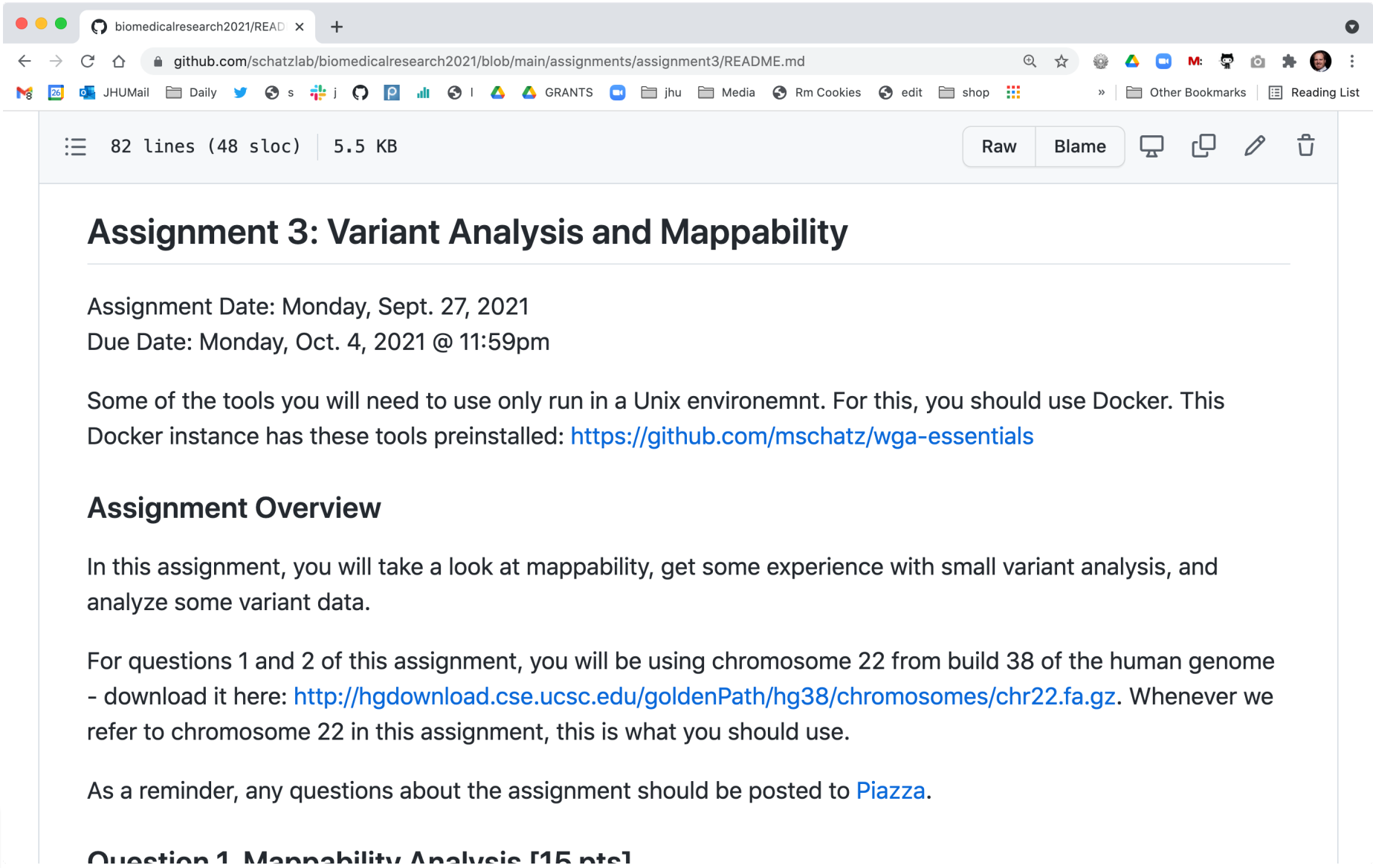
Annotation

Michael Schatz

Oct 6, 2021

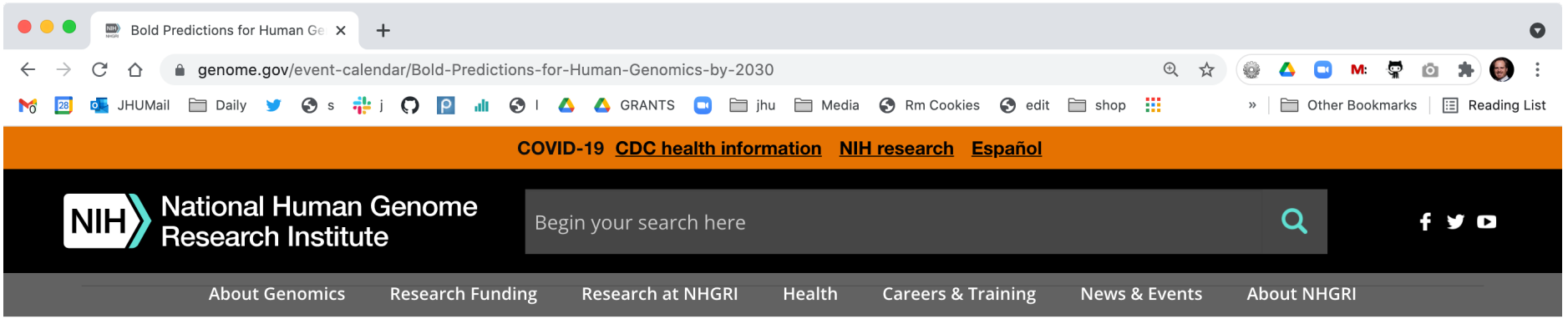
Lecture 11: Computational Biomedical Research

Assignment 3: Variant Analysis & Mappability
Due Oct 4 @ 11:59pm



https://github.com/schatzlab/biomedicalresearch2021

Monday's class



Upcoming

Session 8 - October 4, 2021, 3 p.m. to 4:30 p.m.

Bold Prediction #8: A person's complete genome sequence along with informative annotations can be securely and readily accessible on their smartphone.

Speakers:

Michael Schatz, Ph.D.
Johns Hopkins University & Cold Spring Harbor Laboratory

Gillian Hooker, Ph.D., ScM, LCGC
Concert Genetics

Moderator:

Sarah Bates, M.S.
NHGRI

Session 9 - November 1, 2021, 3 p.m. to 4:30 p.m.

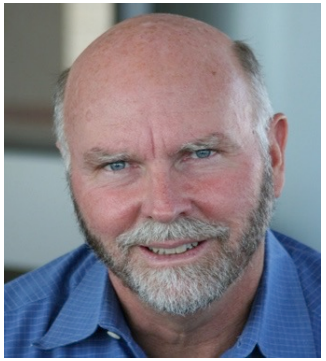
Bold Prediction #9: Individuals from ancestrally diverse backgrounds will benefit equitably from advances in human genomics.

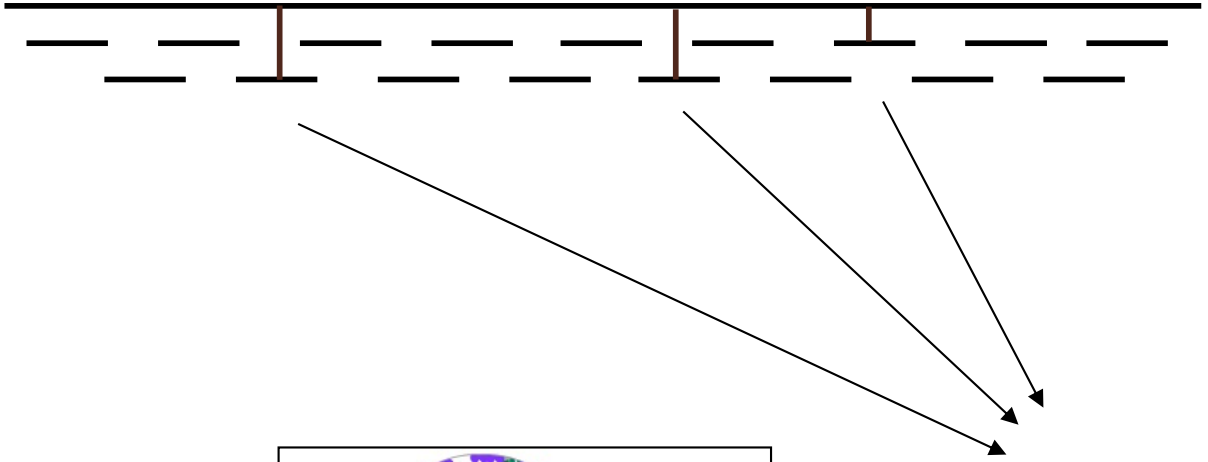
Speakers:

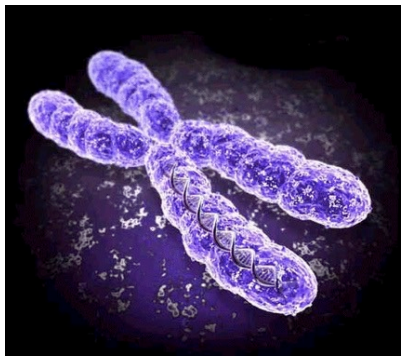
Registration: bit.ly/2XXhLYJ

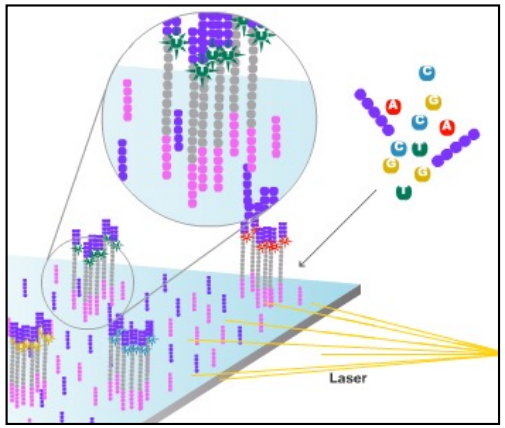
Personal Genomics

How does your genome compare to the reference?







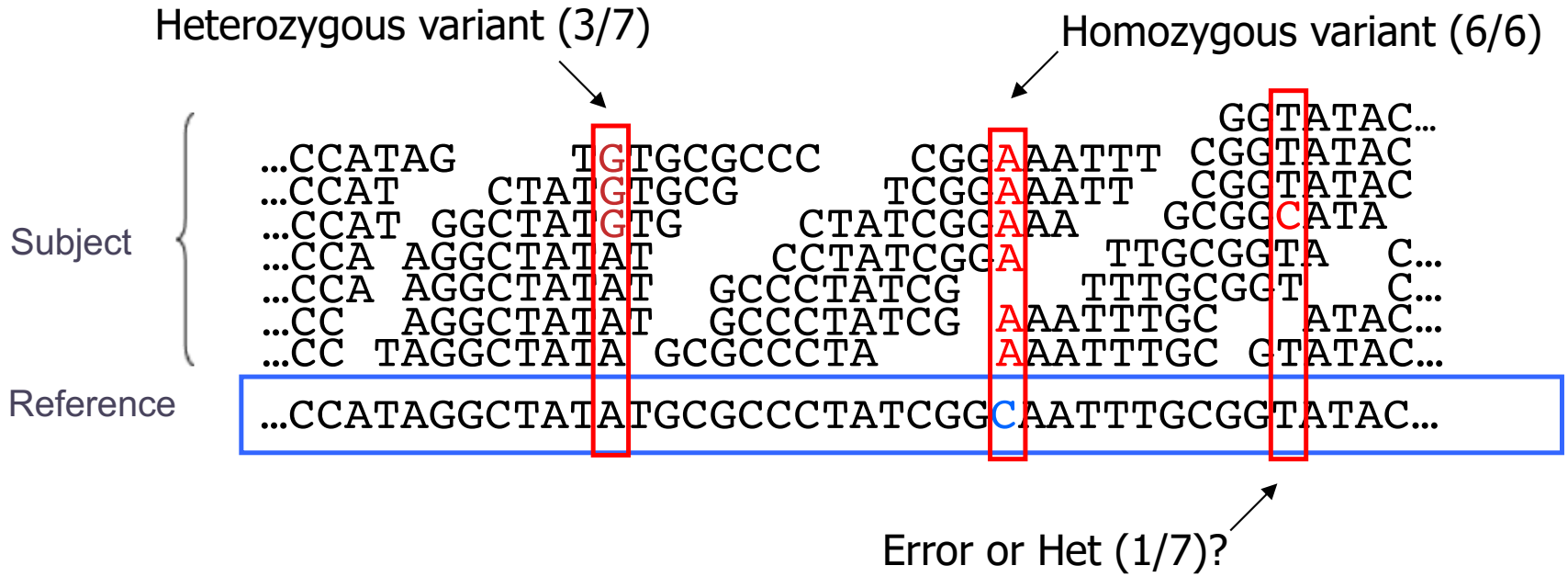


Heart Disease

Cancer

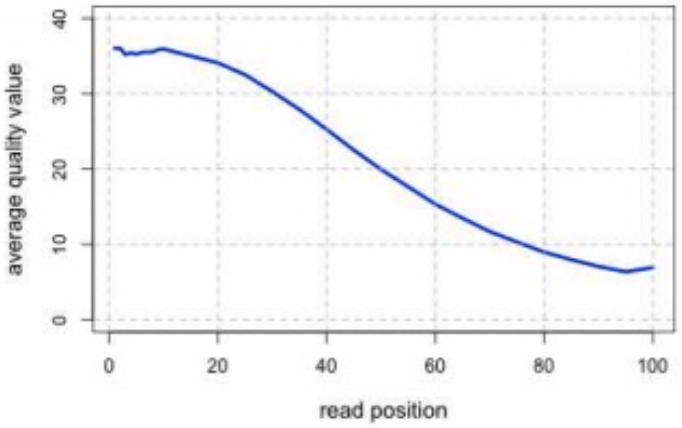
Presidential smile

Genotyping Theory



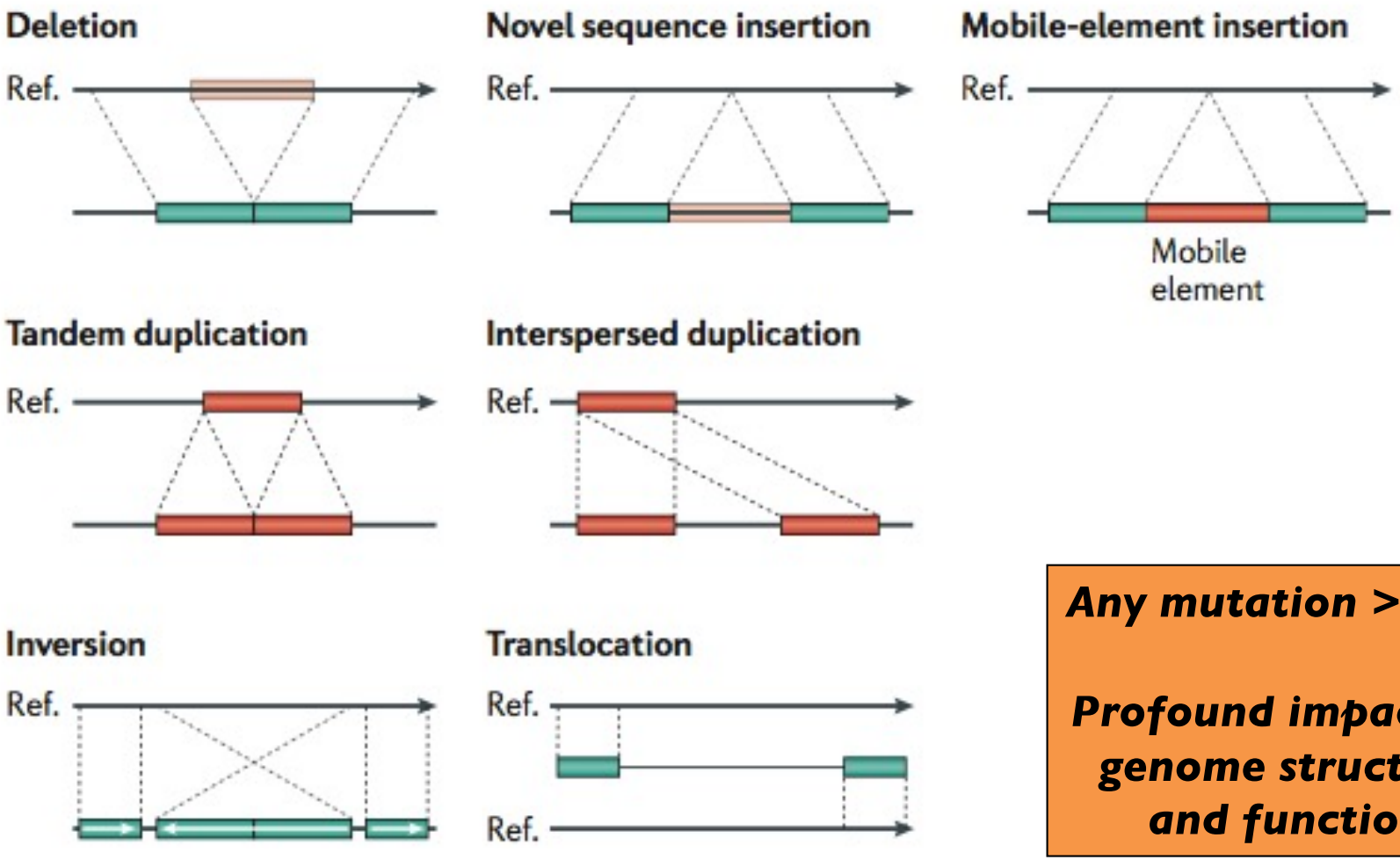
- If there were no sequencing errors, identifying SNPs would be very easy: any time a read disagrees with the reference, it must be a variant!

- Sequencing instruments make mistakes
  - Quality of read decreases over the read length



- A single read differing from the reference is probably just an error, but it becomes more likely to be real as we see it multiple times

Structural Variations



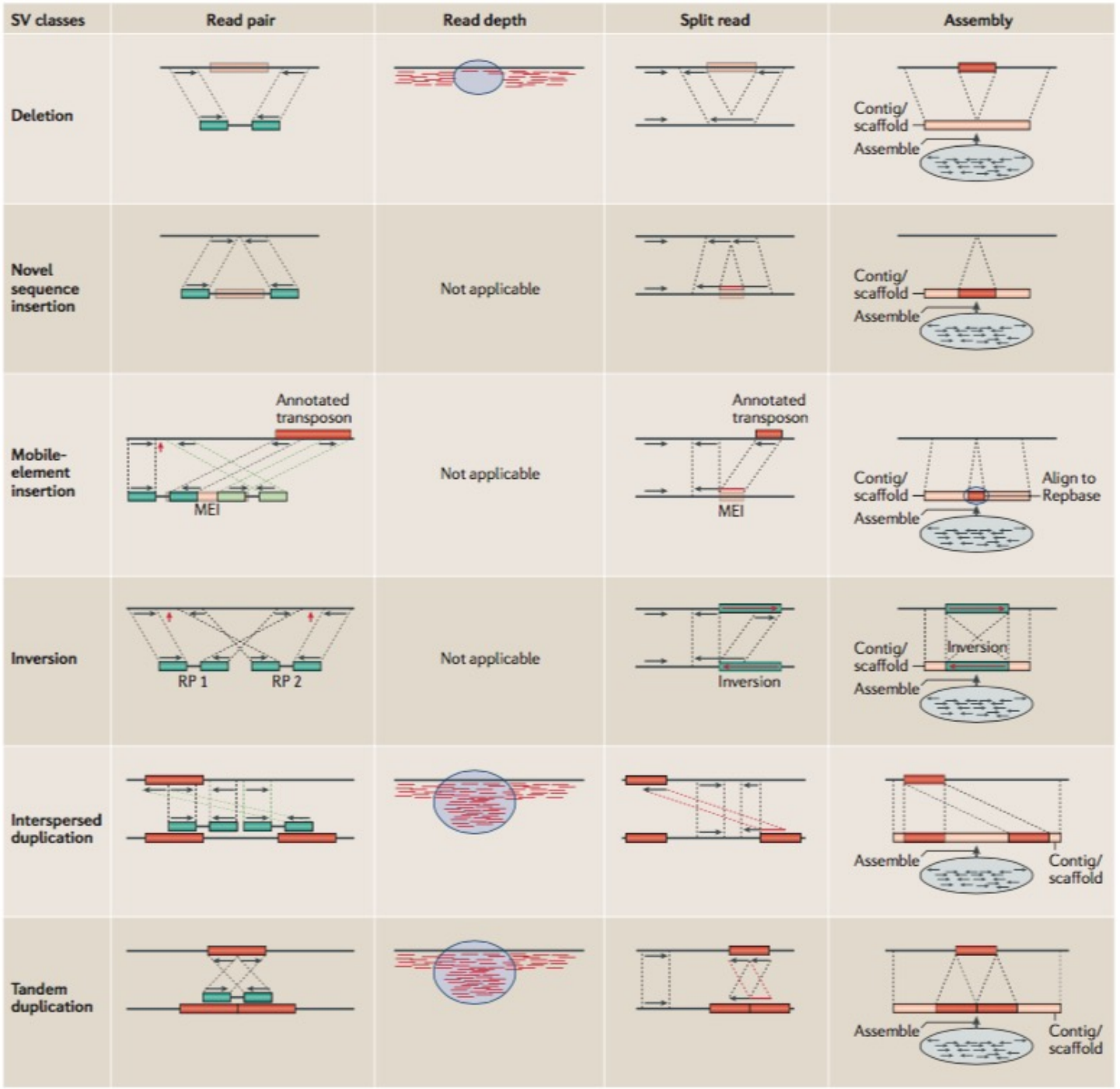
Any mutation >50bp

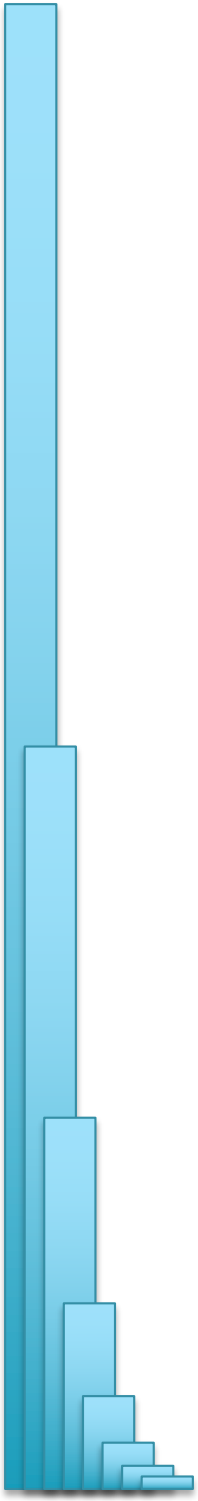
Profound impact on genome structure and function

Genome structural variation discovery and genotyping

Alkan, C, Coe, BP, Eichler, EE (2011) *Nature Reviews Genetics*. May;12(5):363-76. doi: 10.1038/nrg2958.

Structural Variation Sequence Signatures





Annotation

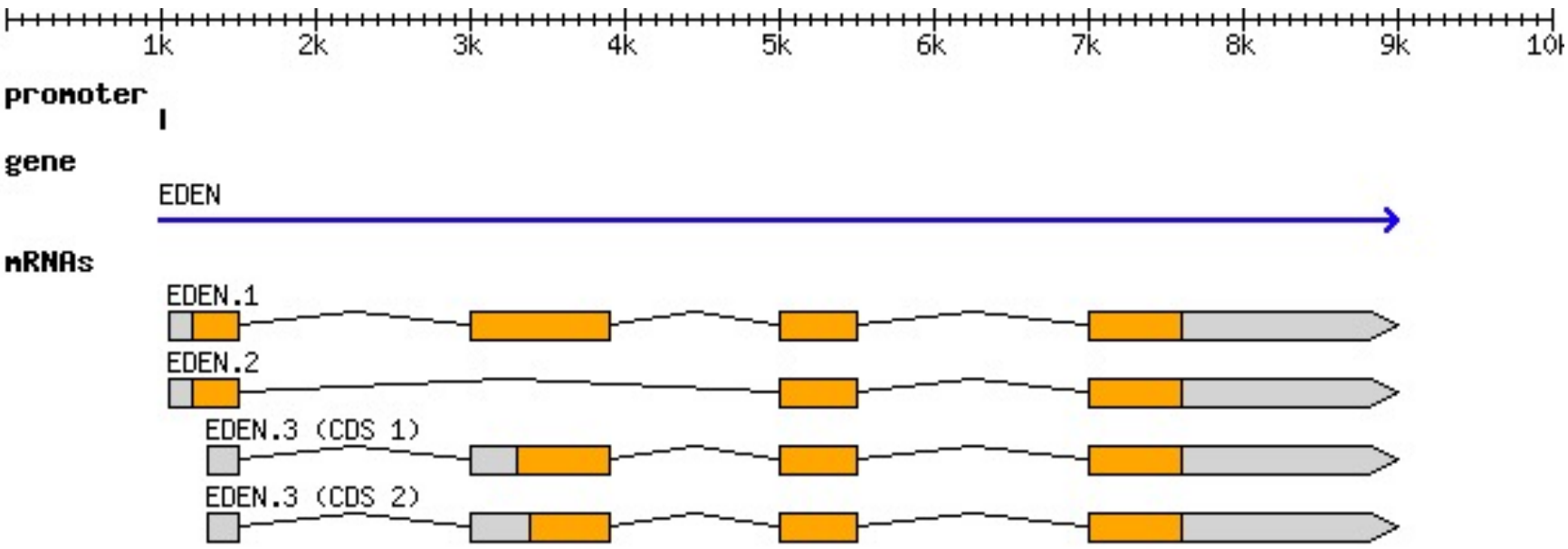
Goal: Genome Annotations

aatgcatgcggctatgctaatgcatgcggctatgctaagctgggatccgatgacaatgcatgcggctatgctaat
gcatgcggctatgcaagctgggatccgatgactatgctaagctgggatccgatgacaatgcatgcggctatgct
aatgaatggtcttgggatttaccttggaatgctaagctgggatccgatgacaatgcatgcggctatgctaatgaa
tggtcttgggatttaccttggaatatgctaatgcatgcggctatgctaagctgggatccgatgacaatgcatgcg
gctatgctaatgcatgcggctatgcaagctgggatccgatgactatgctaagctgcggctatgctaatgcatgcg
gctatgctaagctgggatccgatgacaatgcatgcggctatgctaatgcatgcggctatgcaagctgggatcct
gcggctatgctaatgaatggtcttgggatttaccttggaatgctaagctgggatccgatgacaatgcatgcggct
atgctaatgaatggtcttgggatttaccttggaatatgctaatgcatgcggctatgctaagctgggaatgcatgcg
gctatgctaagctgggatccgatgacaatgcatgcggctatgctaatgcatgcggctatgcaagctgggatccg
atgactatgctaagctgcggctatgctaatgcatgcggctatgctaagctcatgcggctatgctaagctgggaat
gcatgcggctatgctaagctgggatccgatgacaatgcatgcggctatgctaatgcatgcggctatgcaagctg
ggatccgatgactatgctaagctgcggctatgctaatgcatgcggctatgctaagctcggctatgctaatgaatg
gtcttgggatttaccttggaatgctaagctgggatccgatgacaatgcatgcggctatgctaatgaatggtcttgg
gatttaccttggaatatgctaatgcatgcggctatgctaagctgggaatgcatgcggctatgctaagctgggatc
cgatgacaatgcatgcggctatgctaatgcatgcggctatgcaagctgggatccgatgactatgctaagctgcg
gctatgctaatgcatgcggctatgctaagctcatgcgg

Goal: Genome Annotations

Gene!

Gene Models



- "Generic Feature Format" (GFF) records genomic features
  - Coordinates of each exon
  - Coordinates of UTRs
  - Link together exons into transcripts
  - Link together transcripts into gene models

http://www.sequenceontology.org/gff3.shtml

GFF File format

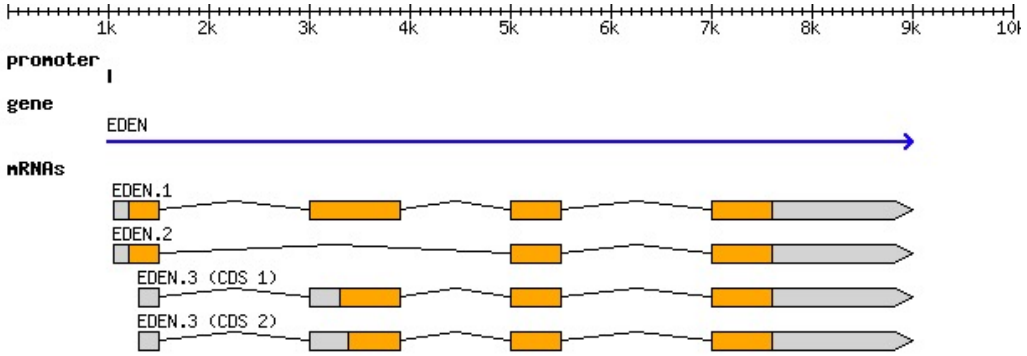
GFF3 files are nine-column, tab-delimited, plain text files

1. ***seqid:*** The ID of the sequence
2. ***source:*** Algorithm or database that generated this feature
3. ***type:*** *gene/exon/CDS/etc…*
4. ***start:*** 1-based coordinate
5. ***end***: 1-based coordinate
6. ***score***: E-values/p-values/index/colors/…
7. ***strand:*** "+' for positive "-" for minus, "." not stranded
8. ***phase:*** For "CDS", where the feature begins with reference to the reading frame (0,1,2)
9. ***attributes:*** A list of tag=value features

   Parent: Indicates the parent of the feature (group exons into transcripts, transcripts into genes, …)

GFF Example

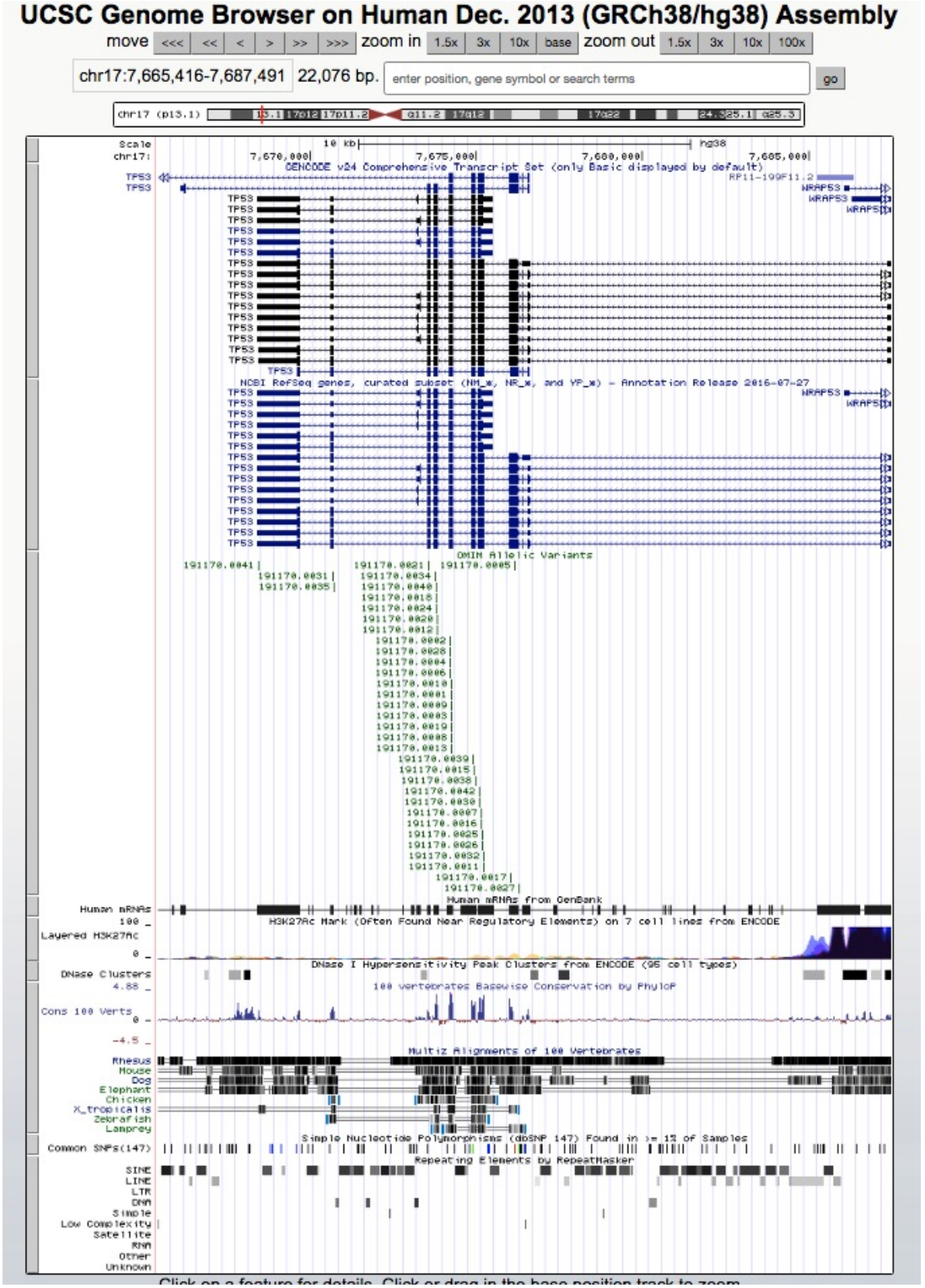
Gene "EDEN" with 3 alternatively spliced transcripts, isoform 3 has two alternative translation start sites



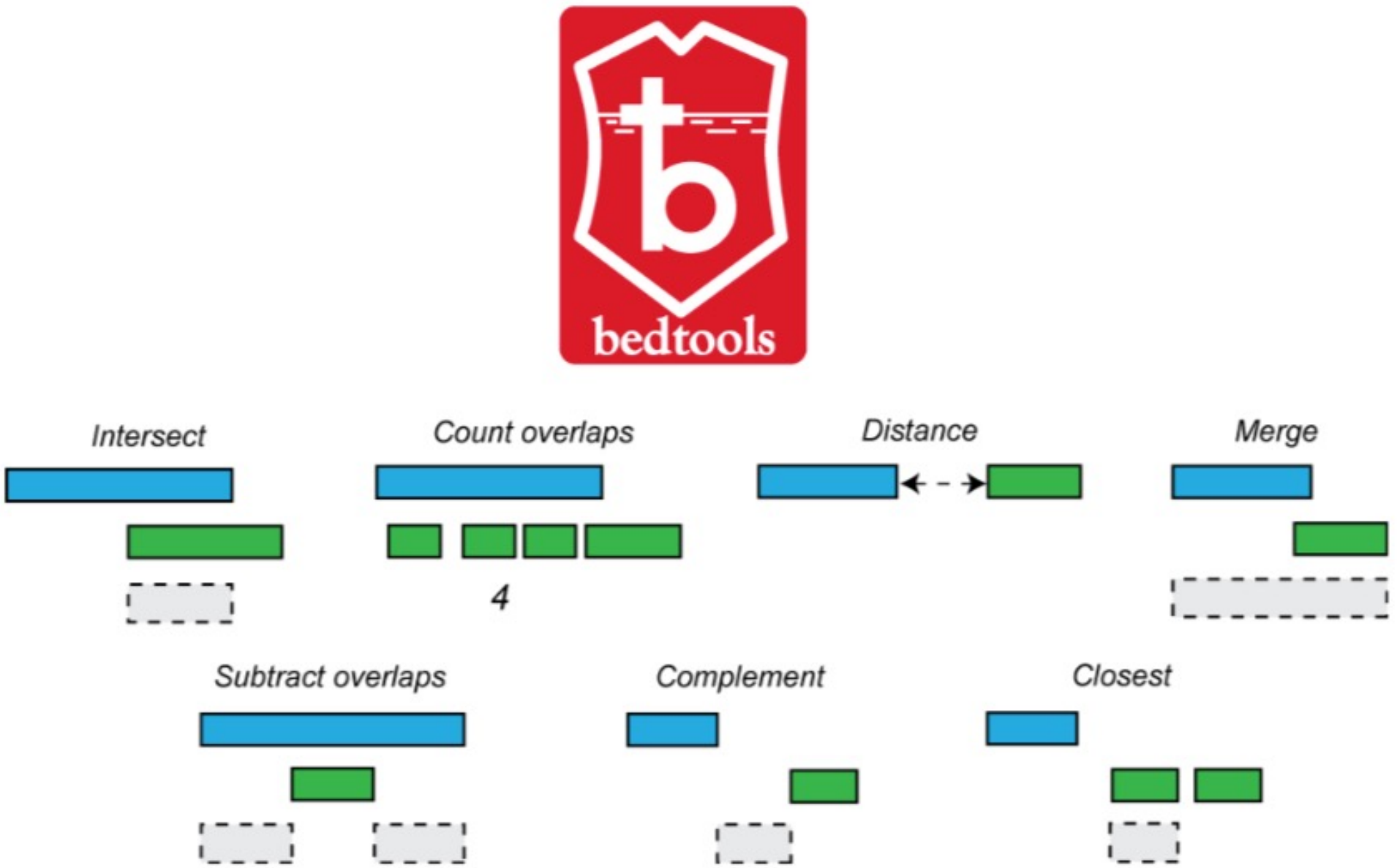

```
##gff-version 3
##sequence-region   ctg123 1 1497228
ctg123 . gene            1000  9000  .  +  .  ID=gene00001;Name=EDEN

ctg123 . TF_binding_site 1000  1012  .  +  .  ID=tfbs00001;Parent=gene00001

ctg123 . mRNA            1050  9000  .  +  .  ID=mRNA00001;Parent=gene00001;Name=EDEN.1
ctg123 . mRNA            1050  9000  .  +  .  ID=mRNA00002;Parent=gene00001;Name=EDEN.2
ctg123 . mRNA            1300  9000  .  +  .  ID=mRNA00003;Parent=gene00001;Name=EDEN.3

ctg123 . exon            1300  1500  .  +  .  ID=exon00001;Parent=mRNA00003
ctg123 . exon            1050  1500  .  +  .  ID=exon00002;Parent=mRNA00001,mRNA00002
ctg123 . exon            3000  3902  .  +  .  ID=exon00003;Parent=mRNA00001,mRNA00003
ctg123 . exon            5000  5500  .  +  .  ID=exon00004;Parent=mRNA00001,mRNA00002,mRNA00003
ctg123 . exon            7000  9000  .  +  .  ID=exon00005;Parent=mRNA00001,mRNA00002,mRNA00003

ctg123 . CDS             1201  1500  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
ctg123 . CDS             3000  3902  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
ctg123 . CDS             5000  5500  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
ctg123 . CDS             7000  7600  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1

ctg123 . CDS             1201  1500  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
ctg123 . CDS             5000  5500  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
ctg123 . CDS             7000  7600  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2

ctg123 . CDS             3301  3902  .  +  0  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
ctg123 . CDS             5000  5500  .  +  1  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
ctg123 . CDS             7000  7600  .  +  1  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3

ctg123 . CDS             3391  3902  .  +  0  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
ctg123 . CDS             5000  5500  .  +  1  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
ctg123 . CDS             7000  7600  .  +  1  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
```

What are genome intervals?

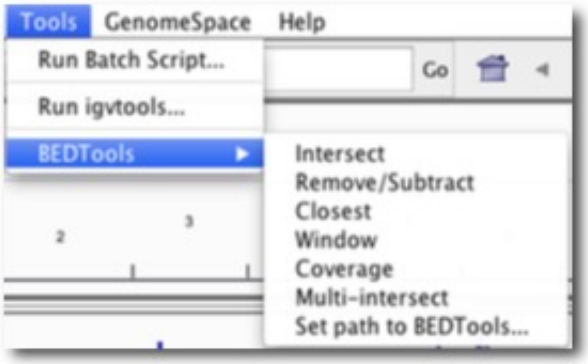
- Genetic variation:
  - SNPs: 1bp
  - Indels: 1-50bp
  - SVs: >50bp
- Genes:
  - exons, introns, UTRs, promoters
- Conservation
- Transposons
- Origins of replication
- TF binding sites
- CpG islands
- Segmental duplications
- Sequence alignments
- Chromatin annotations
- Gene expression data
- …
- ***Your own observations and data: put them into context!***



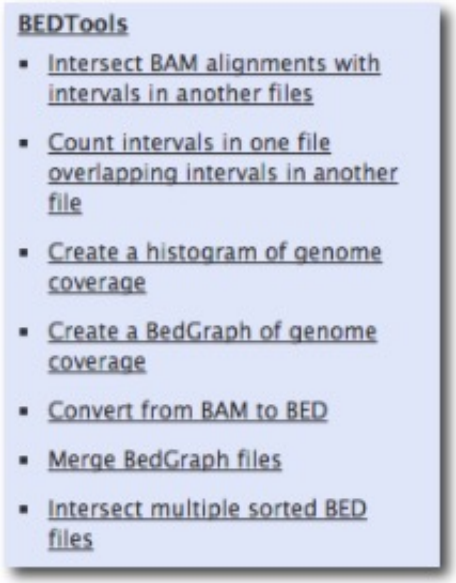
BEDTools to the rescue!



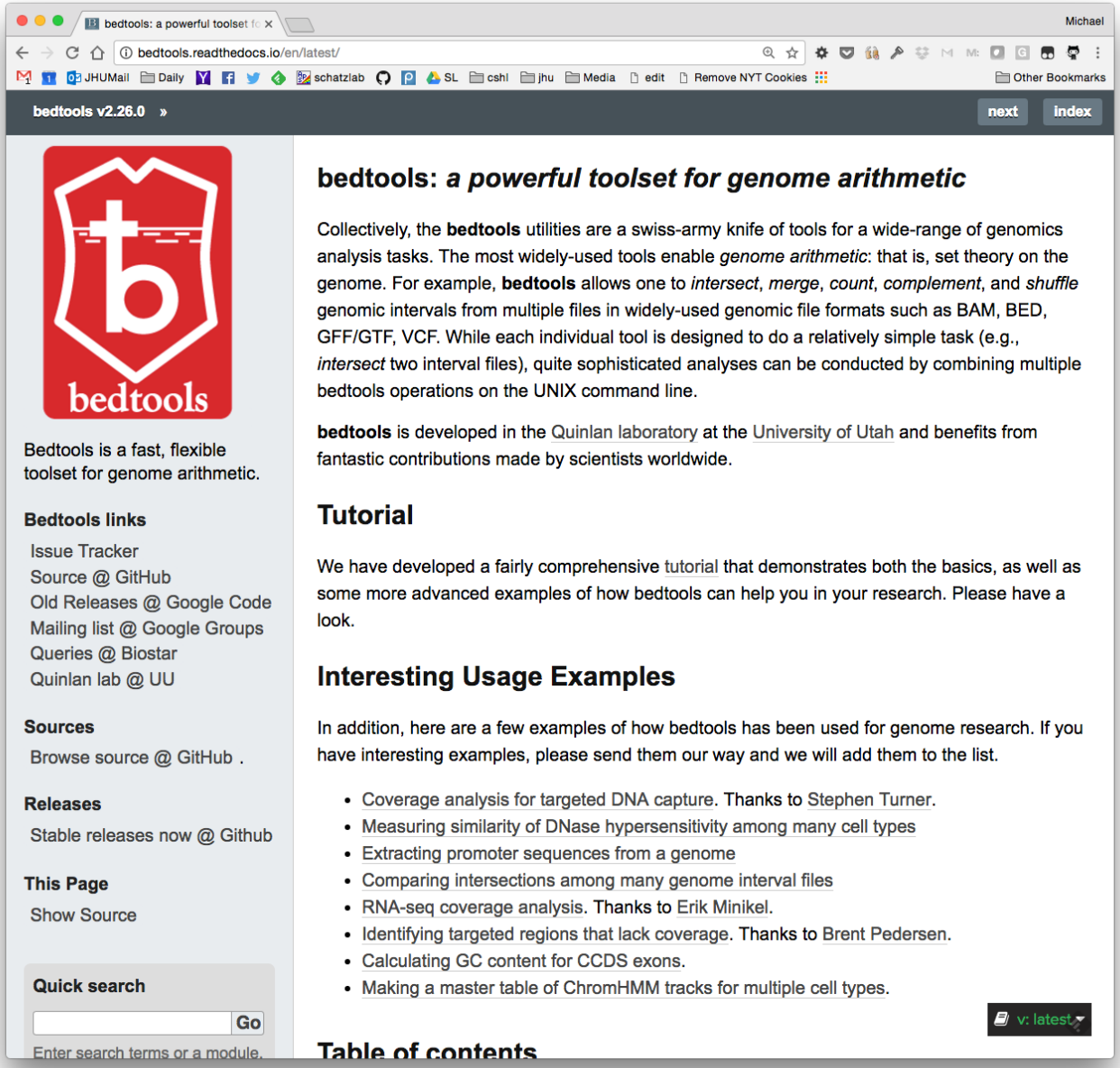
Getting & Using BEDTools



Integrated into **IGV**



In **Galaxy** Toolshed



Extensive Documentation and Examples

BED Format

BED (Browser Extensible Data) format provides a flexible way to define intervals.

The first three required BED fields are:

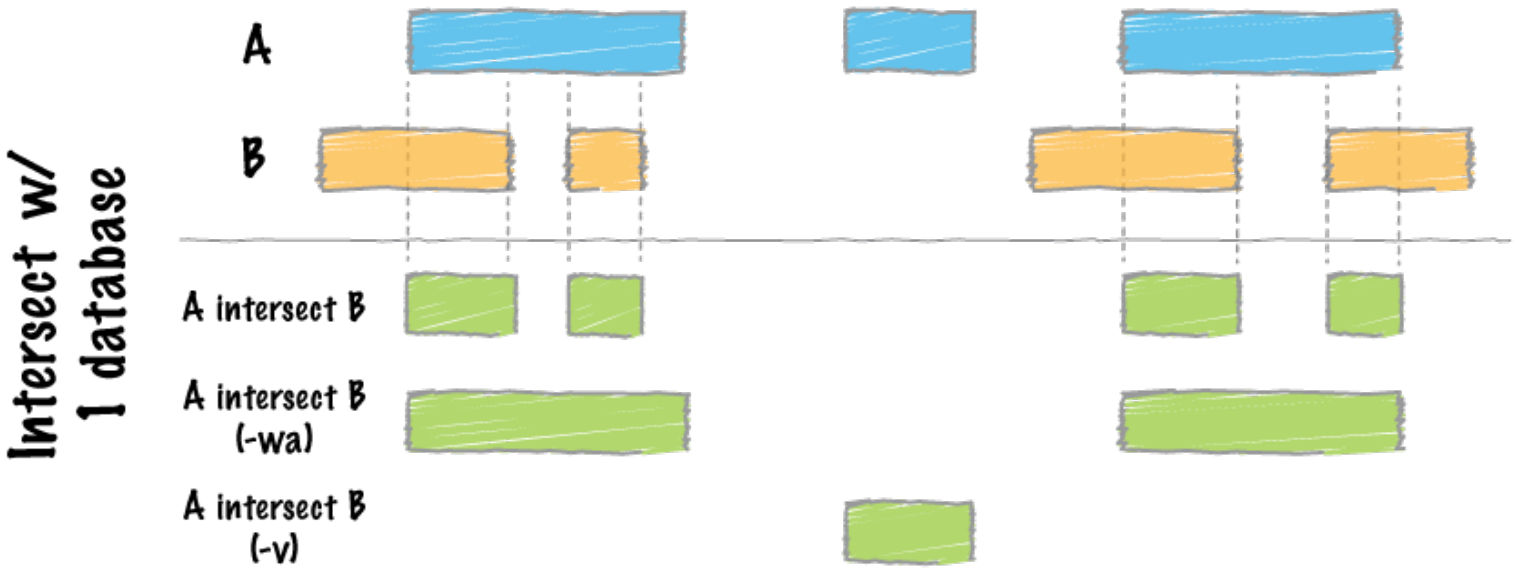
1. chrom — The name of the chromosome (e.g. chr3, chrY, chr2_random) or scaffold (e.g. scaffold10671).
2. chromStart — The starting position of the feature in the chromosome or scaffold. The first base in a sequence is numbered 0.
3. chromEnd — The ending position of the feature in the chromosome or scaffold.
   The chromEnd base is not included in the display of the feature. For example, the first 100 bases of a chromosome are defined as chromStart=0, chromEnd=100, and span the bases numbered 0-99.

The 9 additional optional BED fields are:

1. name - Defines the name of the BED line
2. score - A score between 0 and 1000
3. strand - Defines the strand. Either "." (=no strand) or "+" or "-".
4. thickStart - The starting position at which the feature is drawn thickly
5. thickEnd - The ending position at which the feature is drawn thickly (for example the stop codon in gene displays).
6. itemRgb - An RGB value of the form R,G,B (e.g. 255,0,0).
7. blockCount - The number of blocks (exons) in the BED line.
8. blockSizes - A comma-separated list of the block sizes. The number of items in this list should correspond to blockCount.
9. blockStarts - A comma-separated list of block starts. All of the blockStart positions should be calculated relative to chromStart. The number of items in this list should correspond to blockCount.

```
## genes.bed has: chrom, txStart, txEnd, name, num_exons, and strand
$ head -n4 genes.bed
chr1    134212701    134230065    Nuak2    8    +
chr1    134212701    134230065    Nuak2    7    +
chr1    33510655     33726603     Prim2,   14   -
chr1    25124320     25886552     Bai3,    31   -
```

BEDTools Intersect



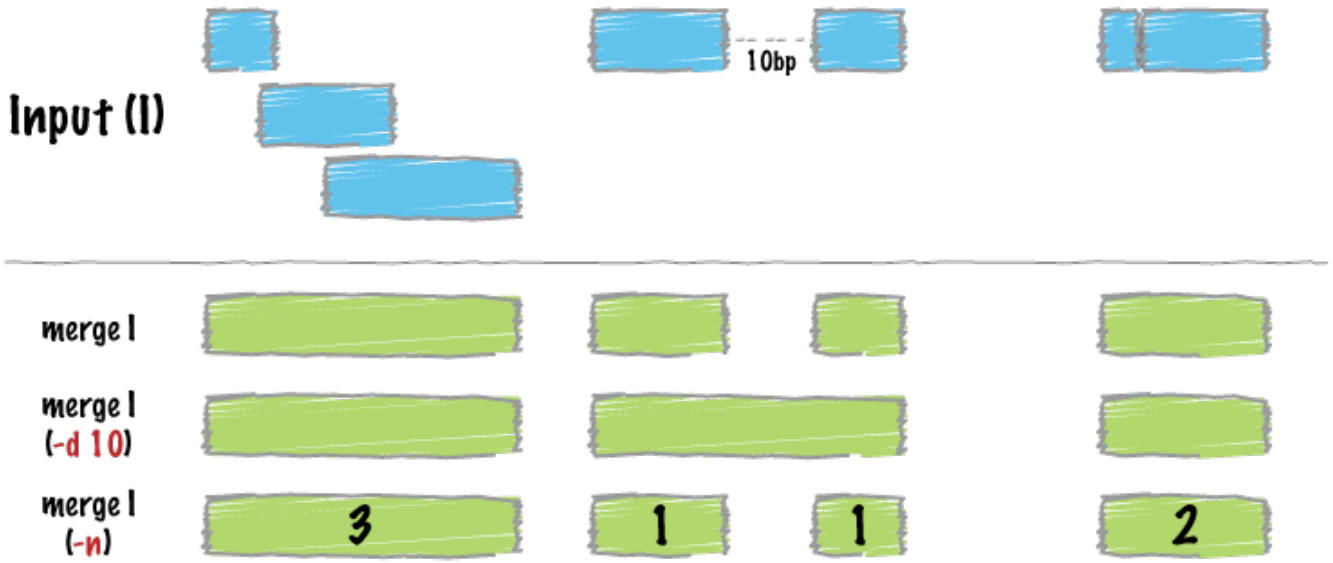
What exons are hit by SVs?

```
$ cat A.bed
chr1  10  20
chr1  30   40

$ cat B.bed
chr1  15  20

$ bedtools intersect -a A.bed -b B.bed -wa
chr1  10   20
```

What parts of exons are hit by SVs?

```
$ cat A.bed
chr1  10  20
chr1  30  40

$ cat B.bed
chr1  15   20

$ bedtools intersect -a A.bed -b B.bed
chr1  15   20
```

BEDTools Merge



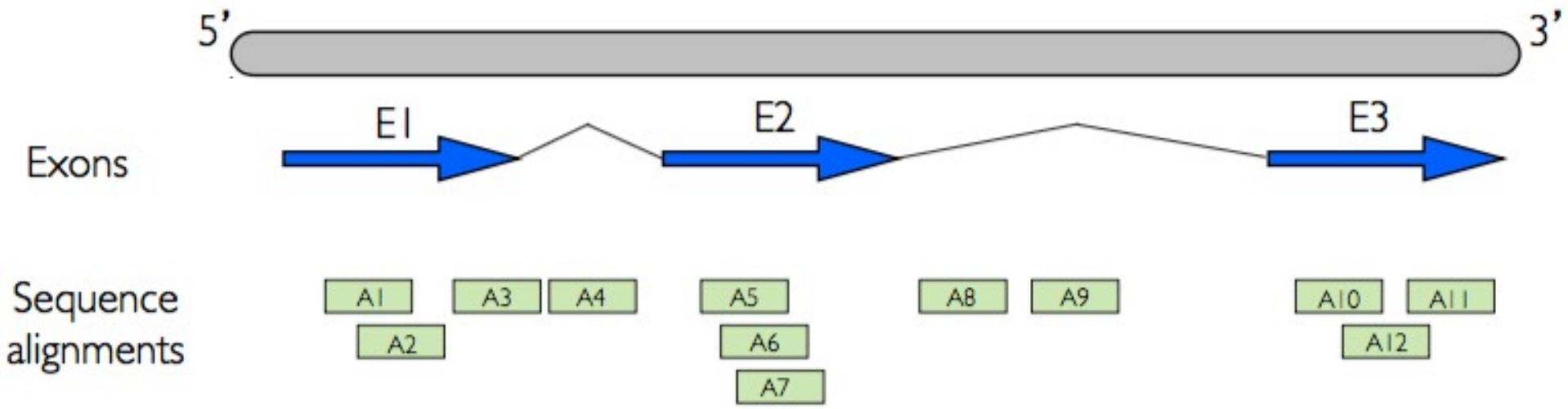
What parts of the genome are exonic?

```
bedtools merge -i exons.bed | head -n 20
chr1    11873   12227
chr1    12612   12721
chr1    13220   14829
chr1    14969   15038
chr1    15795   15947
chr1    16606   16765
chr1    16857   17055
```

Note input must be sorted!

```
sort -k1,1 -k2,2n foo.bed > foo.sort.bed
```

BEDTools Performance



How many reads are aligned to exonic sequences?

```
$ awk '{if ($3=="exon"){print}}' gencode.v21.annotation.gff3 | wc -l
1162114
```

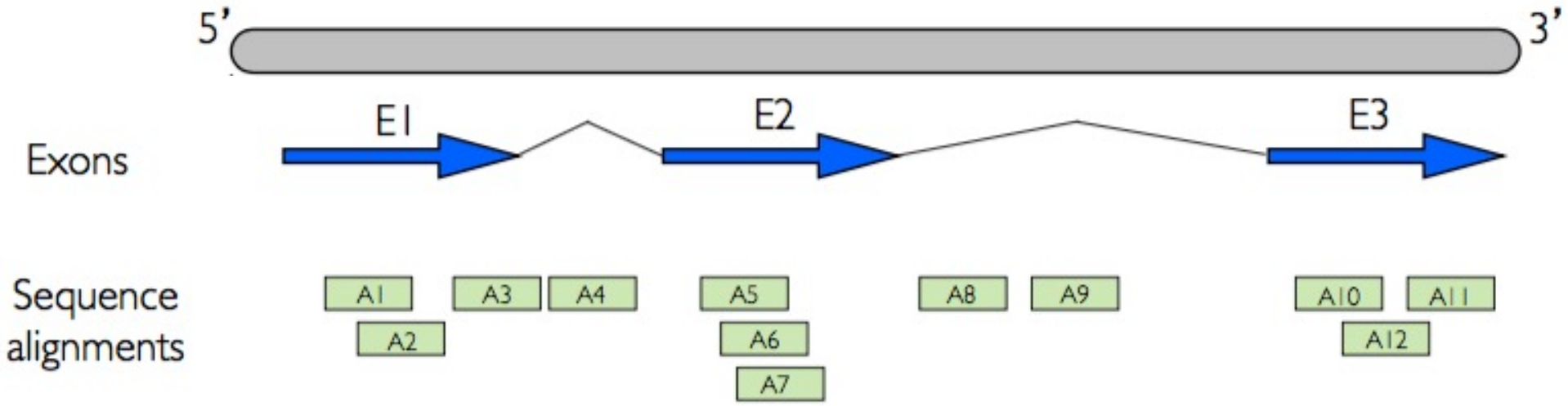
```
if ((read.start <= exon.end) && (read.end >= exon.start)) { print “in exon!”; }
```

How many comparison would a brute force approach take to scan a 30x dataset?

30x3Gb = 90Gbp / 100bp reads = 900M reads

900M reads x 1.1M exons = 990MM comparisons! ☹

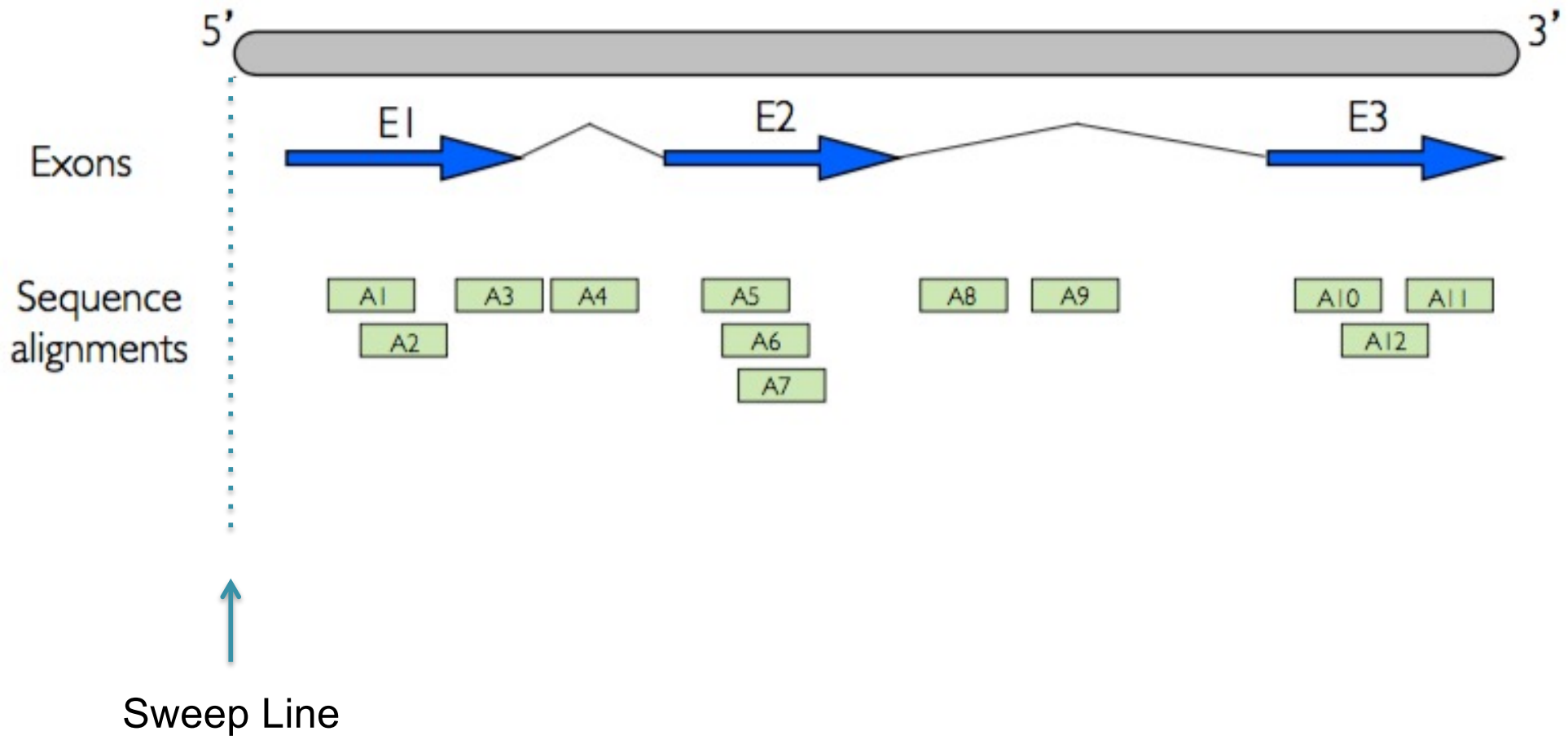
BEDTools Performance



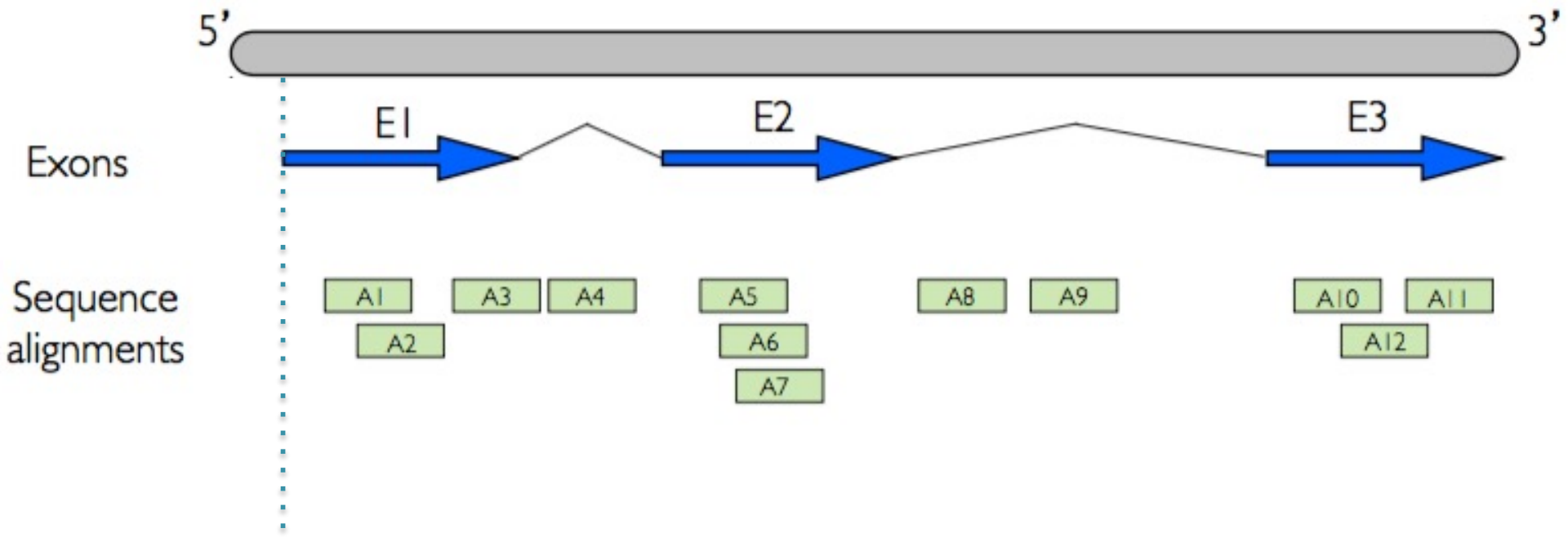
Assume datasets are sorted by chromosome and start position
`samtools sort` to sort alignments, unix `sort` to sort BED file

Any ideas?

Plane Sweep to the Rescue!



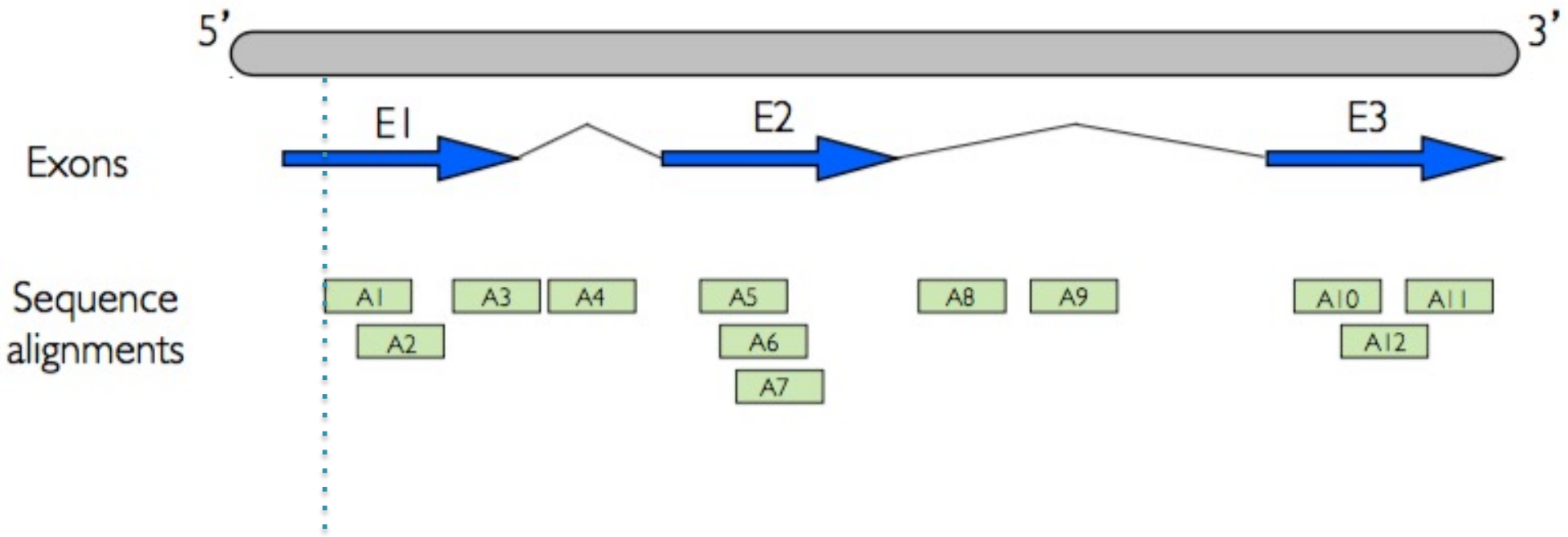
Plane Sweep to the Rescue!



Start of E1
E1 is active

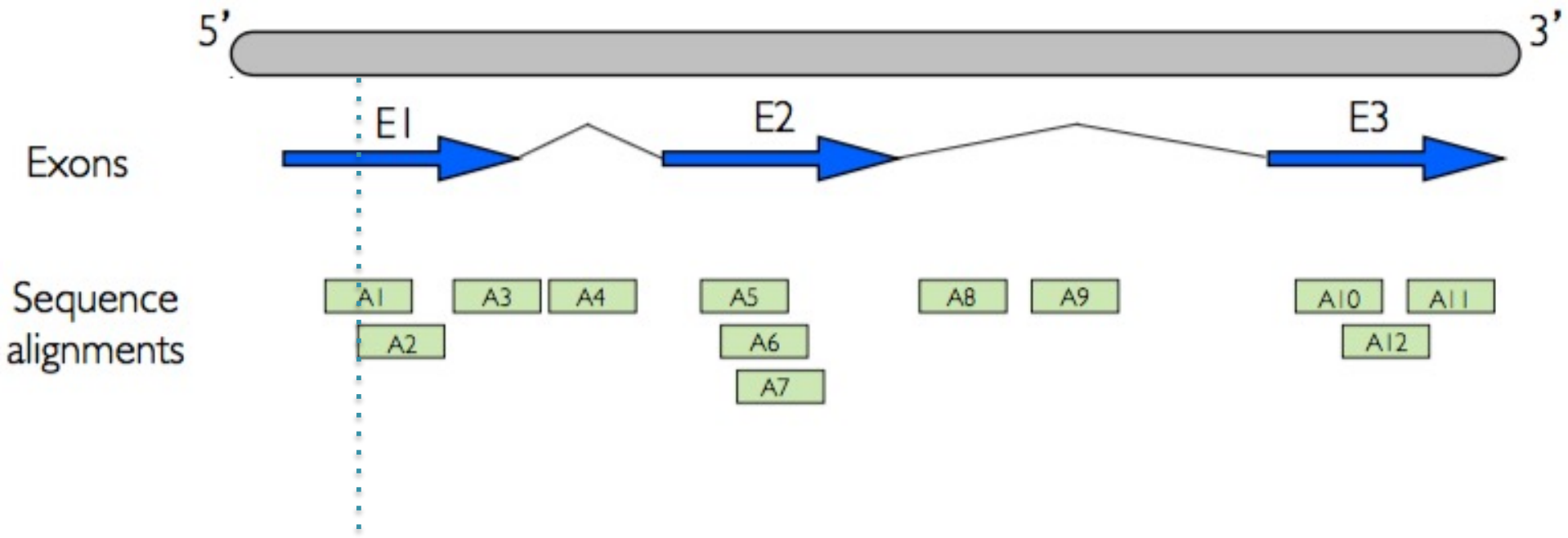
{E1}

Plane Sweep to the Rescue!



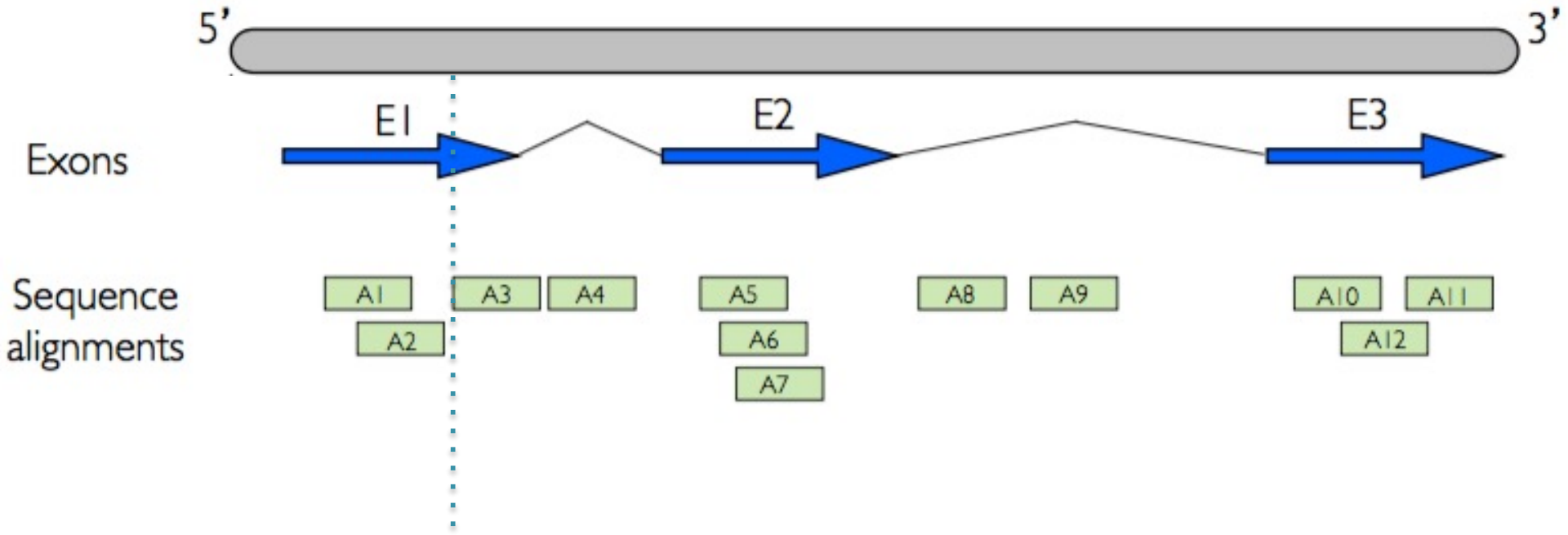

```
A1 overlaps E1

   {E1=(A1)}
```

Plane Sweep to the Rescue!



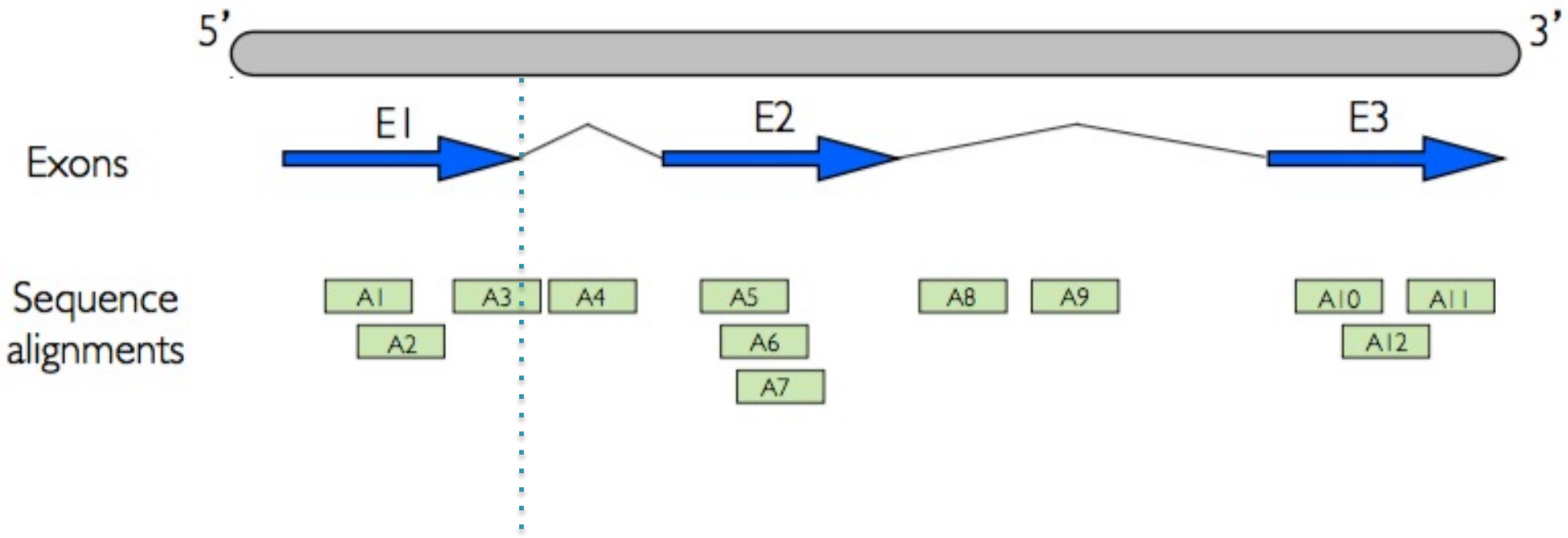

```
A2 overlaps E1

{E1=(A1, A2)}
```

Plane Sweep to the Rescue!




```
A3 overlaps E1

{E1=(A1, A2, A3)}
```

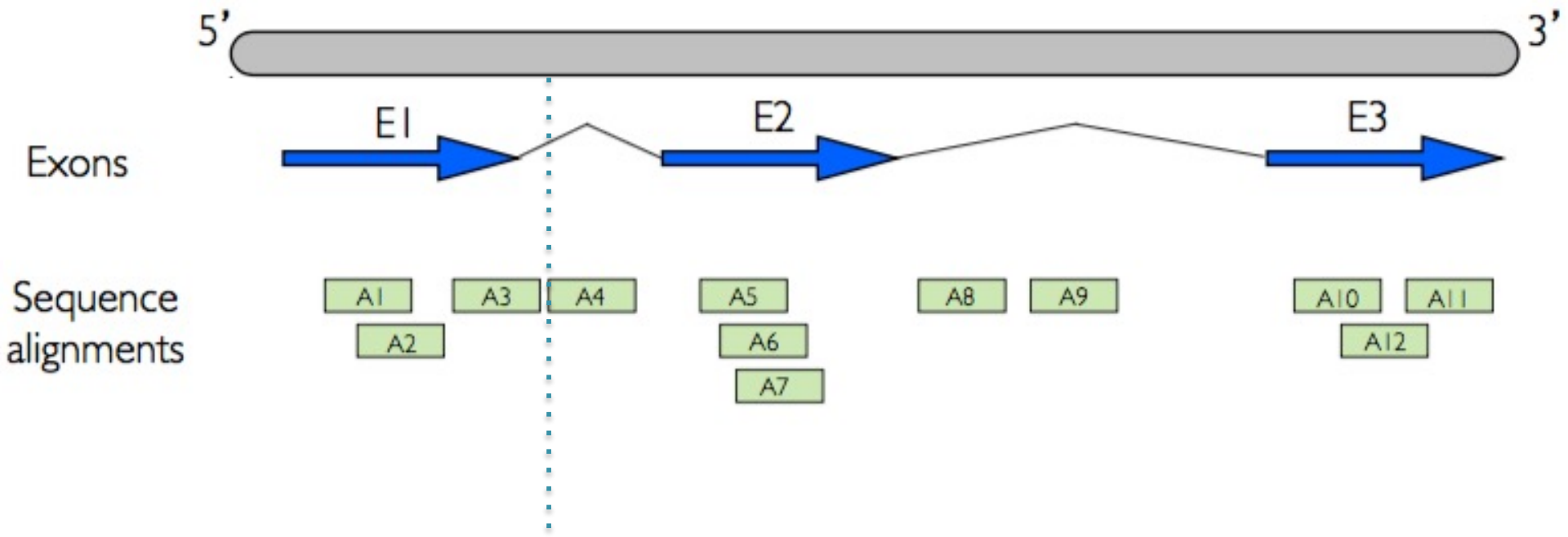
Plane Sweep to the Rescue!



End of E1

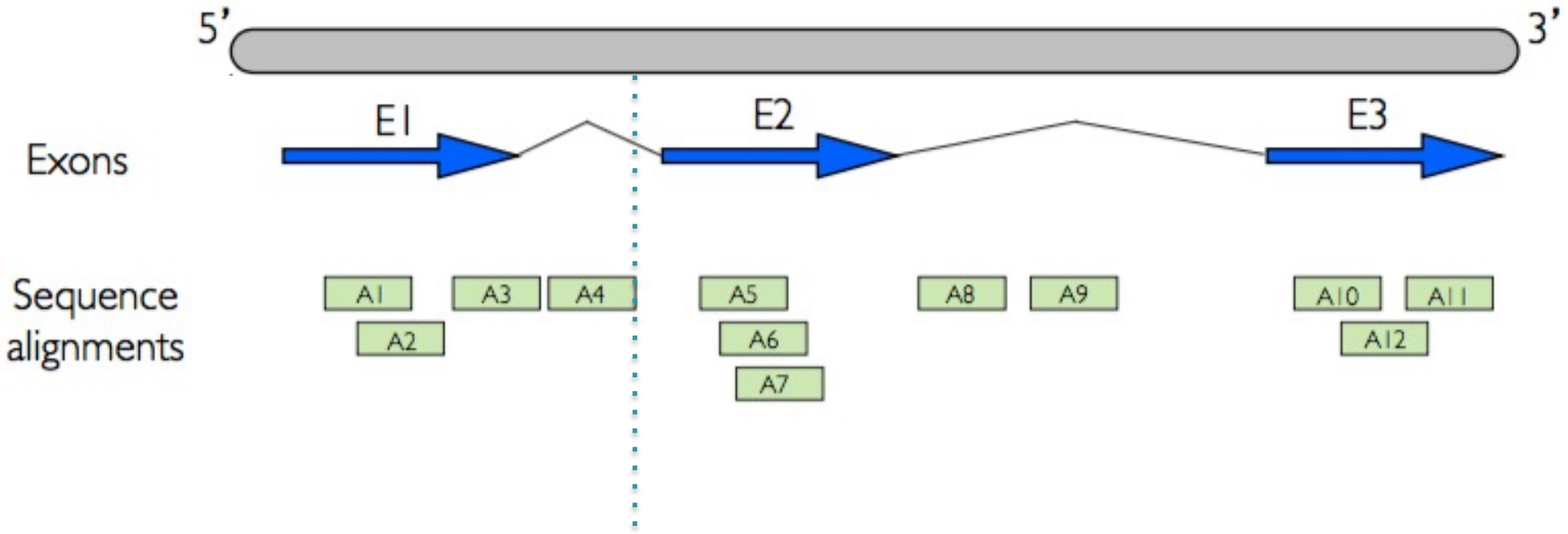
Report:
{E1=(A1, A2, A3)}

Plane Sweep to the Rescue!



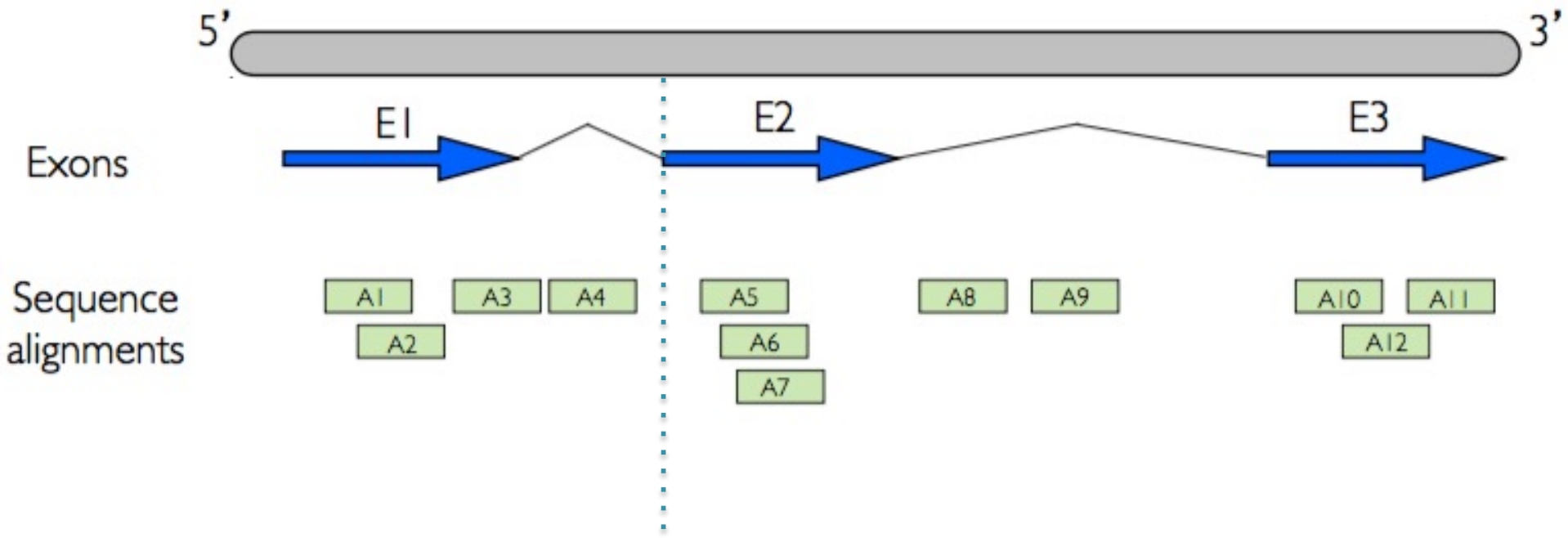
A4 starts,
but nothing is active

Plane Sweep to the Rescue!



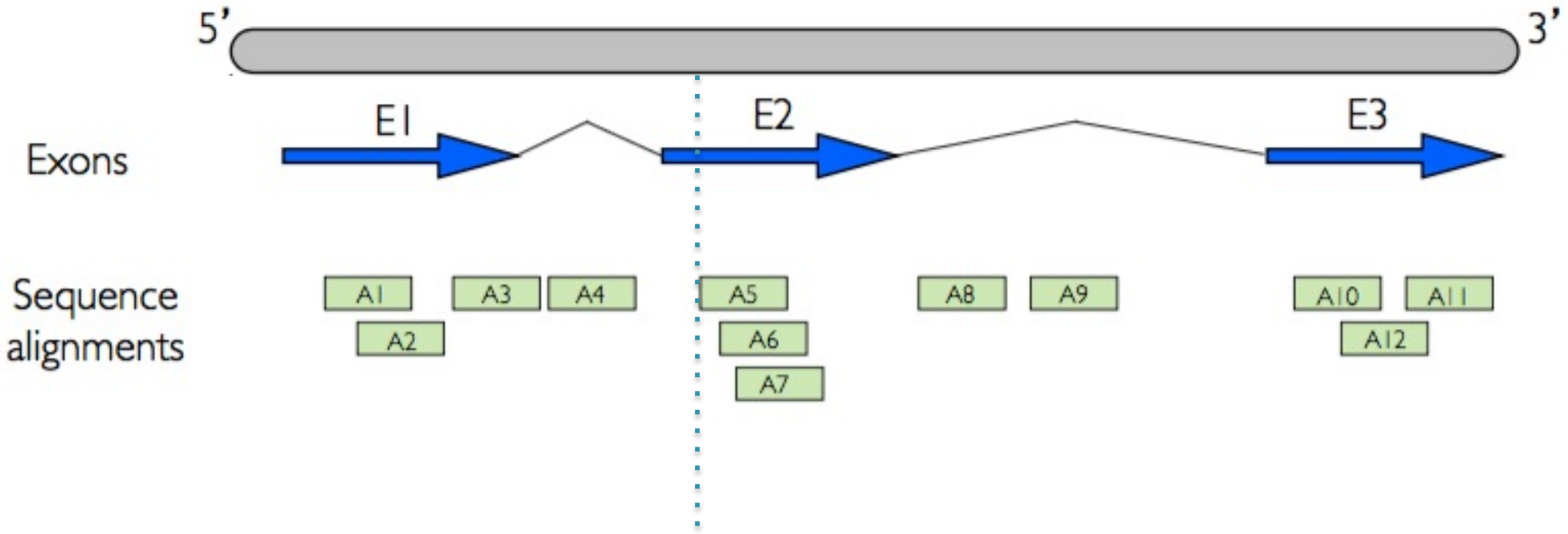
A4 end,
but nothing is active

Plane Sweep to the Rescue!




```
E2 starts

   {E2}
```

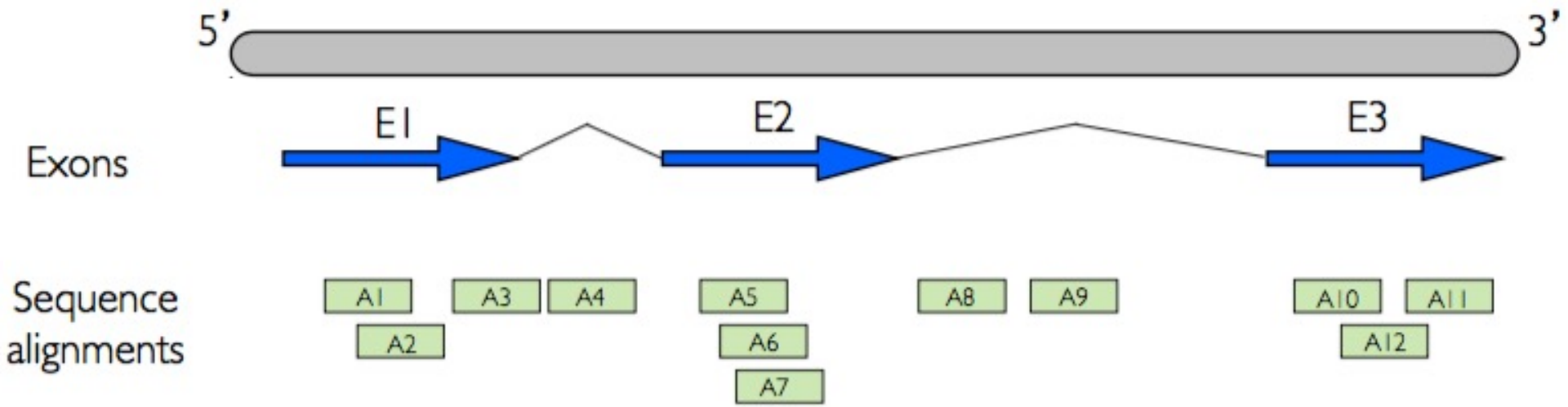
Plane Sweep to the Rescue!



A5 overlaps E2

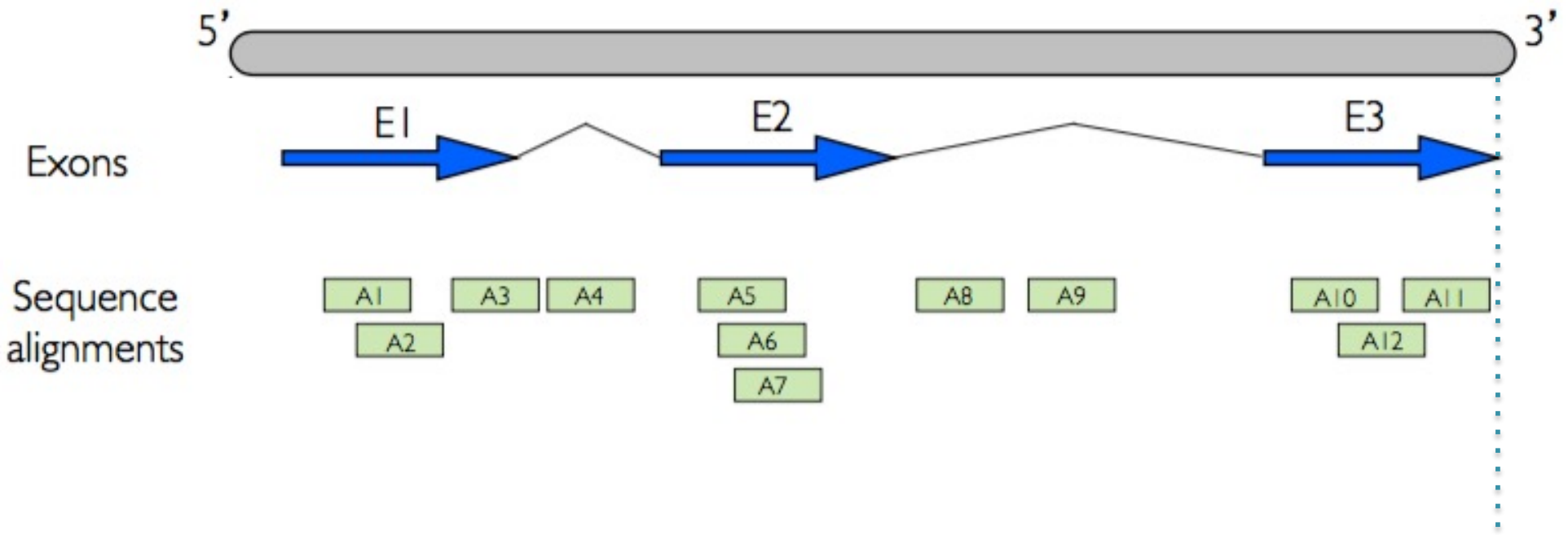
{E2=(A5)}

Plane Sweep to the Rescue!



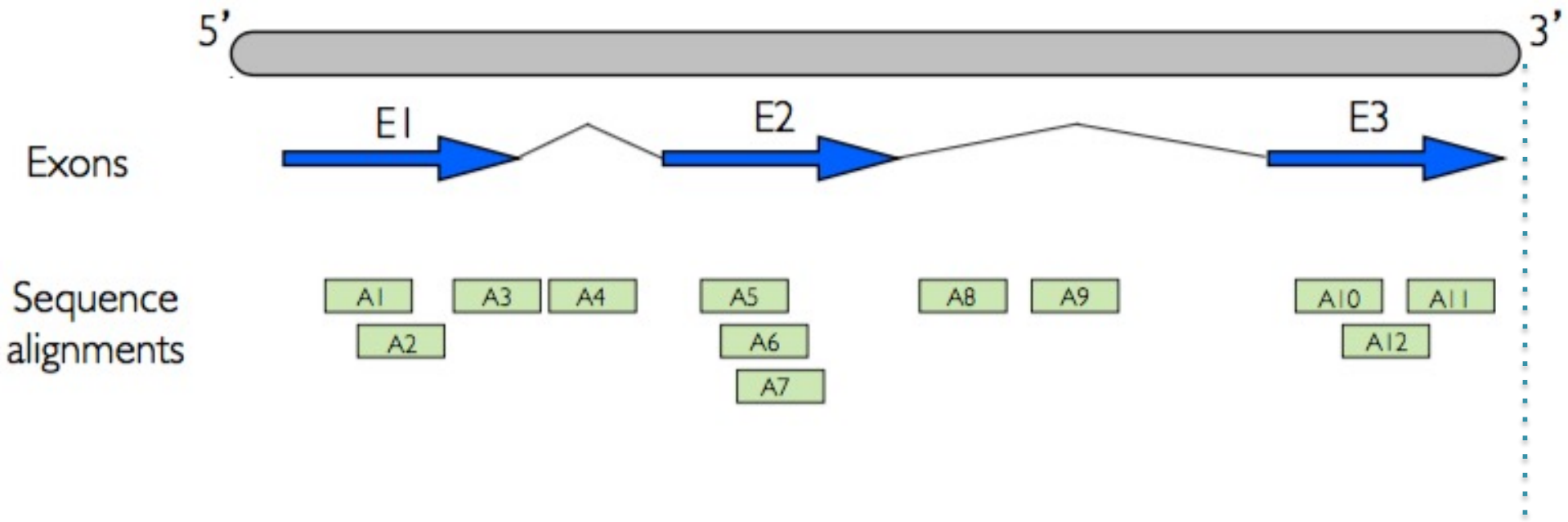
...

Plane Sweep to the Rescue!



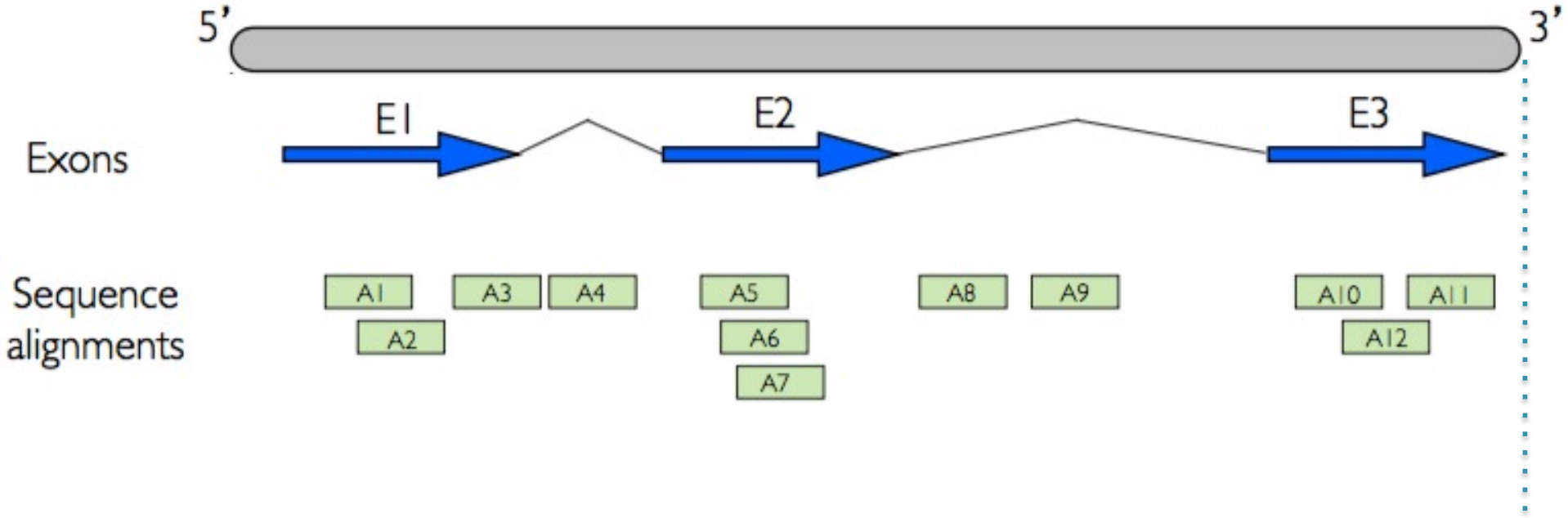
E3 Ends

Plane Sweep to the Rescue!



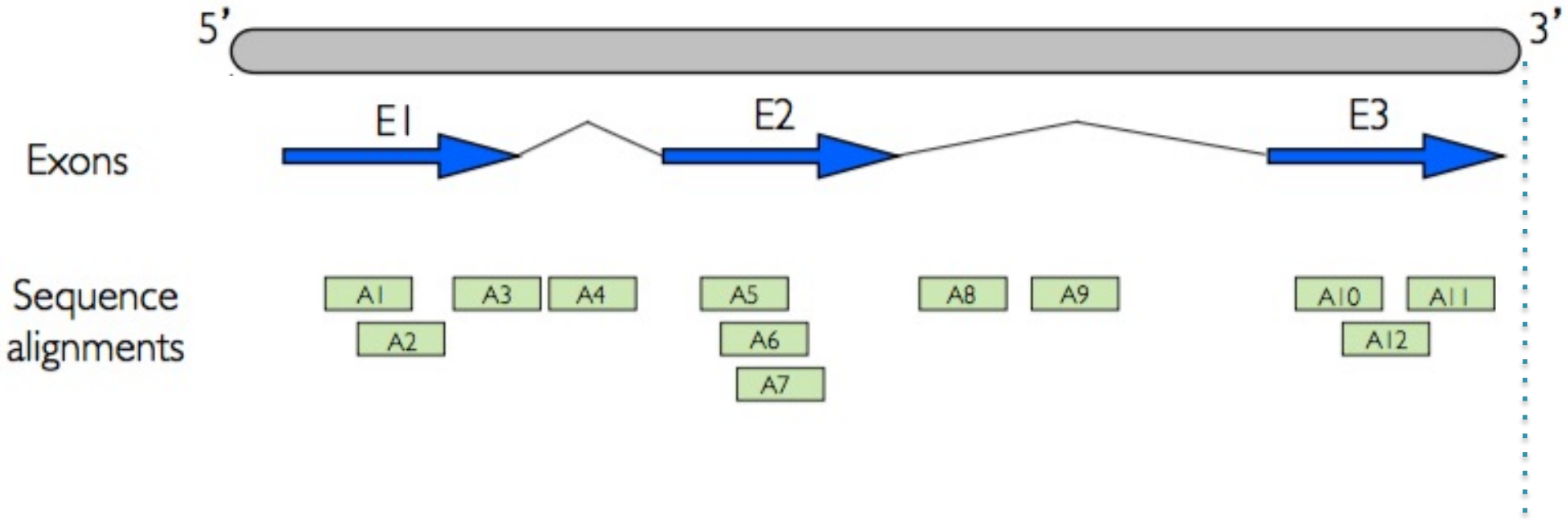
All done!

Plane Sweep to the Rescue!




```
Final Results:

E1=(A1,A2,A3)
E2=(A5,A6,A7)
E3=(A10,A12,A11)
```

Plane Sweep to the Rescue!



How many comparisons does the plane sweep algorithm make?

Each read is compared to the “active set”

Relatively few exons overlap: average ~1.1 active exons/position

Total comparisons: 900M reads * 1.1 “active exons/read” = 990M comparisons ☺

BEDTools commands

```
annotate
bamtobed
bamtofastq
bed12tobed6
bedpetobam
bedtobam
closest
cluster
complement
coverage
expand
flank
fisher
genomecov
getfasta
groupby
groupby
igv
intersect
jaccard
links
makewindows
map
maskfasta
merge
multicov
multiinter
nuc
overlap
pairtobed
pairtopair
random
reldist
shift
shuffle
slop
sort
subtract
tag
unionbedg
window
```

http://bedtools.readthedocs.io/en/latest/content/bedtools-suite.html

Genetic Basis of Autism Spectrum Disorders

Complex disorders of brain development

- Characterized by difficulties in social interaction, verbal and nonverbal communication and repetitive behaviors.
- Have their roots in very early brain development, and the most obvious signs of autism and symptoms of autism tend to emerge between 2 and 3 years of age.

U.S. CDC identify around 1 in 68 American children as on the autism spectrum

- Ten-fold increase in prevalence in 40 years, only partly explained by improved diagnosis and awareness.
- Studies also show that autism is four to five times more common among boys than girls.
- Specific causes remain elusive

What is Autism?
https://autisticadvocacy.org/

Autism is NOT caused by vaccines

Early report

Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith

Summary

Background We investigated a consecutive series of children with chronic enterocolitis and regressive developmental disorder.

Methods 12 children (mean age 6 years [range 3–10], 11 boys) were referred to a paediatric gastroenterology unit with a history of normal development followed by loss of acquired skills, including language, together with diarrhoea and abdominal pain. Children underwent gastroenterological, neurological, and developmental assessment and review of developmental records. Ileocolonoscopy and biopsy sampling, magnetic-resonance imaging (MRI), electroencephalography (EEG), and lumbar puncture were done under sedation. Barium follow-through radiography was done where possible. Biochemical, haematological, and immunological profiles were examined.

Findings Onset of behavioural symptoms was associated by the parents, with measles, mumps, and rubella vaccination in eight of the 12 children, with measles infection in one child, and otitis media in another. All 12 children had intestinal abnormalities ranging from lymphoid nodular hyperplasia to aphthoid ulceration. Histology showed patchy chronic inflammation in 11 children and reactive ileal lymphoid hyperplasia in seven, but no granulomas. Behavioural disorders included autism (nine), disintegrative psychosis (one), and possible postviral or vaccinal encephalitis (two). There were no focal neurological abnormalities and MRI and EEG tests were normal. Abnormal laboratory results were significantly raised urinary methylmalonic acid compared with age-matched controls ($p=0.003$), low haemoglobin in four children, and low serum IgA in four children.

Interpretation We identified associated gastrointestinal disease and developmental regression in a group of previously normal children, which was generally associated in time with possible environmental triggers.

Lancet 1998; **351:** 637–41
See Commentary page

Inflammatory Bowel Disease Study Group, University Departments of Medicine and Histopathology (A J Wakefield FRCS, A Anthony MB, J Linnell PhD, A P Dhillon MRCPath, S E Davies MRCPath) **and the University Departments of Paediatric Gastroenterology** (S H Murch MB, D M Casson MRCP, M Malik MRCP, M A Thomson FRCP, J A Walker-Smith FRCP,), **Child and Adolescent Psychiatry** (M Berelowitz FRCPsych), **Neurology** (P Harvey FRCP), **and Radiology** (A Valentine FRCR), **Royal Free Hospital and School of Medicine, London NW3 2QG, UK**

Correspondence to: Dr A J Wakefield

Introduction

We saw several children who, after a period of apparent normality, lost acquired skills, including communication. They all had gastrointestinal symptoms, including abdominal pain, diarrhoea, and bloating and, in some cases, food intolerance. We describe the clinical findings, and gastrointestinal features of these children.

Patients and methods

12 children, consecutively referred to the department of paediatric gastroenterology with a history of a pervasive developmental disorder with loss of acquired skills and intestinal symptoms (diarrhoea, abdominal pain, bloating and food intolerance), were investigated. All children were admitted to the ward for 1 week, accompanied by their parents.

Clinical investigations

We took histories including details of immunisations and exposure to infectious diseases, and assessed the children. In 11 cases the history was obtained by the senior clinician (JW-S). Neurological and psychiatric assessments were done by consultant staff (PH, MB) with HMS-4 criteria.[1] Developmental studies included a review of prospective developmental records from parents, health visitors, and general practitioners. Four children did not undergo psychiatric assessment in hospital; all had been assessed professionally elsewhere, so these assessments were used as the basis for their behavioural diagnosis.

After bowel preparation, ileocolonoscopy was performed by SHM or MAT under sedation with midazolam and pethidine. Paired frozen and formalin-fixed mucosal biopsy samples were taken from the terminal ileum; ascending, transverse, descending, and sigmoid colons, and from the rectum. The procedure was recorded by video or still images, and were compared with images of the previous seven consecutive paediatric colonoscopies (four normal colonoscopies and three on children with ulcerative colitis), in which the physician reported normal appearances in the terminal ileum. Barium follow-through radiography was possible in some cases.

Also under sedation, cerebral magnetic-resonance imaging (MRI), electroencephalography (EEG) including visual, brain stem auditory, and sensory evoked potentials (where compliance made these possible), and lumbar puncture were done.

Laboratory investigations

Thyroid function, serum long-chain fatty acids, and cerebrospinal-fluid lactate were measured to exclude known causes of childhood neurodegenerative disease. Urinary methylmalonic acid was measured in random urine samples from eight of the 12 children and 14 age-matched and sex-matched normal controls, by a modification of a technique described previously.[2] Chromatograms were scanned digitally on computer, to analyse the methylmalonic-acid zones from cases and controls. Urinary methylmalonic-acid concentrations in patients and controls were compared by a two-sample t test. Urinary creatinine was estimated by routine spectrophotometric assay.

Children were screened for antiendomyseal antibodies and boys were screened for fragile-X if this had not been done

Increasing Exposure to Antibody-Stimulating Proteins and Polysaccharides in Vaccines Is Not Associated with Risk of Autism

Frank DeStefano, MD, MPH[1], Cristofer S. Price, ScM[2], and Eric S. Weintraub, MPH[1]

Objective To evaluate the association between autism and the level of immunologic stimulation received from vaccines administered during the first 2 years of life.

Study design We analyzed data from a case-control study conducted in 3 managed care organizations (MCOs) of 256 children with autism spectrum disorder (ASD) and 752 control children matched on birth year, sex, and MCO. In addition to the broader category of ASD, we also evaluated autistic disorder and ASD with regression. ASD diagnoses were validated through standardized in-person evaluations. Exposure to total antibody-stimulating proteins and polysaccharides from vaccines was determined by summing the antigen content of each vaccine received, as obtained from immunization registries and medical records. Potential confounding factors were ascertained from parent interviews and medical charts. Conditional logistic regression was used to assess associations between ASD outcomes and exposure to antigens in selected time periods.

Results The aOR (95% CI) of ASD associated with each 25-unit increase in total antigen exposure was 0.999 (0.994-1.003) for cumulative exposure to age 3 months, 0.999 (0.997-1.001) for cumulative exposure to age 7 months, and 0.999 (0.998-1.001) for cumulative exposure to age 2 years. Similarly, no increased risk was found for autistic disorder or ASD with regression.

Conclusion In this study of MCO members, increasing exposure to antibody-stimulating proteins and polysaccharides in vaccines during the first 2 years of life was not related to the risk of developing an ASD. *(J Pediatr 2013;163:561-7).*

The initial concerns that vaccines may cause autism were related to the measles, mumps, and rubella vaccine[1] and thimerosal-containing vaccines.[2] In 2004, a comprehensive review by the Institute of Medicine concluded that the evidence favors rejection of possible causal associations between each of these vaccine types and autism.[3] Nonetheless, concerns about a possible link between vaccines and autism persist,[4] with the latest concern centering on the number of vaccines administered to infants and young children.[5] A recent survey found that parents' top vaccine-related concerns included administration of too many vaccines during the first 2 years of life, administration of too many vaccines in a single doctor visit, and a possible link between vaccines and learning disabilities, such as autism.[6] All of the foregoing concerns were reported by 30%-36% of all survey respondents, and were reported by 55%-90% of parents who indicated that their children would receive some, but not all, of the vaccines on the recommended schedule. Another recent survey found that more than 10% of parents of young children refuse or delay vaccinations, with most believing that delaying vaccine doses is safer than providing them in accordance with the Centers for Disease Control and Prevention's recommended vaccination schedule.[7]

Using the number of antibody-stimulating proteins and polysaccharides contained in vaccines as a measure, we evaluated the association between the level of immunologic stimulation received from vaccines during the first 2 years of life and the risk of developing an autism spectrum disorder (ASD), including specific ASD subtypes.

Methods

We performed a secondary analysis of publicly available data from a case-control study designed to examine potential associations between exposure to thimerosal-containing injections and ASD.[8] The study was conducted in 3 managed care organizations (MCOs). Data sources for the original study included MCO computerized data files, abstraction of biological mothers' and children's medical charts, and standardized telephone interviews with biological mothers. Case children underwent standardized in-person assessment to verify case status.

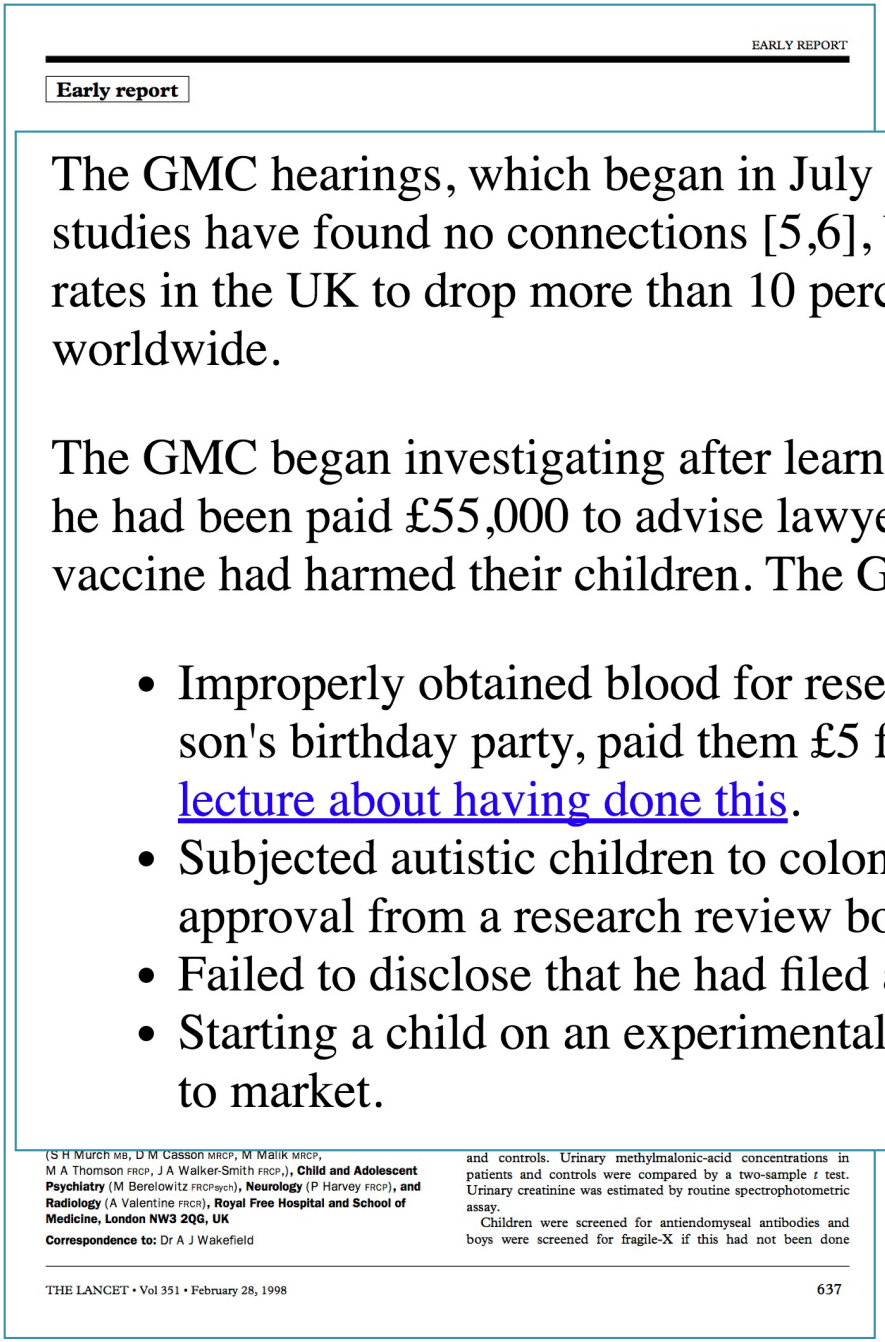
AD	Autistic disorder
ADI-R	Autism Diagnostic Interview–Revised
ADOS	Autism Diagnostic Observation Schedule
ASD	Autism spectrum disorder
MCO	Managed care organization
SCQ	Social Communication Questionnaire

From the [1]Immunization Safety Office, Centers for Disease Control and Prevention, Atlanta, GA and [2]Abt Associates Inc, Bethesda, MD

Funded by a contract from the Centers for Disease Control and Prevention to America's Health Insurance Plans (AHIP), and by subcontracts from AHIP to Abt Associates, Inc. The findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. The authors declare no conflicts of interest.

 http://dx.doi.org/10.1016/j.jpeds.2013.02.001

Autism is NOT caused by vaccines



EARLY REPORT

Early report

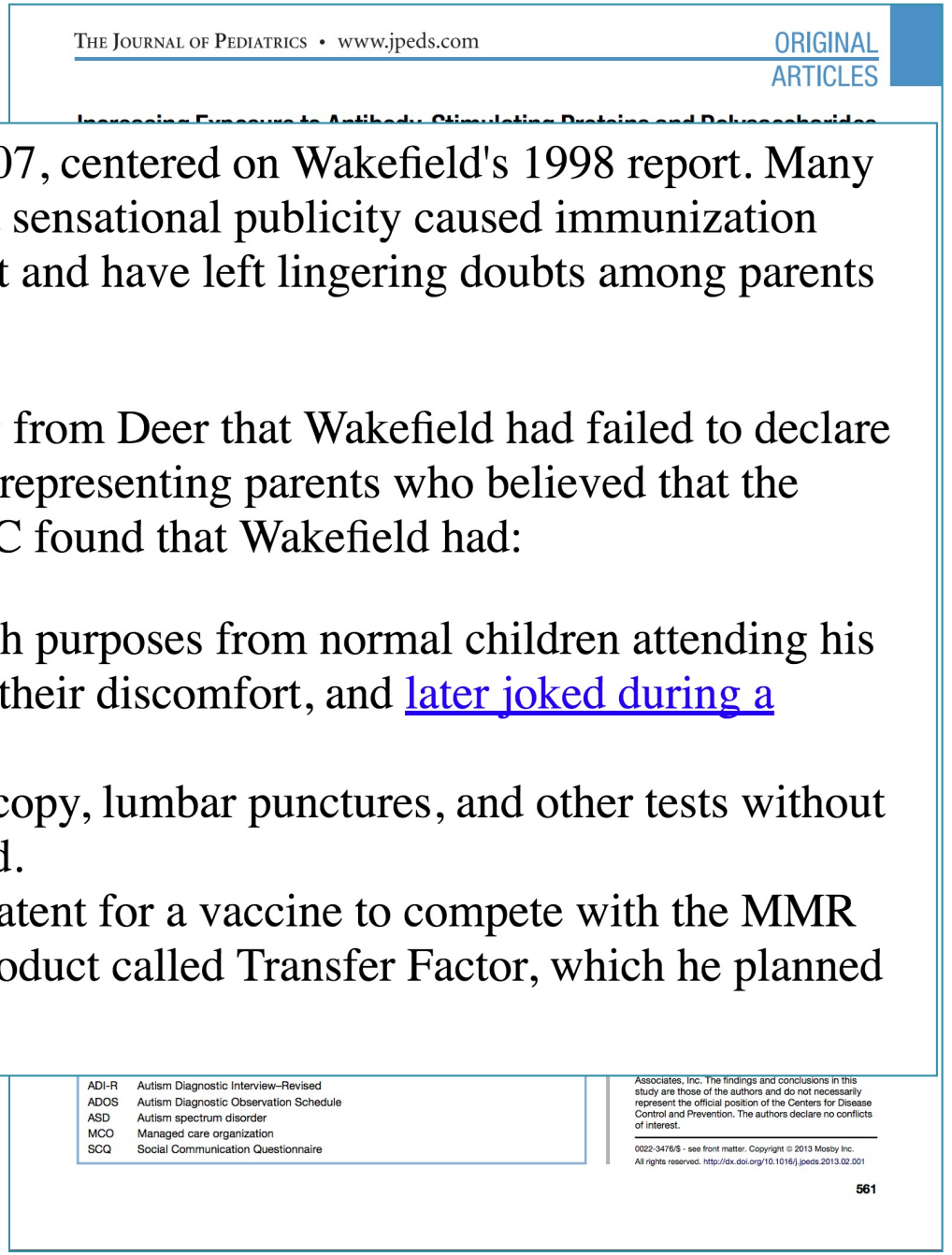
(S H Murch MB, D M Casson MRCP, M Malik MRCP, M A Thomson FRCP, J A Walker-Smith FRCP,), **Child and Adolescent Psychiatry** (M Berelowitz FRCPsych), **Neurology** (P Harvey FRCP), **and Radiology** (A Valentine FRCR), **Royal Free Hospital and School of Medicine, London NW3 2QG, UK**

Correspondence to: Dr A J Wakefield

and controls. Urinary methylmalonic-acid concentrations in patients and controls were compared by a two-sample *t* test. Urinary creatinine was estimated by routine spectrophotometric assay.

Children were screened for antiendomyseal antibodies and boys were screened for fragile-X if this had not been done

THE LANCET • Vol 351 • February 28, 1998 637



THE JOURNAL OF PEDIATRICS • www.jpeds.com ORIGINAL ARTICLES

ADI-R	Autism Diagnostic Interview–Revised
ADOS	Autism Diagnostic Observation Schedule
ASD	Autism spectrum disorder
MCO	Managed care organization
SCQ	Social Communication Questionnaire

Associates, Inc. The findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. The authors declare no conflicts of interest.

561

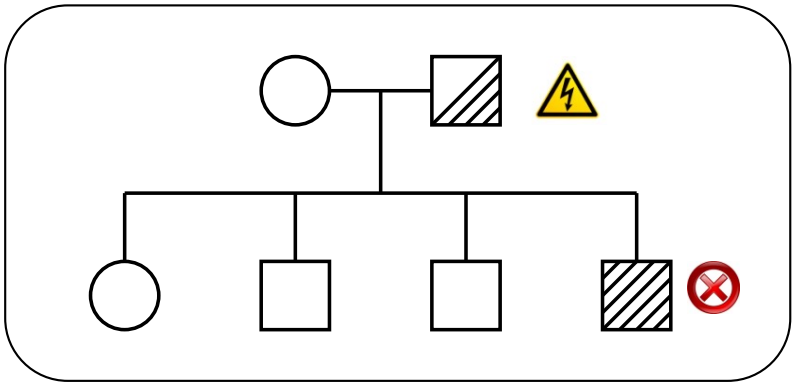
The GMC hearings, which began in July 2007, centered on Wakefield's 1998 report. Many studies have found no connections [5,6], but sensational publicity caused immunization rates in the UK to drop more than 10 percent and have left lingering doubts among parents worldwide.

The GMC began investigating after learning from Deer that Wakefield had failed to declare he had been paid £55,000 to advise lawyers representing parents who believed that the vaccine had harmed their children. The GMC found that Wakefield had:

- Improperly obtained blood for research purposes from normal children attending his son's birthday party, paid them £5 for their discomfort, and later joked during a lecture about having done this.
- Subjected autistic children to colonoscopy, lumbar punctures, and other tests without approval from a research review board.
- Failed to disclose that he had filed a patent for a vaccine to compete with the MMR
- Starting a child on an experimental product called Transfer Factor, which he planned to market.

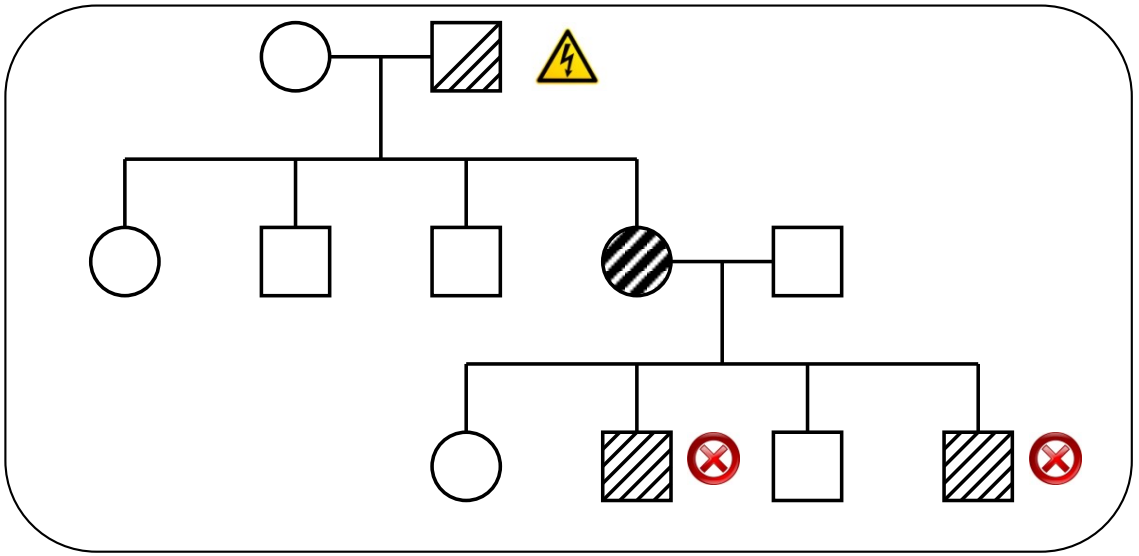
Unified Model of Autism

Sporadic Autism



De novo mutations of high penetrance contributes to autism, especially in low risk families with no history of autism.

Familial Autism



Legend


Sporadic mutation

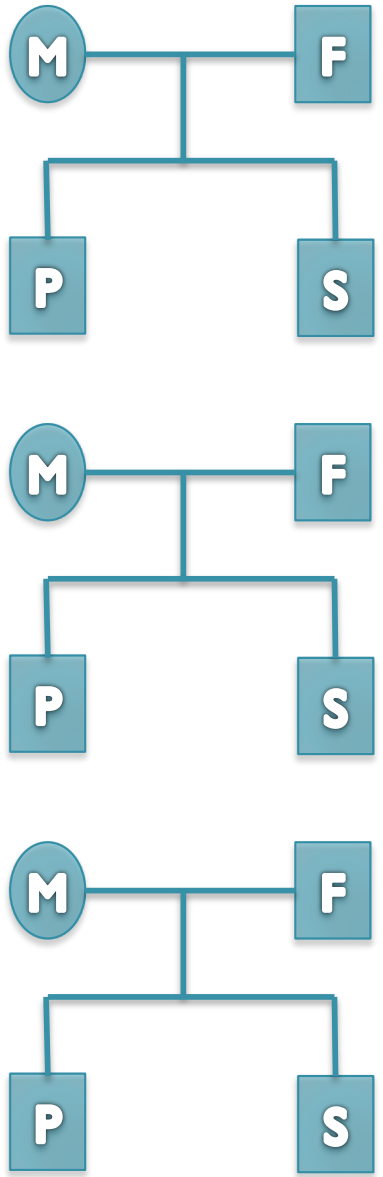

Fails to procreate

A unified genetic theory for sporadic and inherited autism
Zhao *et al.* (2007) *PNAS. 104(31)12831-12836.*

Searching for the genetic risk factors

Search Strategy

- Thousands of families identified from a dozen hospitals around the United States
- Large scale genome sequencing of "simplex" families: mother, father, affected child, unaffected sibling
- Unaffected siblings provide a natural control for environmental factors

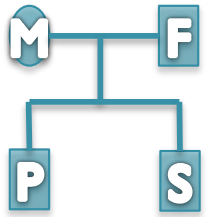


Are there any genetic variants present in affected children, that are not in their parents or unaffected siblings?

SFARI SIMONS FOUNDATION AUTISM RESEARCH INITIATIVE

De novo mutation discovery and validation

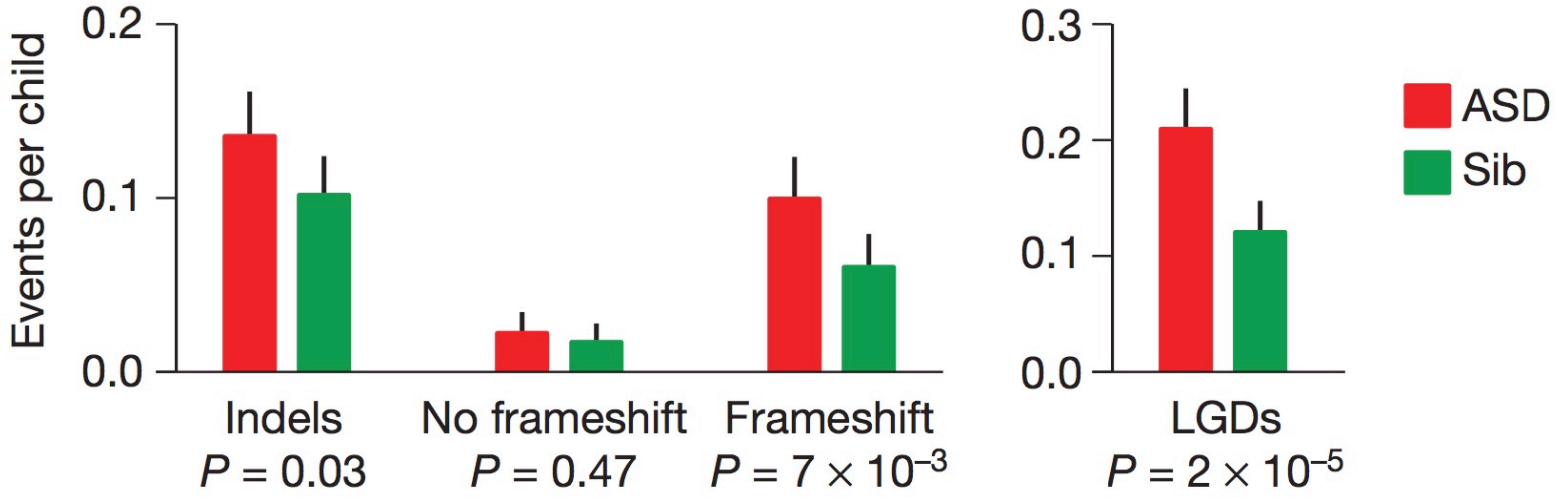
De novo mutations:
Sequences not inherited from your parents.




```
Reference:   ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...

Father(1):   ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...
Father(2):   ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...

Mother(1):   ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...
Mother(2):   ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...

Sibling(1):  ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...
Sibling(2):  ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...

Proband(1):  ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...
Proband(2):  ...TCAAATCCTTTTAAT****AAGAGCTGACA...
```

4bp heterozygous deletion at chr15:93524061 CHD2

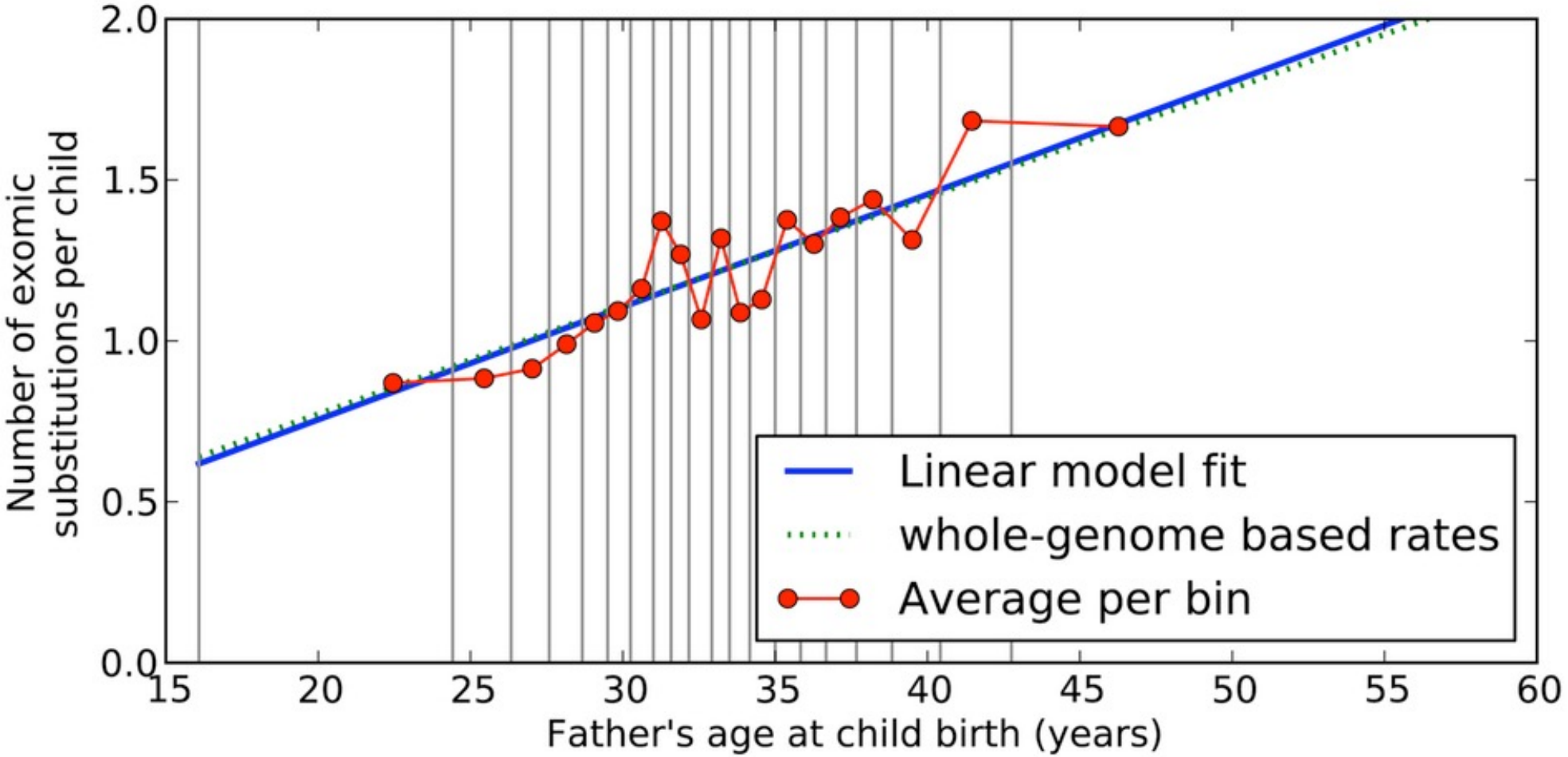
De novo Genetics of Autism



- In 2,500 family quads we see significant enrichment in de novo ***likely gene disruptions (LGDs)*** in the autistic children
  - Overall rate of de novo mutations basically 1:1
  - 2:1 enrichment in frameshift indels, nonsense mutations
  - Contributed dozens of new autism candidate genes, highly enriched for neuron development or chromatin modifiers

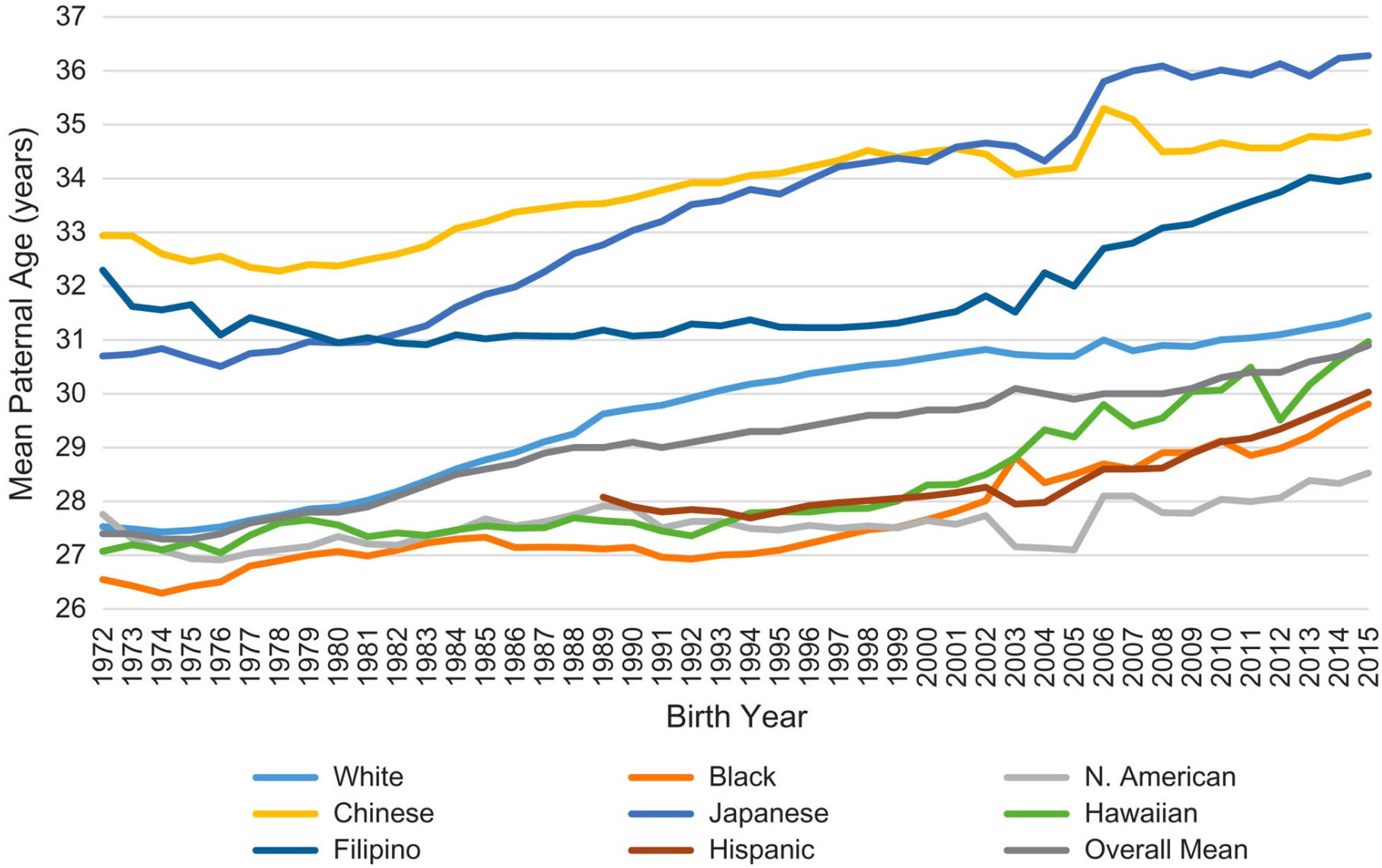
The contribution of de novo coding mutations to autism spectrum disorder
Iossifov *et al* (2014) *Nature*. doi:10.1038/nature13908

De novo Mutations in Men



The contribution of de novo coding mutations to autism spectrum disorder
Iossifov *et al* (2014) *Nature*. doi:10.1038/nature13908

Age of Fatherhood



The age of fathers in the USA is rising: an analysis of 168 867 480 births from 1972 to 2015
Khandwala *et al* (2017) *Human Reproduction. https://doi.org/10.1093/humrep/dex267*