CSC 334: TOPICS IN COMPUTATIONAL BIOLOGY

"Algorithms for Genomic Data"

Fall 2015

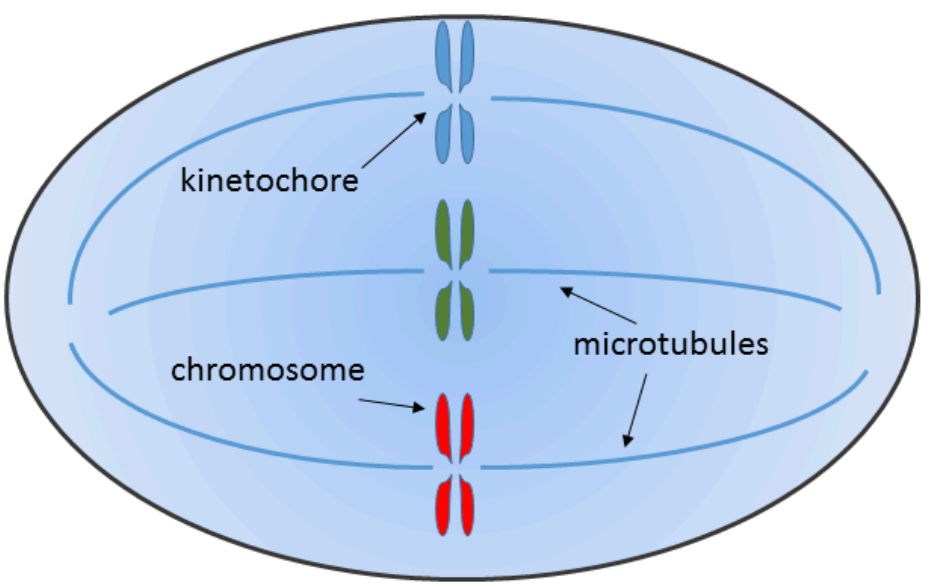
Smith College

Instructor: Prof. Sara Sheehan

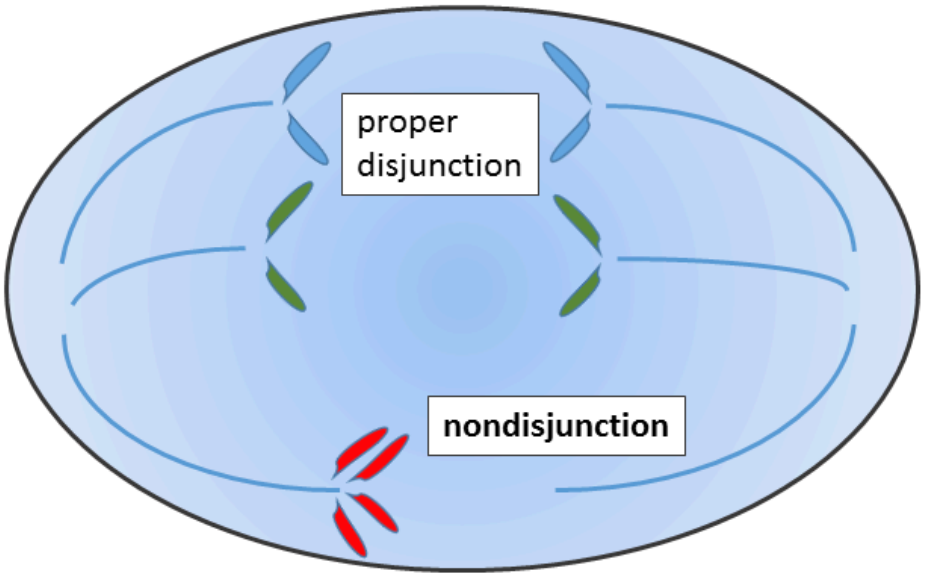
Prenatal Testing Background

Causes of aneuploidy

↑ abnormal number of chromosomes



metaphase



anaphase

"Mitotic nondisjunction" by Wpeissner - https://commons.wikimedia.org

Motivation: aneuploidy testing

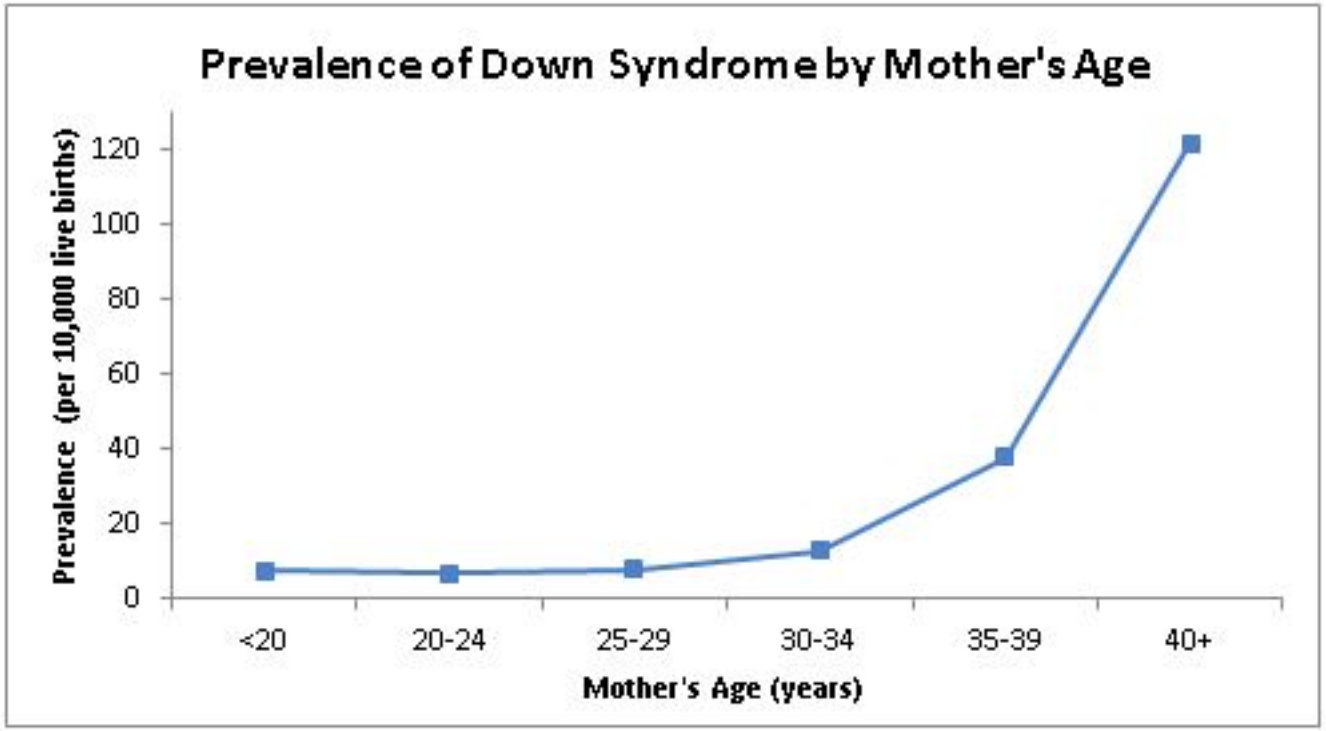
Autosomal chromosomes:

- Trisomy 13: Patau syndrome
- Trisomy 18: Edwards syndrome
- Trisomy 21: Down syndrome

Sex chromosomes:

- X0 (one X chromosome): Turner's syndrome
- XYY or XXX: normal male or female phenotype
- XXY: Klinefelter Syndrome

Older women are more at risk



Motivation: aneuploidy testing

- Historically common approach: amniocentesis (invasive)
  - 16-22 weeks
  - serum biochemical assays used as well
- 1981: proof of fetal DNA in maternal blood
  - Y chromosome detected with fluorescence
- ≈2011: companies offer non-invasive prenatal diagnosis
  - Verinata
  - Harmony
  - NIFTY
  - All using next-generation sequencing

Basic Procedure

1) Sample maternal blood

- contains cell-free fetal DNA (cffDNA)

2) Low-coverage sequencing (0.1x - 4x)

3) Read alignment/mapping to human genome

- save reads ("tags") that map uniquely

4) Compute coverage for each chromosome

5) *t*-test for aneuploidy (coverage differences)

Related Work

verinata

Continuous innovation

Increased safety and peace of mind for your patients

Swift acceptance of the verifi® prenatal test has made a world of difference to high-risk patients across the country:

- SAFE—Routine blood draw, **just one tube** (7-10 ml)
- ACCURATE—Directly analyzes cell-free fetal DNA with our proprietary SAFeR™ algorithm
- EASY—Test as early as 10 weeks, no limitations in reference to patient ethnicity, BMI, ART, or egg donor cases
- FAST—Results reported in 3-6 business days after sample receipt

The basic verifi® test detects:

- T21 (Down syndrome)
- T18 (Edwards syndrome)
- T13 (Patau syndrome)

→ Trisomy 21, 18, 13

Now a wider option is available for sex chromosomes at no extra charge:

- Monosomy X (MX; Turner syndrome)
- XXX (Triple X)
- XXY (Klinefelter syndrome)
- XYY (Jacobs syndrome)
- Fetal sex (XX or XY)—aids in stratifying the risk for X-linked disorders such as hemophilia, Duchenne muscular dystrophy, or cases of ambiguous genitalia, such as congenital adrenal hyperplasia

verinata

t-test comparing chroms 1 and 2:

coverage difference between two chromosomes

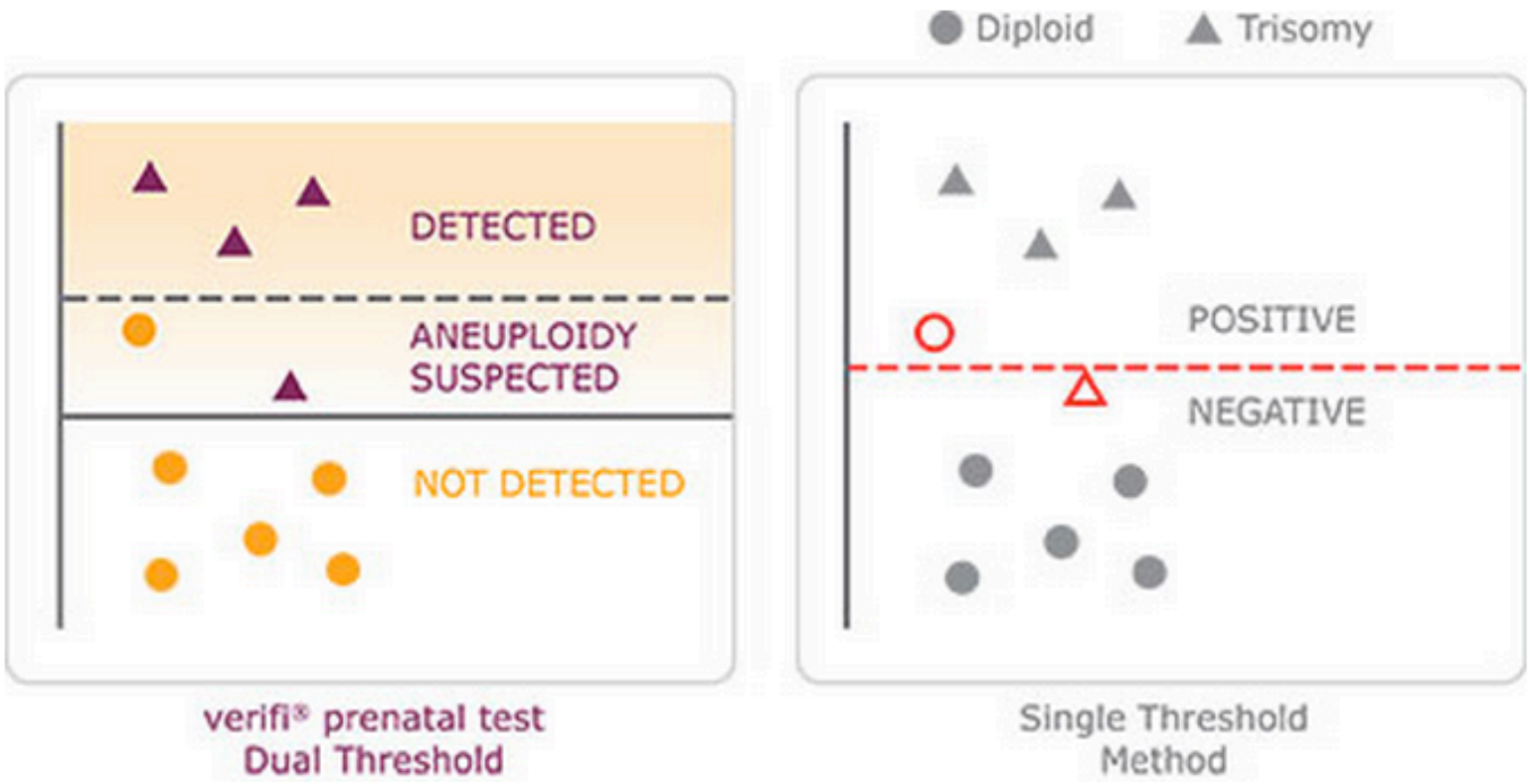
$$t = \frac{\overline{y_2} - \overline{y_1}}{\sqrt{\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}}}$$

standard deviation

number of windows

threshold for reporting aneuploidy: |t| > 3.09
(99% confidence interval)

verinata


Diploid
Trisomy
DETECTED
ANEUPLOIDY SUSPECTED
NOT DETECTED
POSITIVE
NEGATIVE
verifi® prenatal test
Dual Threshold
Single Threshold
Method

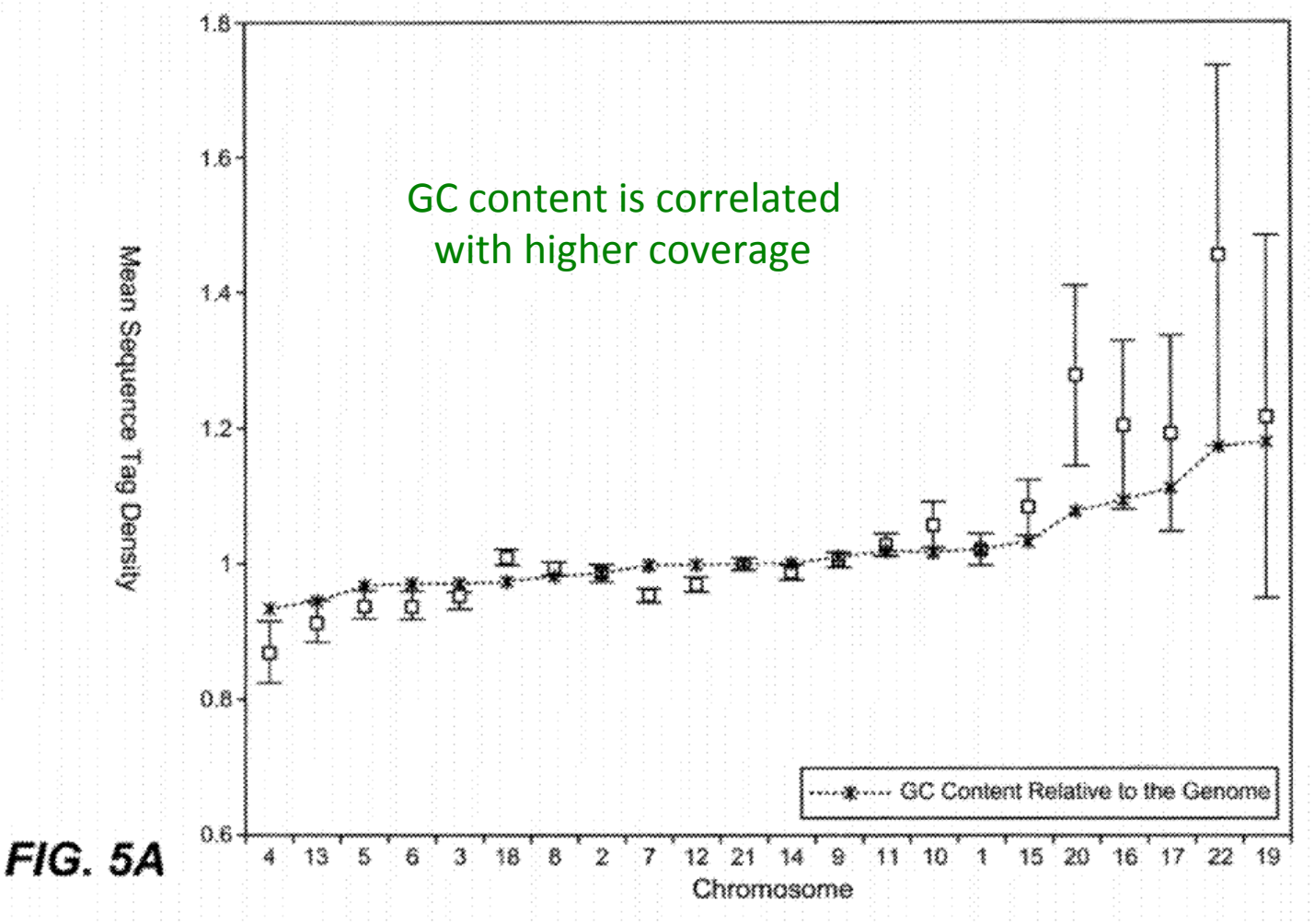

GC content is correlated
with higher coverage
Mean Sequence Tag Density
1.8
1.6
1.4
1.2
1
0.8
0.6
4 13 5 6 3 18 8 2 7 12 21 14 9 11 10 1 15 20 16 17 22 19
Chromosome
GC Content Relative to the Genome

FIG. 5A

verinata

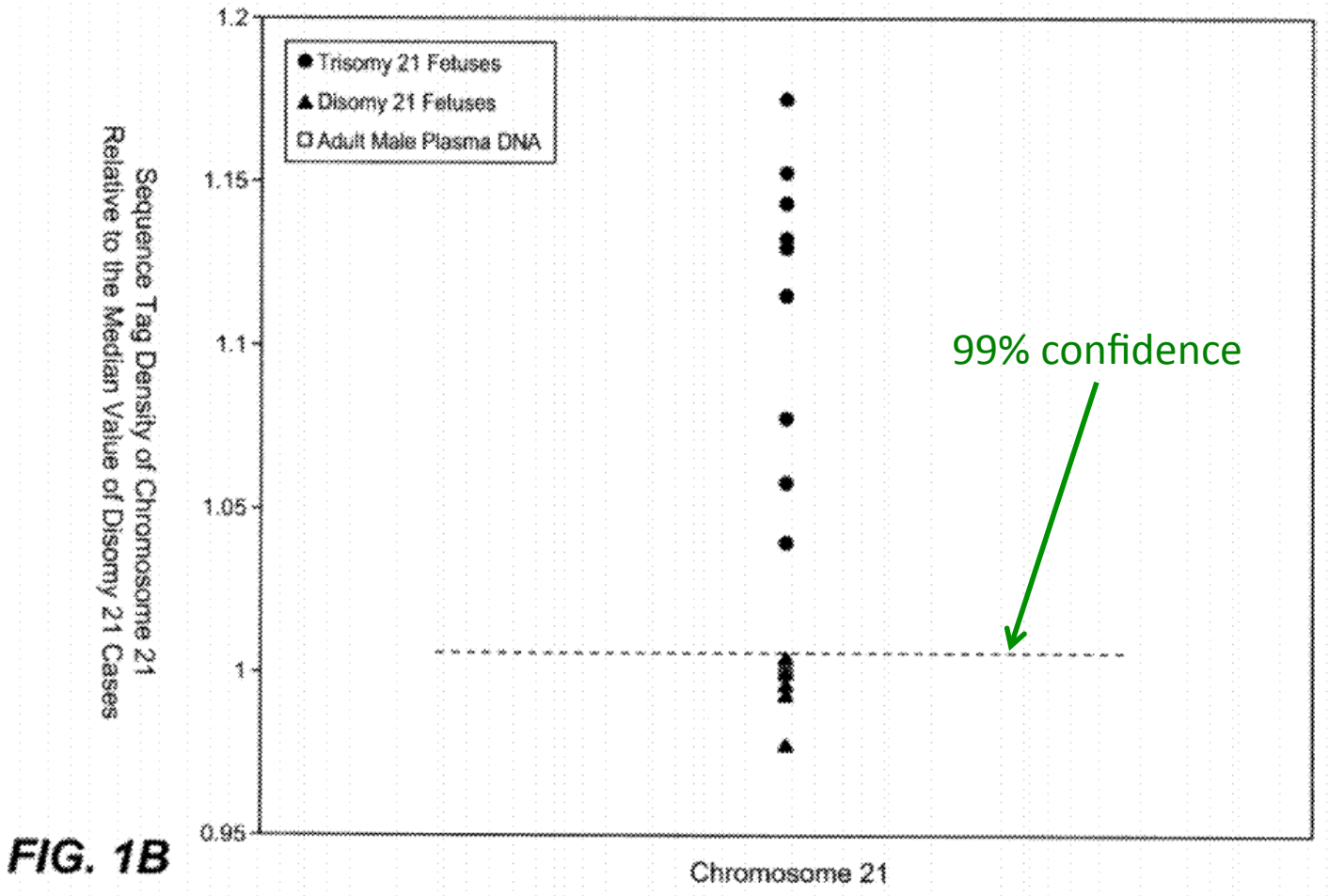


FIG. 1B

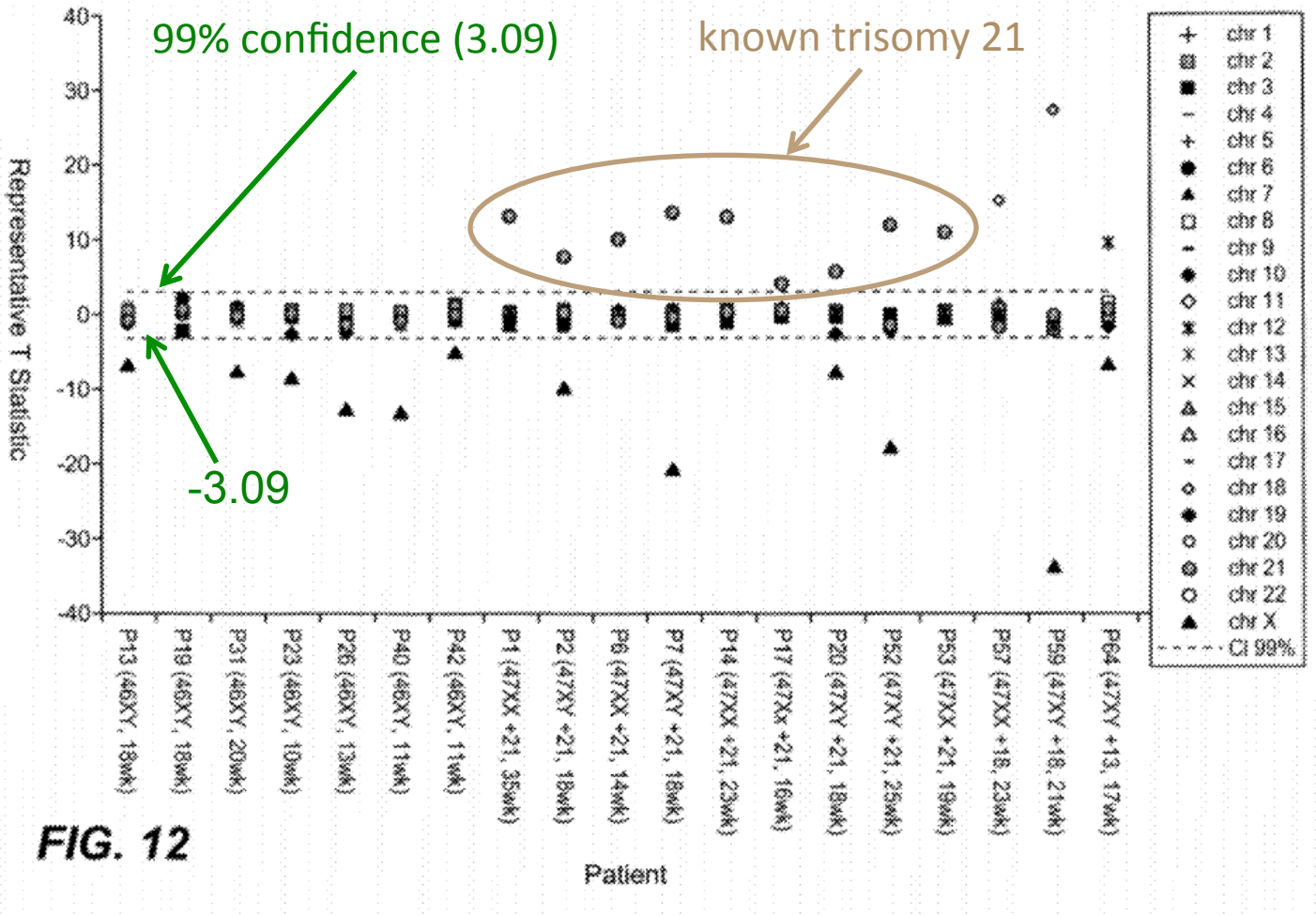

99% confidence (3.09)
known trisomy 21
-3.09
Representative T Statistic
40
30
20
10
0
-10
-20
-30
-40
P13 (46XY, 19wk)
P19 (46XY, 18wk)
P31 (46XY, 20wk)
P23 (46XY, 10wk)
P26 (46XY, 13wk)
P40 (46XY, 11wk)
P42 (46XY, 11wk)
P1 (47XX +21, 35wk)
P2 (47XY +21, 18wk)
P6 (47XX +21, 14wk)
P7 (47XY +21, 18wk)
P14 (47XX +21, 23wk)
P17 (47Xx +21, 16wk)
P20 (47XY +21, 18wk)
P52 (47XY +21, 25wk)
P53 (47XX +21, 19wk)
P57 (47XX +18, 23wk)
P59 (47XY +18, 21wk)
P64 (47XY +13, 17wk)
Patient
chr 1
chr 2
chr 3
chr 4
chr 5
chr 6
chr 7
chr 8
chr 9
chr 10
chr 11
chr 12
chr 13
chr 14
chr 15
chr 16
chr 17
chr 18
chr 19
chr 20
chr 21
chr 22
chr X
CI 99%

FIG. 12

verinata

DNA Sequencing versus Standard Prenatal Aneuploidy Screening

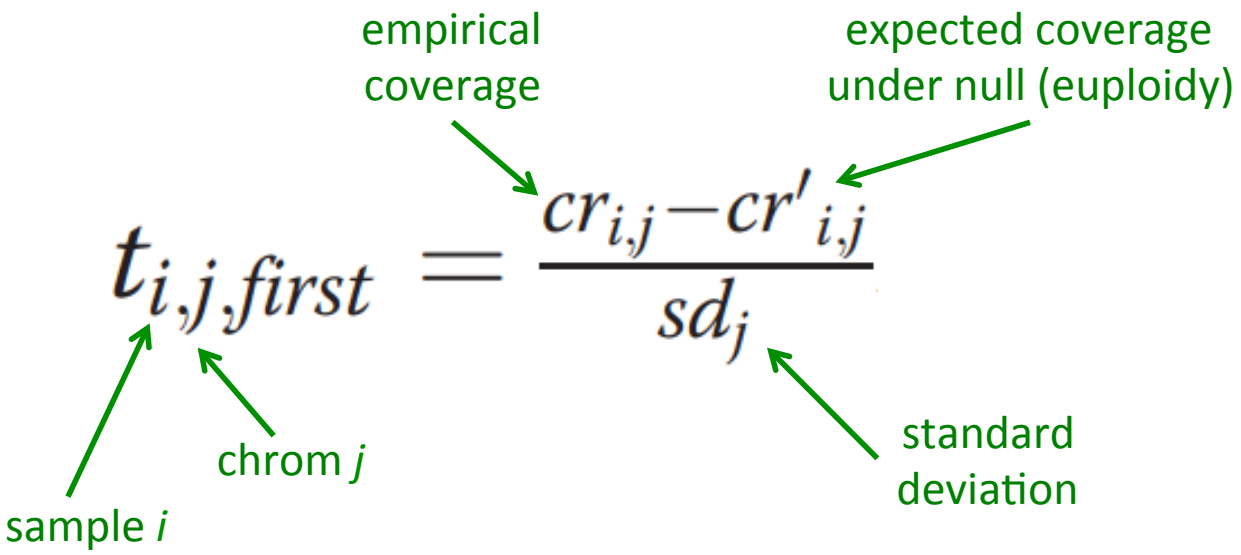
Diana W. Bianchi, M.D., R. Lamar Parker, M.D., Jeffrey Wentworth, M.D., Rajeevi Madankumar, M.D., Craig Saffer, M.D., Anita F. Das, Ph.D., Joseph A. Craig, M.D., Darya I. Chudova, Ph.D., Patricia L. Devers, M.S., C.G.C., Keith W. Jones, Ph.D., Kelly Oliver, B.S., Richard P. Rava, Ph.D., and Amy J. Sehnert, M.D. for the CARE Study Group

N Engl J Med 2014; 370:799-808 | February 27, 2014 | DOI: 10.1056/NEJMoa1311037

	false positive (sequencing)	false positive (standard)	false negative (sequencing)	num positives	PPV (sequencing)	PPV (standard)
T21	0.30%	3.60%	0	5	45.50%	4.20%
T18	0.20%	0.60%	0	2	40.00%	8.30%
T13			0	1		

NIFTY

- *t*-test
- critical threshold: $|t_{i,j,first}| > 3$



"Noninvasive Fetal Trisomy (NIFTY) test: an advanced noninvasive prenatal diagnosis methodology for fetal autosomal and sex chromosomal aneuploidies."

Jiang et al. *BMC Medical Genomics* 2012, 5:57

NIFTY

Non-invasive prenatal testing for fetal chromosomal abnormalities by low-coverage whole-genome sequencing of maternal plasma DNA: review of 1982 consecutive cases in a single center

T. K. LAU*, S. W. CHEUNG†, P. S. S. LO*, A. N. PURSLEY†, M. K. CHAN‡, F. JIANG‡, H. ZHANG‡, W. WANG‡, L. F. J. JONG*, O. K. C. YUEN*, H. Y. C. CHAN*, W. S. K. CHAN* and K. W. CHOY§¶

**Fetal Medicine Centre, Paramount Medical Centre, Hong Kong, China; †Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; ‡Clinical Laboratory of BGI Health, Shenzhen, China; §Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong, China; ¶Shenzhen Research Institute, The Chinese University of Hong Kong, China*

Chr	*Sensitivity*	*Specificity*
21	23/23 (100 (85.7–100))	1959/1959 (100 (99.8–100))
18	4/4 (100 (51.0–100))	1978/1978 (100 (99.8–100))
13	2/2 (100 (34.2–100))	1980/1980 (100 (99.8–100))

Values are given as no. positive/total no. (% (95% CI)).
Chr, chromosome.

Bayesian Model

Clinical Trials Example

- Disease affects 1/100 people: $P(\text{disease}) = 0.01$
- Test for the disease with 90% accuracy
  - $P(\text{positive} \mid \text{disease}) = 0.9$
  - $P(\text{negative} \mid \text{healthy}) = 0.9$

$$P(\text{disease}|\text{positive}) = \frac{P(\text{positive}|\text{disease})\, P(\text{disease})}{P(\text{positive})}$$

$$\approx 8.3\%$$

Bayesian Model

Input data are read counts for each chromosome (1,2,…,n):

$$q_1, q_2, \cdots, q_n = \vec{q} \qquad \sum_{i=1}^{n} q_i = N$$

Goal:

$$\mathbb{P}(T_{21}|\vec{q}\,) = \frac{\mathbb{P}(\vec{q}\,|T_{21}) \cdot \mathbb{P}(T_{21})}{\mathbb{P}(\vec{q}\,)}$$

$$= \frac{\mathbb{P}(\vec{q}\,|T_{21}) \cdot \mathbb{P}(T_{21})}{\mathbb{P}(\vec{q}\,|T_{21}) \cdot \mathbb{P}(T_{21}) + \mathbb{P}(\vec{q}\,|T_{21}^{C}) \cdot \mathbb{P}(T_{21}^{C})}$$

Prior:

$P(T_{21})$

Maternal Age	Trisomy 21	All Trisomies
20	1 in 1,667	1 in 526
21	1 in 1,429	1 in 526
22	1 in 1,429	1 in 500
23	1 in 1,429	1 in 500
24	1 in 1,250	1 in 476
25	1 in 1,250	1 in 476
26	1 in 1,176	1 in 476
27	1 in 1,111	1 in 455
28	1 in 1,053	1 in 435
29	1 in 1,000	1 in 417
30	1 in 952	1 in 384
31	1 in 909	1 in 384
32	1 in 769	1 in 323
33	1 in 625	1 in 286
34	1 in 500	1 in 238
35	1 in 385	1 in 192
36	1 in 294	1 in 156
37	1 in 227	1 in 127
38	1 in 175	1 in 102
39	1 in 137	1 in 83
40	1 in 106	1 in 66
41	1 in 82	1 in 53
42	1 in 64	1 in 42
43	1 in 50	1 in 33
44	1 in 38	1 in 26
45	1 in 30	1 in 21
46	1 in 23	1 in 16
47	1 in 18	1 in 13
48	1 in 14	1 in 10
49	1 in 11	1 in 8

Estimating ε

(fetal DNA fraction)

- Use Y chromosome
- If trisomy suspected, use extra chrom counts
- Centrifuge maternal DNA
- Look at polymorphism
- Use a constant or don't rely on ε