

# Second-Generation *p*-values: Introduction and Applications

---

Jeffrey D. Blume, PhD

Departments of Biostatistics,  
Biomedical Informatics

VANDERBILT  UNIVERSITY

- Fix Footnotes!!!!
- Revise outline
- Add SGPV package demonstration
- Add slide on suggested use of p-value and CIs...this is essentially the SPGV from ASA statement.

## Synopsis

- Classical  $p$ -values are
  - Ubiquitous, Sacrosanct, Imperfect, Misused
  - Misunderstood (Significance vs. Hypothesis testing roles)
  - Openly debated in practice and theory
- Trend toward estimation in reporting of results
  - Report a estimation interval (e.g. confidence interval)
  - Does interval contain only clinically significant values?
- Second-generation  $p$ -values (SGPVs)
  - Embody and formalize this trend
  - Maintain and improve error rate control
  - Define clinically significant before looking at the data

Second-generation  $p$ -values

Blume, Nov 2019

## Outline

- Evidential Metrics
- Second-generation  $p$ -value
- Introductory examples (4)
- High-dimensional example, 7128 Genes
  - $\alpha=0.05$  vs  $\alpha=0.05/7128$  vs SG  $p$ -value
- High-dimensional example
  - Prostate Cancer SNP data (~247,000)
  - 3,894 subjects: 2,511 cases & 1,383 controls
- Outrageous claim

Second-generation  $p$ -values

Blume, Nov 2019

## Evidential metrics

Example:  
Diagnostic Test

### 1. Measure of the strength evidence

- Axiomatic and intuitive justification
- Summary statistic, yardstick

Positive Test  
Negative Test

### 2. Propensity to collect data that will yield a misleading #1

- Error rates
- Properties of the study design (!)

Sensitivity  
Specificity

### 3. Probability that an observed #1 is misleading

- False Discovery rate, False Confirmation rate
- Chance that an observed result is mistaken
- Properties of the observed data (!)

PPV  
NPV

Second-generation p-values

Blume, Nov 2019

## Testing

Evidential Metric	What it measures	Hypothesis Testing	Significance Testing
1	strength of the evidence	Absent	Tail-area probability ( <i>p</i> -value)
2	propensity for study to yield misleading evidence	Tail-area probability (error rates)	Absent
3	propensity for observed results to be misleading	misinterpret #2	misinterpret #1

- The *tail-area probability* is used to measure *three* distinct metrics

Second-generation p-values

Blume, Nov 2019

## Second-generation *p*-value

- [StatisticalEvidence.com](http://StatisticalEvidence.com)
- Examine statistical properties in module 2
- Retains strict error control

Evidential Metric	What it measures	SGPV
1	Summary measure	$p_\delta$
2	Operating characteristics	$P(p_\delta = 0   H_0)$ $P(p_\delta = 1   H_1)$ $P(0 < p_\delta < 1   H)$
3	False discovery rates	$P(H_0   p_\delta = 0)$ $P(H_1   p_\delta = 1)$

Second-generation *p*-values

Blume, Nov 2019

## The *p*-value (what it is)

- Number between 0 and 1
- Smaller  $\Rightarrow$  support for an alternative hypothesis
- Larger  $\Rightarrow$  data are inconclusive
- Clinical significance is ignored
- Sample size confounds comparisons
- Interpretation
  - awkward
  - assumes null hypothesis true
  - rooted in inductive reasoning
- Not clear if/when ‘adjustments’ are necessary

Second-generation *p*-values

Blume, Nov 2019

*2<sup>nd</sup>-generation*  
**The  $\hat{p}$ -value (what it is)**

*Version 2.0*

- ✓ Number between 0 and 1 → near 0 supports alt  
near 1 supports null  
near  $\frac{1}{2}$  inconclusive
- ✓ Smaller ⇒ support for an alternative hypothesis
  - Larger ⇒ data ~~are inconclusive~~ support null
  - Clinical significance is ~~ignored~~ incorporated
- ✗ Sample size confounds comparisons
  - Interpretation → Fraction of data-supported hypotheses that are null
    - ~~awkward~~ straightforward
    - assumes ~~null hypothesis true~~ conditions on observed data
    - ~~rooted in inductive reasoning~~ descriptive, summarizes
  - ~~Not~~ clear if/when ‘adjustments’ are necessary

Ideally, never

Second-generation p-values      Blume, Nov 2019

**Illustration**

The diagram illustrates two types of null hypotheses: Point Null Hypothesis and Interval Null Hypothesis.

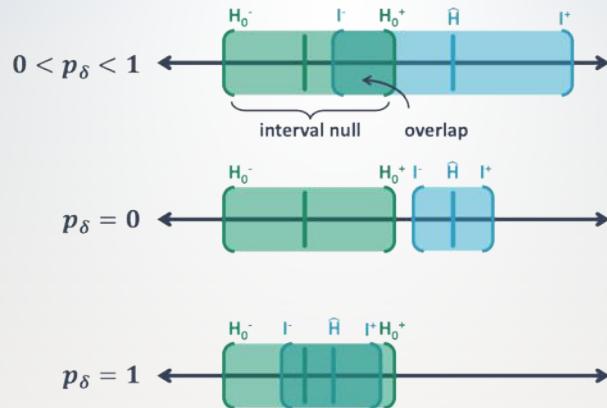
**Point Null Hypothesis:** A single vertical line labeled  $H_0$  on a horizontal axis. A light blue shaded interval labeled  $\hat{H}$  is centered around  $H_0$ . A bracket below the axis is labeled "confidence interval".

**Interval Null Hypothesis:** Two vertical lines labeled  $H_0^-$  and  $H_0^+$  define a central interval. A light blue shaded interval labeled  $CI^-$  and  $CI^+$  is centered around this central interval. A blue bracket below the axis is labeled "overlap". A blue arrow labeled  $\delta$  points to the distance from the central interval to the boundaries of the confidence interval.

Point null hypothesis  $H_0$  and interval null hypothesis  $[H_0^-, H_0^+]$   
Data-supported hypothesis  $\hat{H}$  and confidence interval  $[CI^-, CI^+]$

Second-generation p-values      Blume, Nov 2019

## Illustration



Works with confidence, credible, and support intervals

Second-generation p-values

Blume, Nov 2019

## Second-generation p-value



- SGPV is in  $[0,1]$  and denoted by  $p_\delta$
- $\delta$  for scientific significance
  - $p_\delta = 0 \Rightarrow$  null **incompatible** with data
  - $p_\delta = 1 \Rightarrow$  null **compatible** with data
  - $0 < p_\delta < 1 \Rightarrow$  data are **inconclusive**
- Fraction of data-supported hypotheses that are null
- Retains strict error control, all rates  $\rightarrow 0$

Second-generation p-values

Blume, Nov 2019

## Definition

**Second-generation  
p-value (SGPV)**

$$p_\delta = \frac{|I \cap H_0|}{|I|} \times \max\left\{\frac{|I|}{2|H_0|}, 1\right\}$$

**Proportion** of data-supported hypotheses that are also null hypotheses

**Small-sample correction factor**  
shrinks proportion to  $\frac{1}{2}$  when  $|I|$  wide

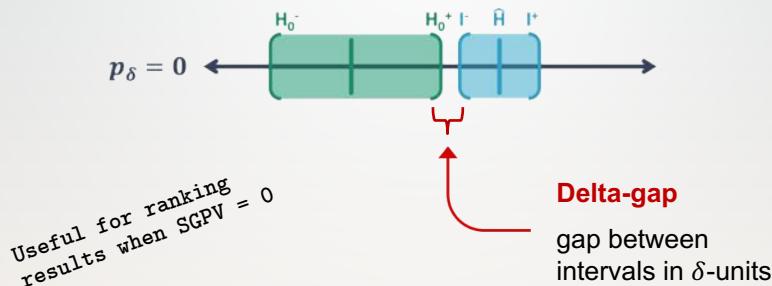
when  $|I| > 2|H_0|$

Second-generation p-values

Blume, Nov 2019

## The Delta-gap

When SGPV=0, there is a gap between the intervals. The length of that gap, in  $\delta$ -units is the **delta-gap**.



Second-generation p-values

Blume, Nov 2019

## Steps

- SGPV ~ the fraction of data-supported hypotheses that are null or practically null
1. Specify an the interval null hypothesis or a point null with indifference zone
  2. Find confidence, support or credible interval
  3. Measure the fraction of interval (#2) that is in the null interval
  4. Apply small-sample correction factor, as necessary

Second-generation p-values

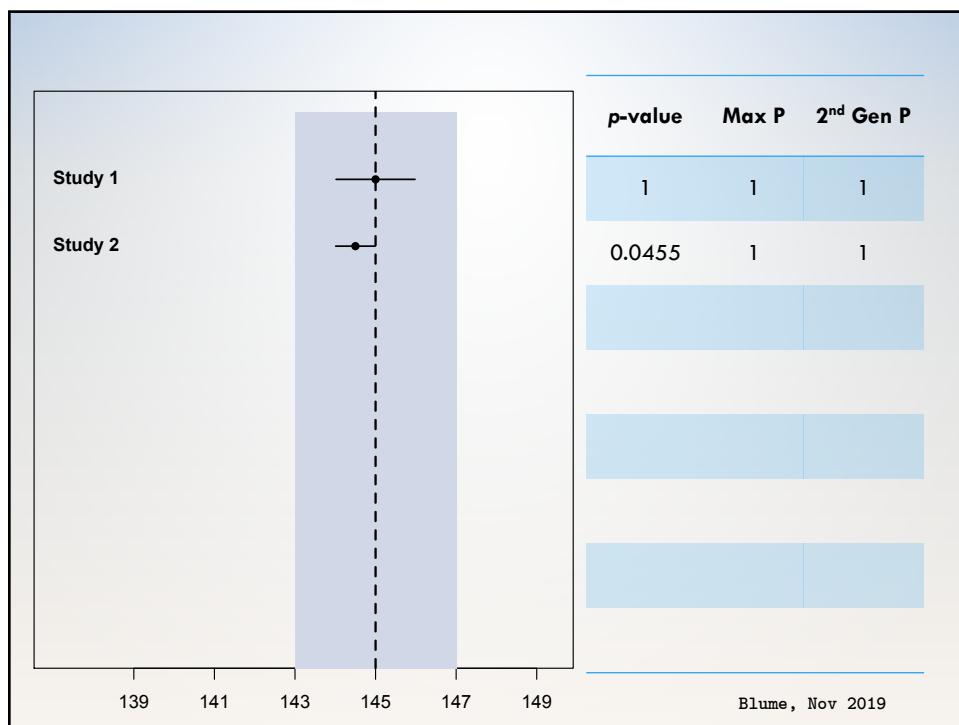
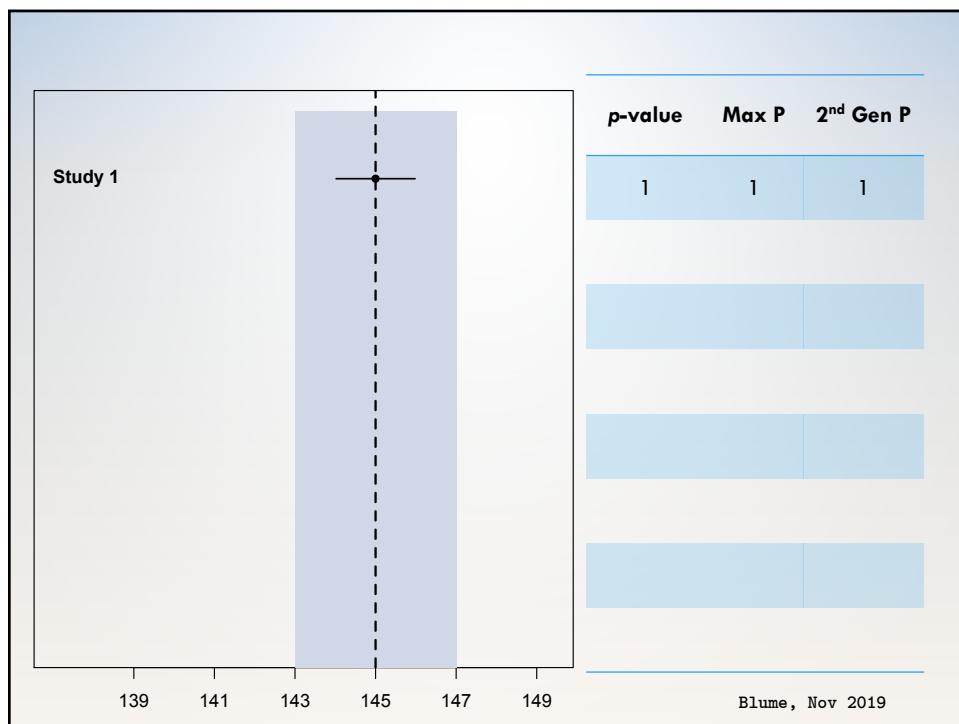
Blume, Nov 2019

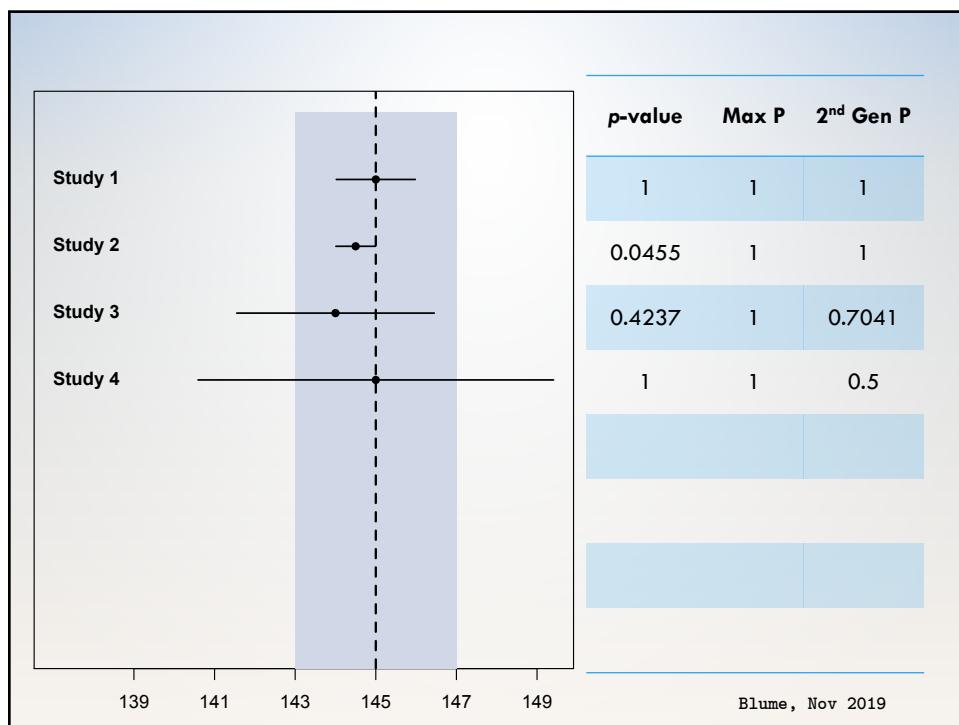
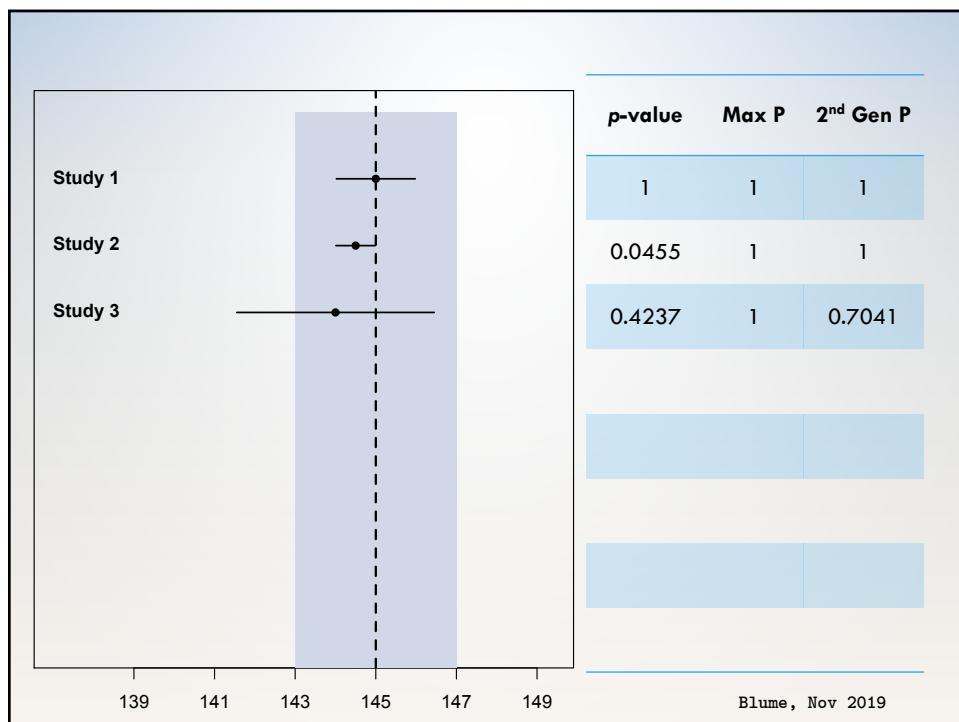
## Systolic Blood Pressure

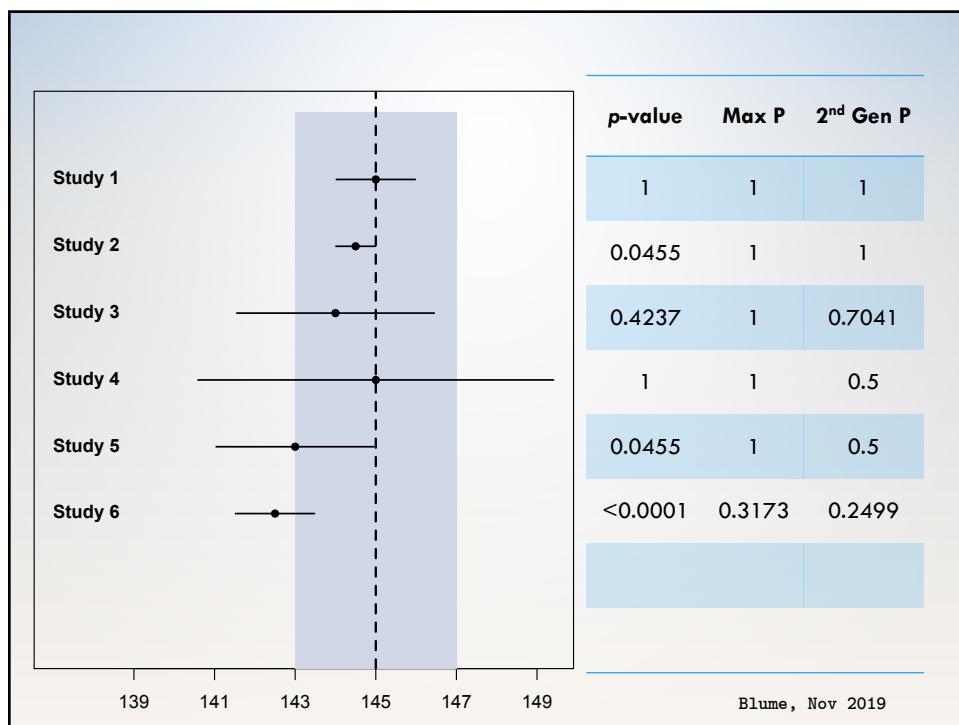
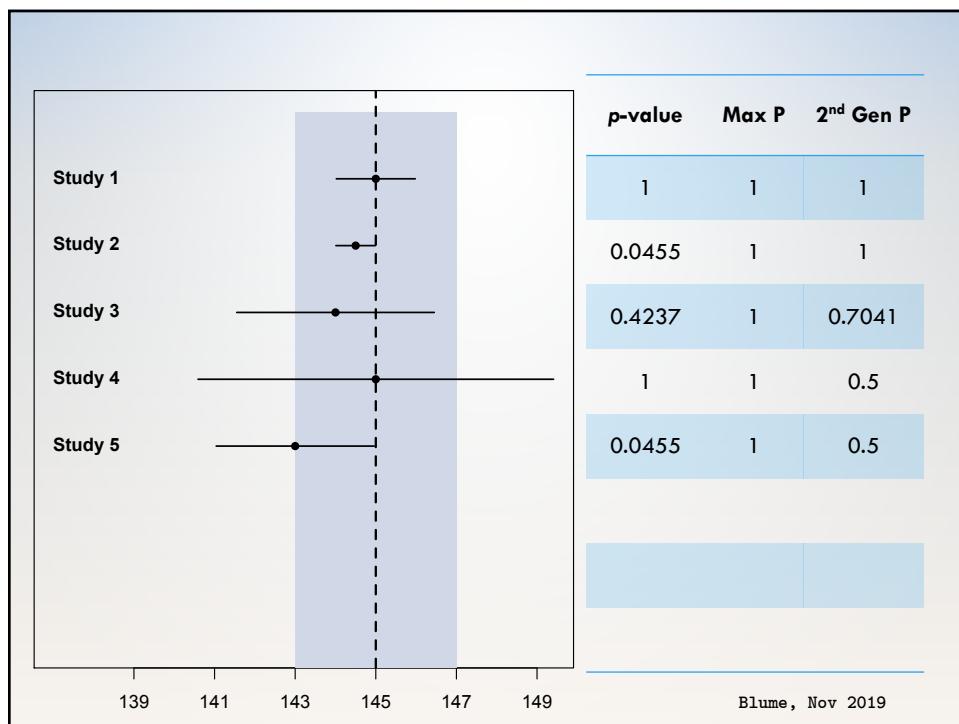
- SBP is reported to the nearest 2 mmHg
- Null Hypothesis: mean SPB is 145 mmHg
- Interval Null hypothesis: mean is 143 to 147 mmHg
- Results from 8 mock studies

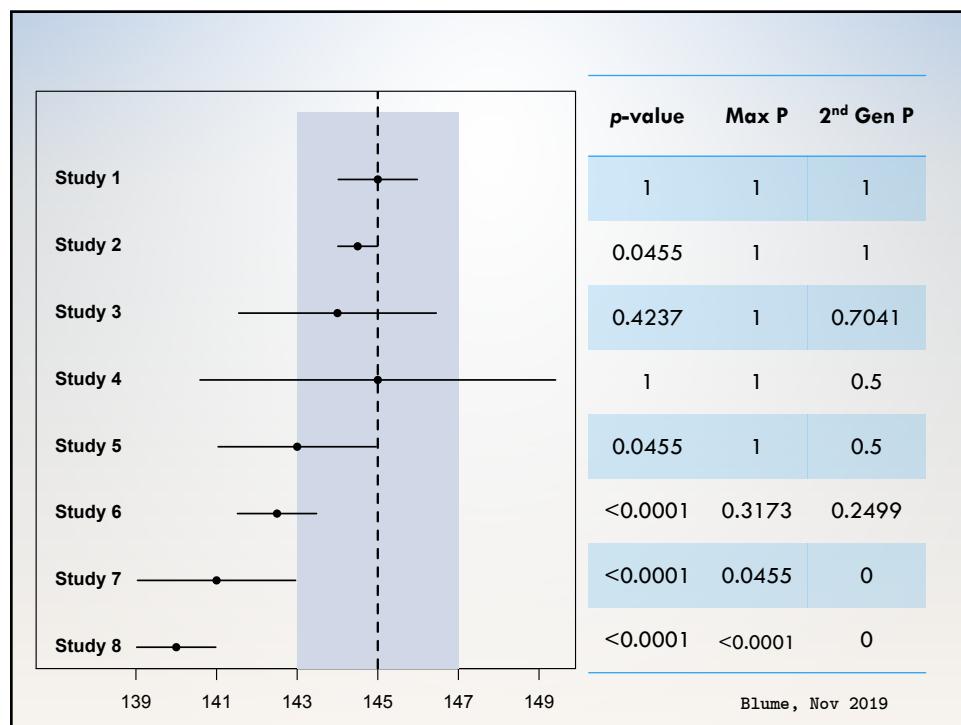
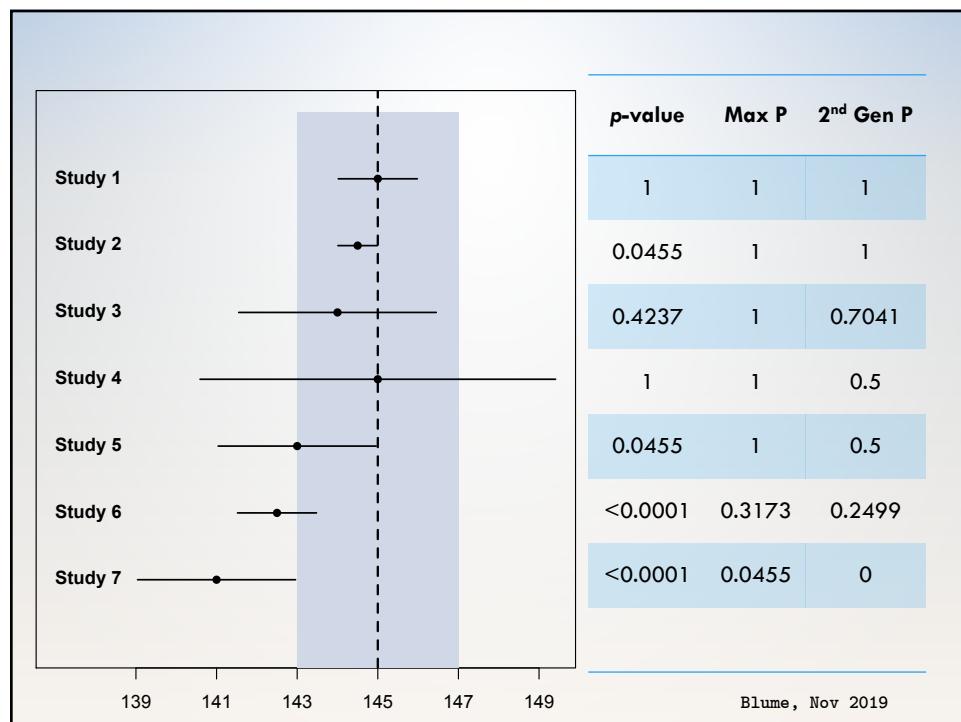
Second-generation p-values

Blume, Nov 2019









## 2x2 Tables & Odds Ratios

Exposure	Outcome	
	No	Yes
Exposed	35	65
Unexposed	50	50

OR = 1.86  
95% CI: (1.05, 3.29)

Null: (0.9, 1.11)

$$p_{\delta} = \frac{(1.11 - 1.05)}{(3.29 - 1.05)}(1) = 0.024$$

---

log(or) = 0.62  
95% CI: (0.05, 1.19)

Null: (-0.1, 0.1)

$$p_{\delta} = \frac{(0.1 - 0.05)}{(1.19 - 0.05)}(1) = 0.044$$

Second-generation p-values

Blume, Nov 2019

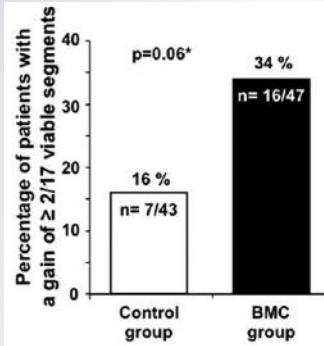
## Bone Marrow in Acute Myocardial Infarction (BOMAMI)

- European Heart Journal (2011)
- Randomized multicenter study
- Intracoronary administration of autologous bone marrow cells (BMCs) can lead to a modest improvement in cardiac function
- Aim: Evaluate the effect of BMC therapy on myocardial viability in patients with decreased left ventricular ejection fraction (LVEF) after acute myocardial infarction (AMI)

Second-generation p-values

Blume, Nov 2019

## BONAMI Trial



Odds ratios	95% confidence interval	P-value
2.654	0.967 - 7.286	0.06

Null Interval: (0.9, 1.11)

$$p_{\delta} = \frac{(1.11 - 0.967)}{(7.286 - 0.967)} (15) = 0.34$$

Second-generation p-values

Blume, Nov 2019

## Effect Measures for BOMAMI

	BMC	Control	Total
Gain	16	7	23
No Gain	31	36	67
Total	47	43	90
Risk	0.34	0.16	

Null Hypotheses

OR/RR: (0.9, 1.11)

RD: (-0.05, 0.05)

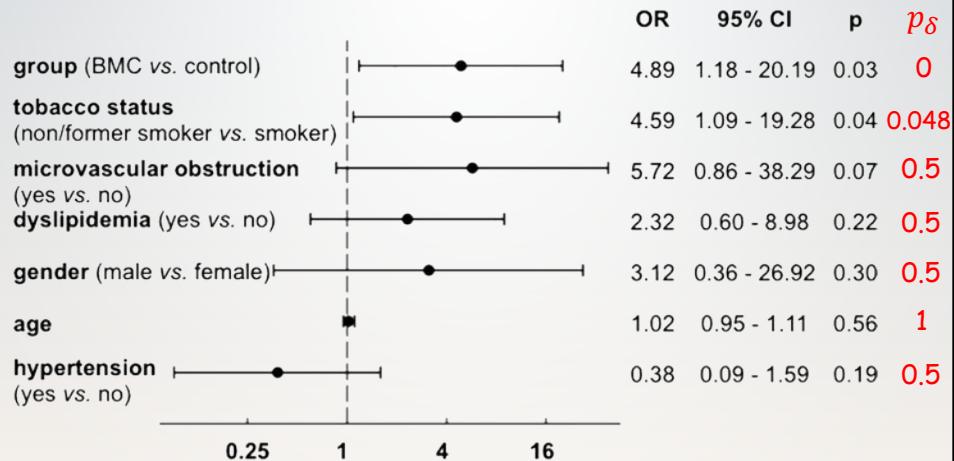
	Estimate	CI Lower	CI Upper	SGPV
Odds Ratio	2.65	0.967	7.286	0.34
Risk Ratio	2.09	0.953	4.589	0.37
Risk Difference	0.18	0.003	0.352	0.24

Second-generation p-values

Blume, Nov 2019

## BONAMI Trial

Logistic Regression with Null Zone: (0.9, 1.11)

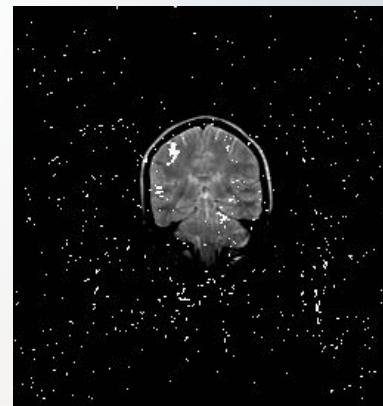


Second-generation p-values

Blume, Nov 2019

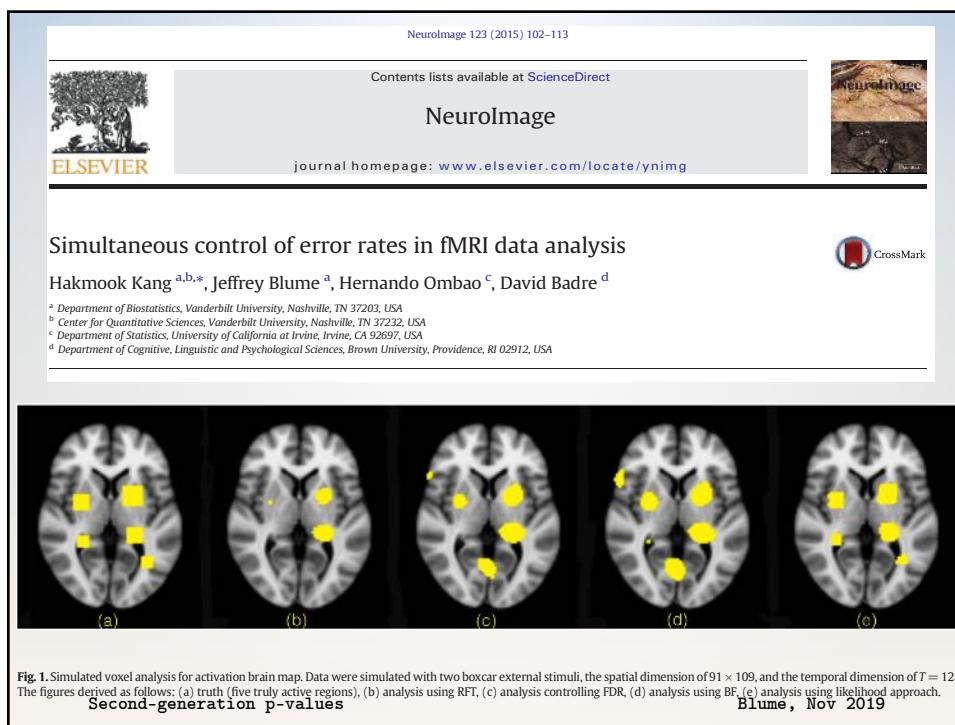
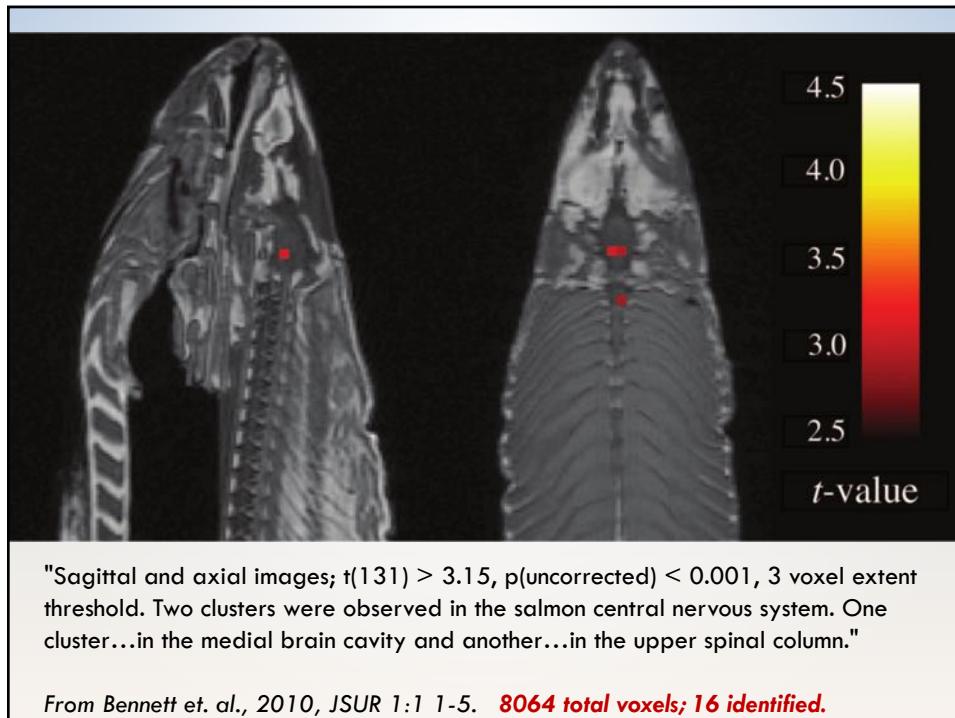
## Setting interval null

- Before analyzing data (!)
- Measurement error
- Subject matter knowledge
- Impact of findings
- Community standard
- Get creative (fMR example)
- Width not critical, buffer
- *The Atlantic salmon imaging saga*



Second-generation p-values

Blume, Nov 2019

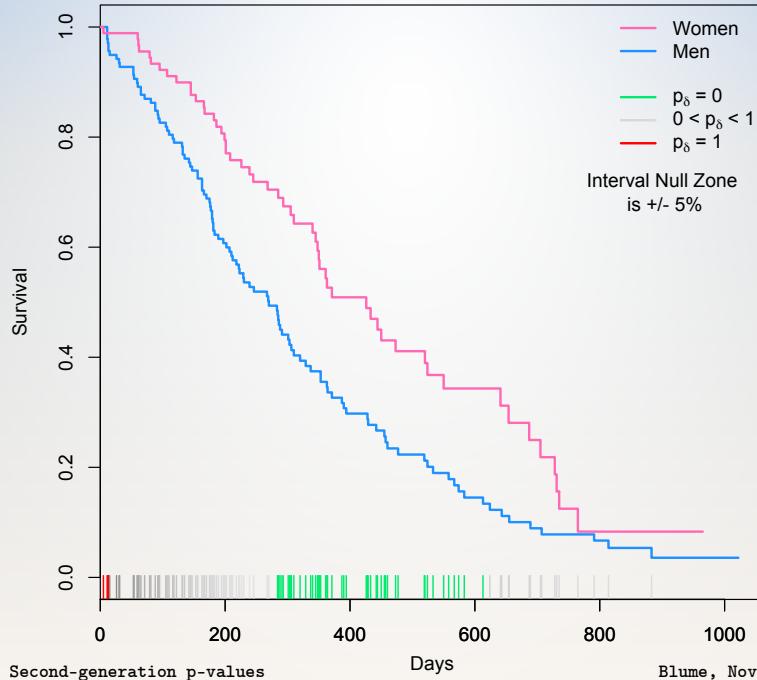


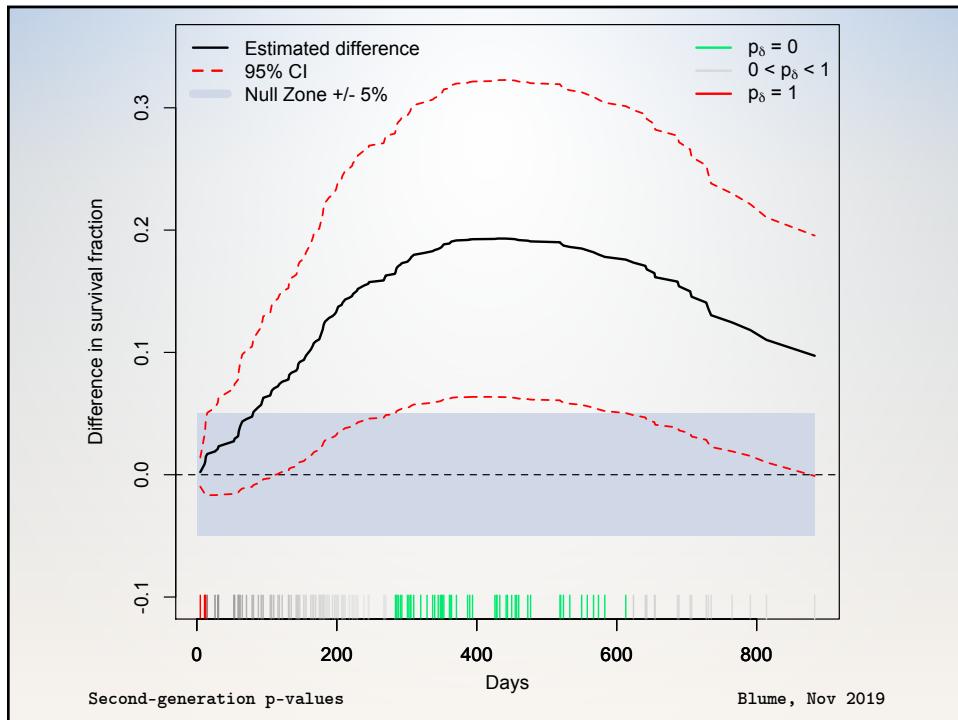
## Lung Cancer Survival

- Survival time in patients with advanced lung cancer (days)
- Potential for gender dissimilarities
- Trial by North Central Cancer Treatment Group (1994)

Second-generation p-values

Blume, Nov 2019



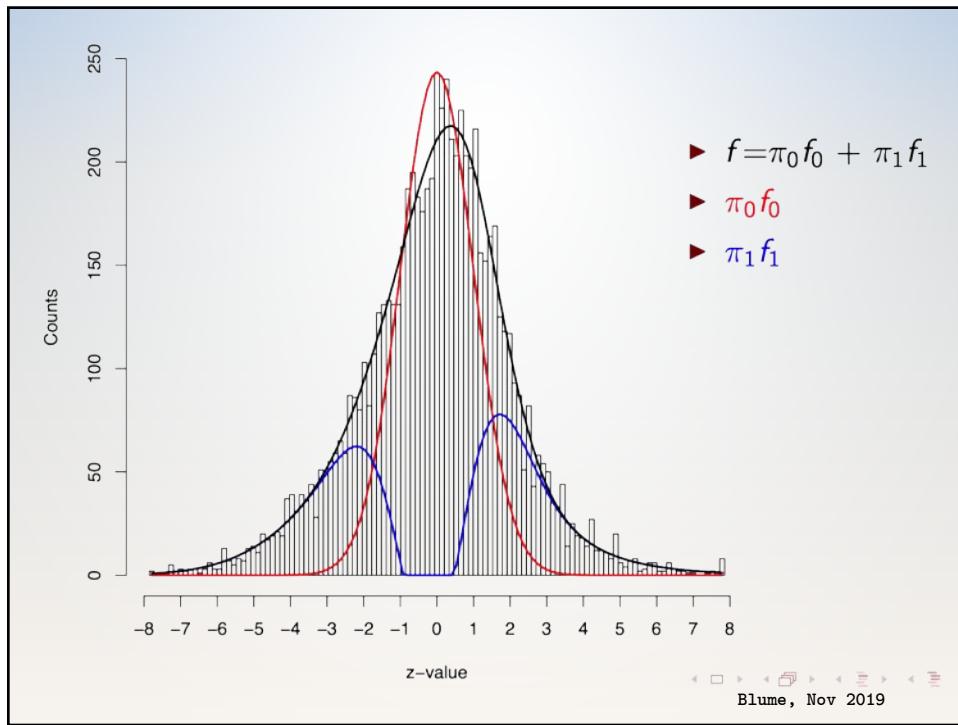
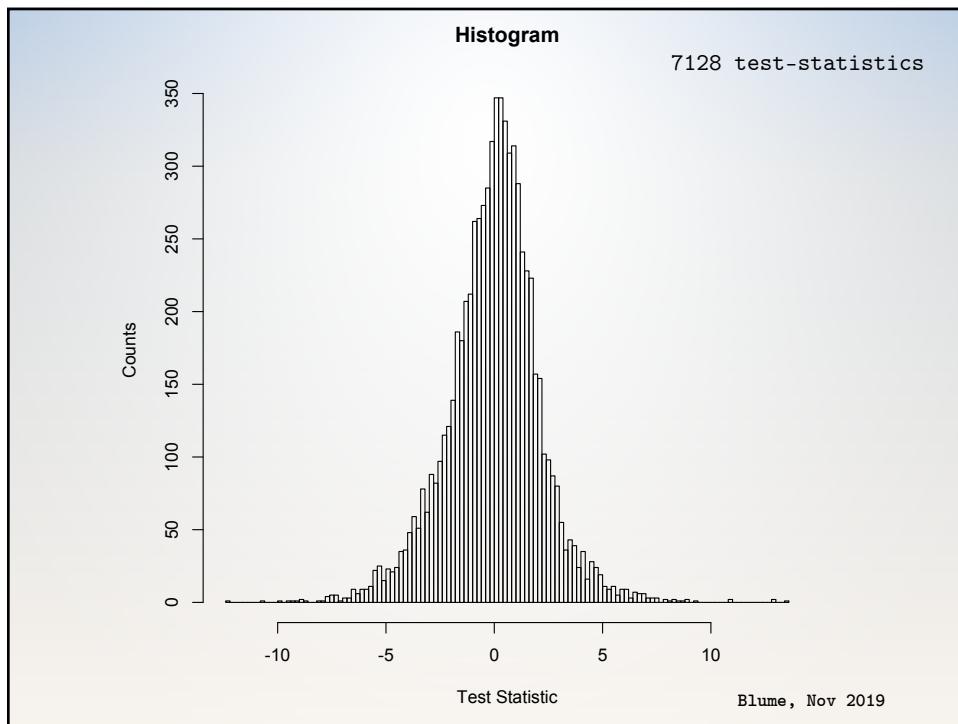


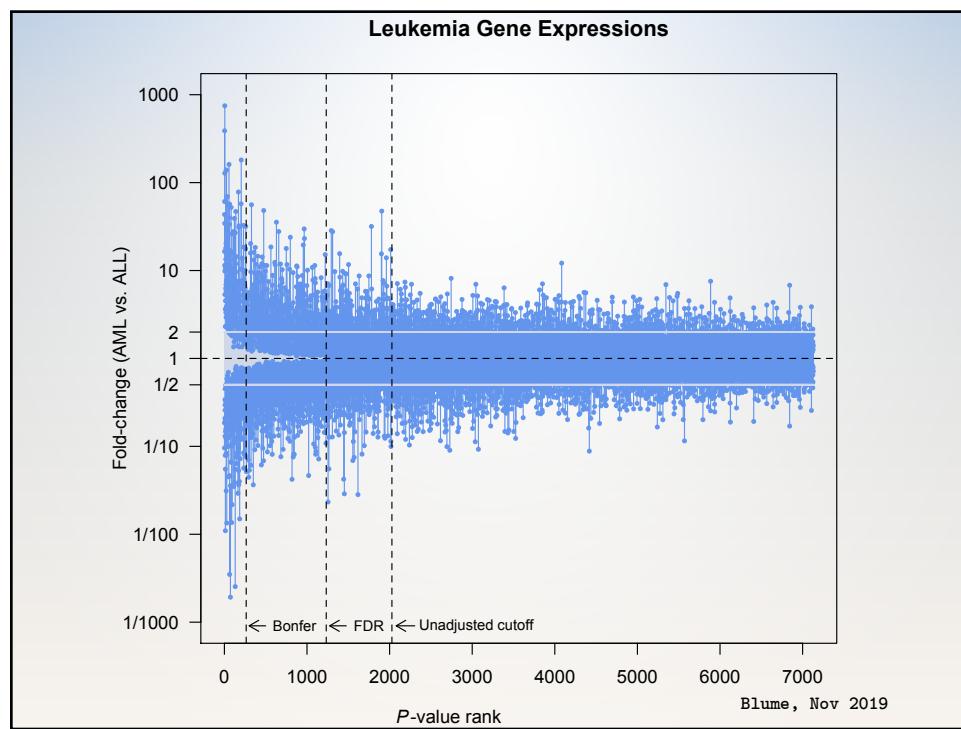
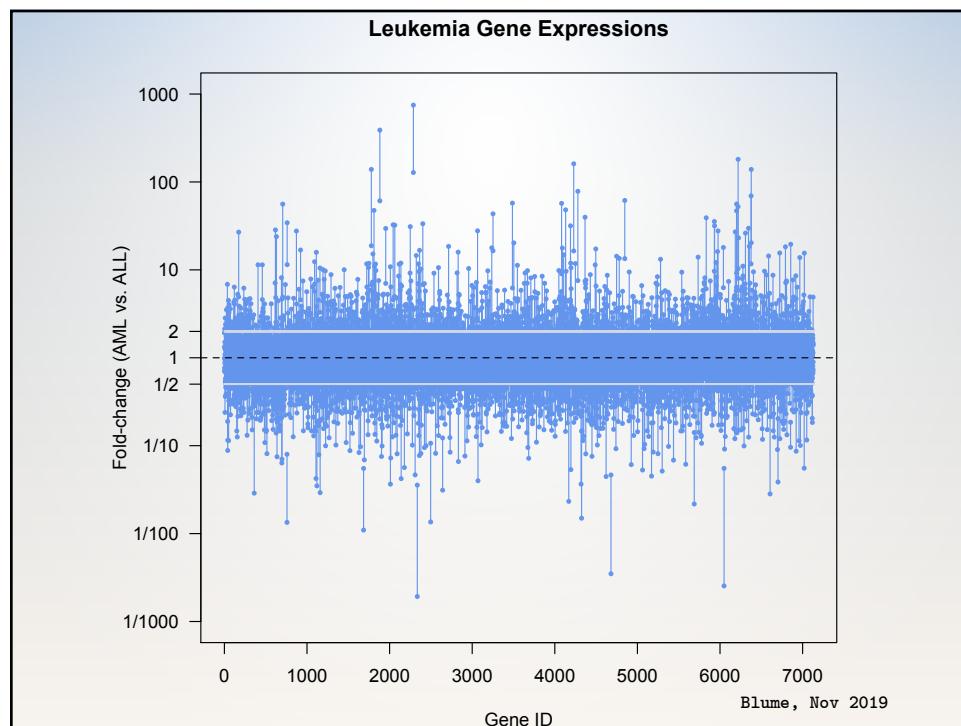
## Leukemia gene expression

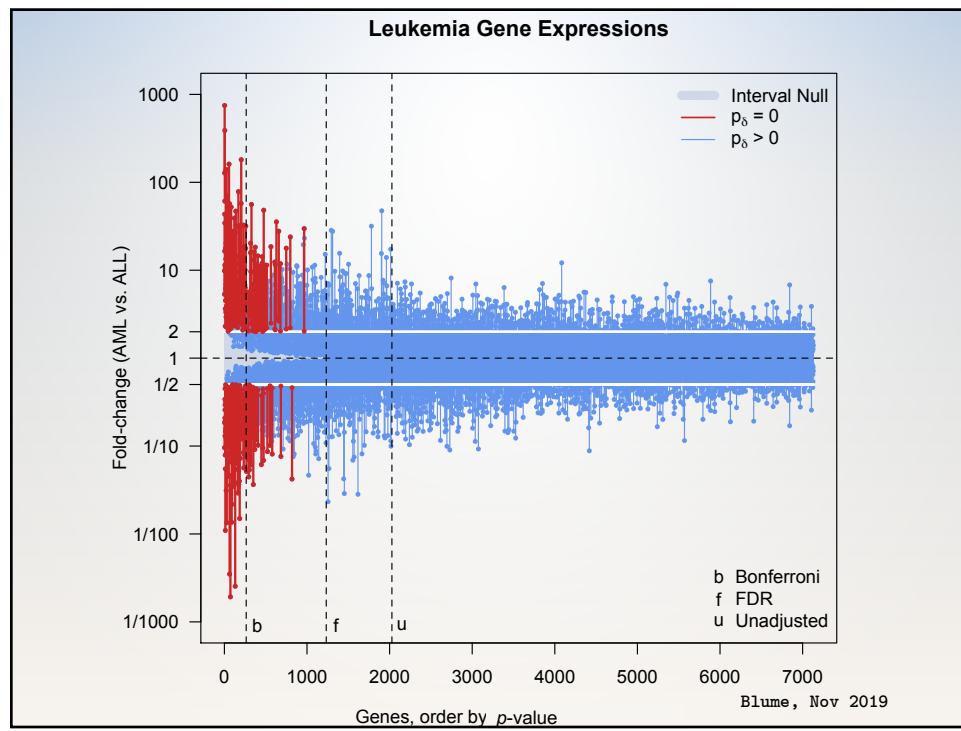
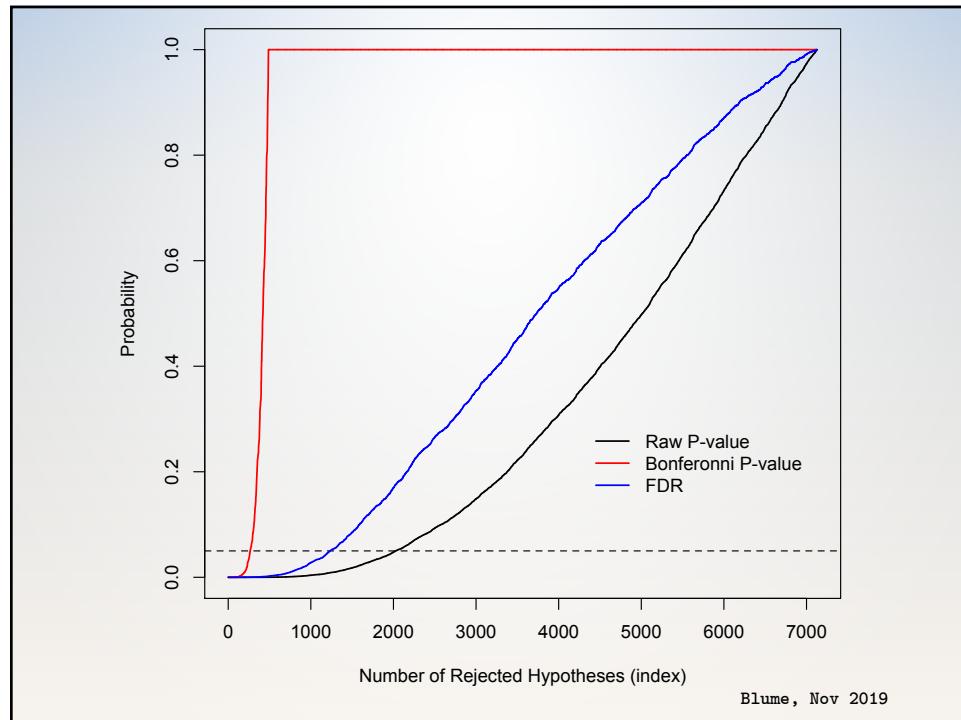
- Classifying acute leukemia by precursors  
(Golub 1999, *Science*)
- 7128 genes ; 72 patients (47 ALL and 25 AML)
- Affymetrix chip collected expression levels
- Goal: Identify 'interesting' genes whose expression levels differ between All and AML subjects.
- Looking for fold changes of 2 or more

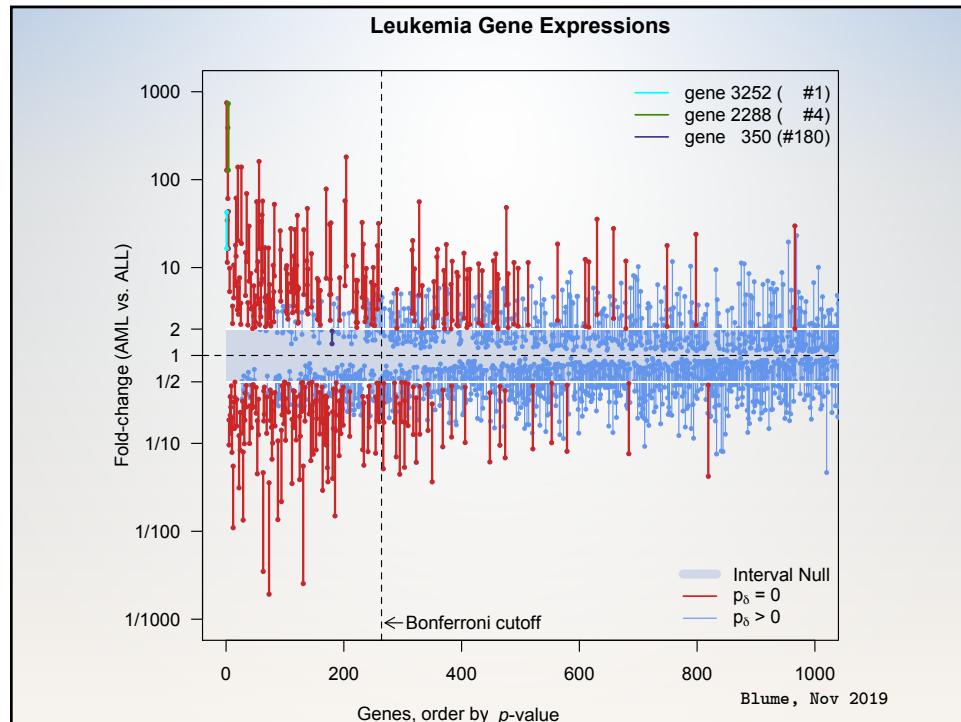
Second-generation p-values

Blume, Nov 2019









**Cross-Tabulation of Leukemia Results**

- Bonferroni vs Second-Generation *p*-values

	1/2 < Fold Change < 2 ( $\delta = 0.3$ )	1/1.915 < Fold Change < 1.915 ( $\delta = 0.282$ )		
	$p_\delta = 0$	$p_\delta > 0$	$p_\delta = 0$	$p_\delta > 0$
$p_{bon} < 0.05$	164	100	182	82
$p_{bon} > 0.05$	65	6799	82	6782
<b>Total</b>	229	6899	264	6864

Second-generation *p*-values

Blume, Nov 2019

## Leukemia study findings

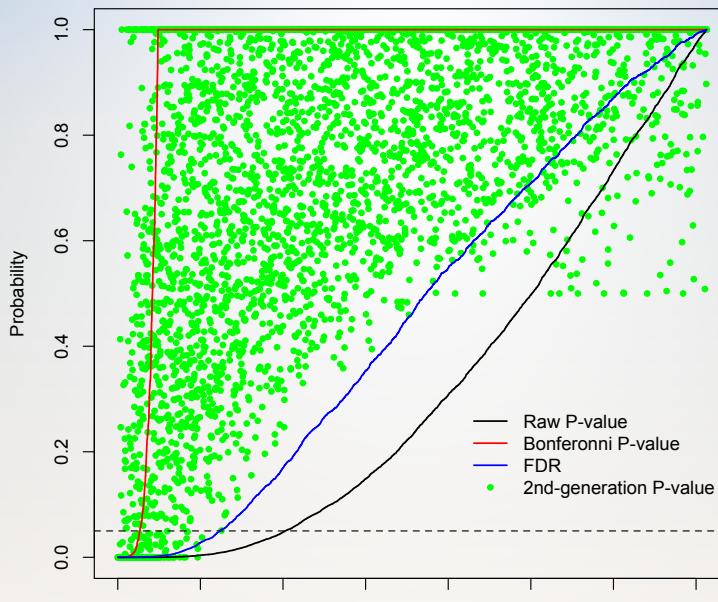
- Findings: Bonferroni 264, SGPV 229
  - Agree on 164 findings
  - Bonferroni +100, SGPV +65
- Effective Type I error rate: 0.037 vs. 0.032
- FDR of 2.45% captures all  $p_\delta = 0$ , 737 others
- Moving cutoff trades Type I for Type II errors
- SGPV changes the *ranking* of findings
  - Three categories now: null, alt, inconclusive
  - Null findings not illustrated here

Some SGPV findings  
have a priori  
published validation

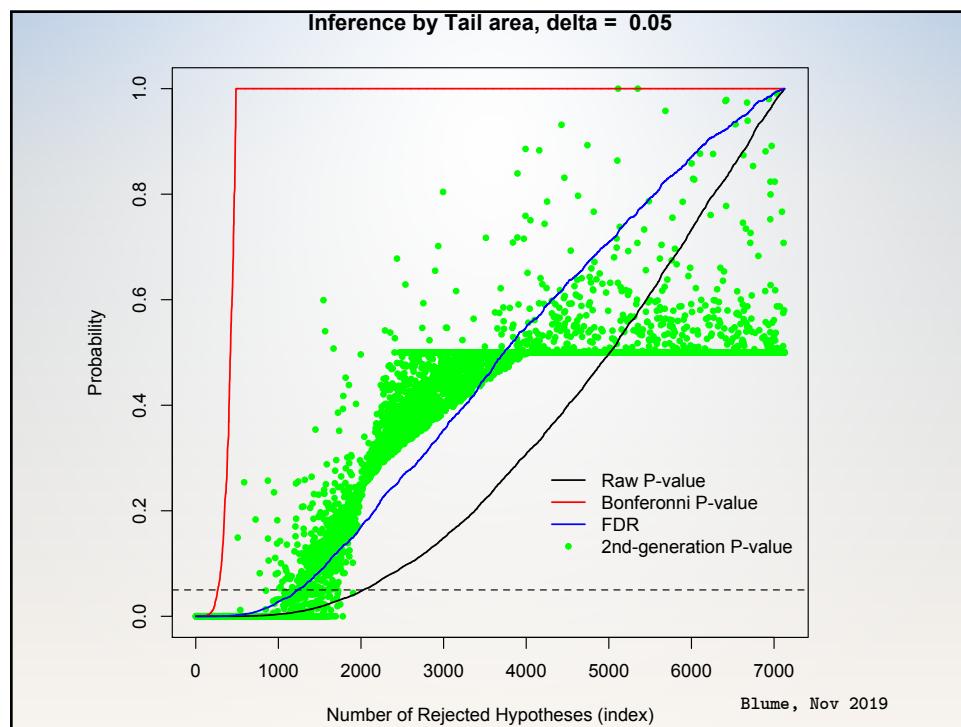
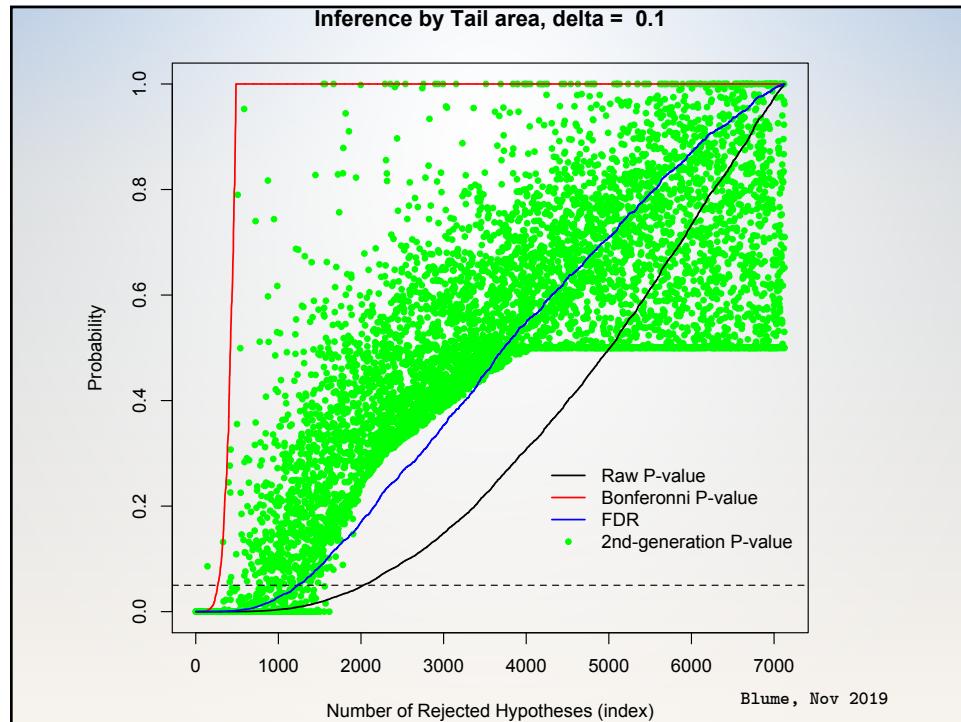
Second-generation p-values

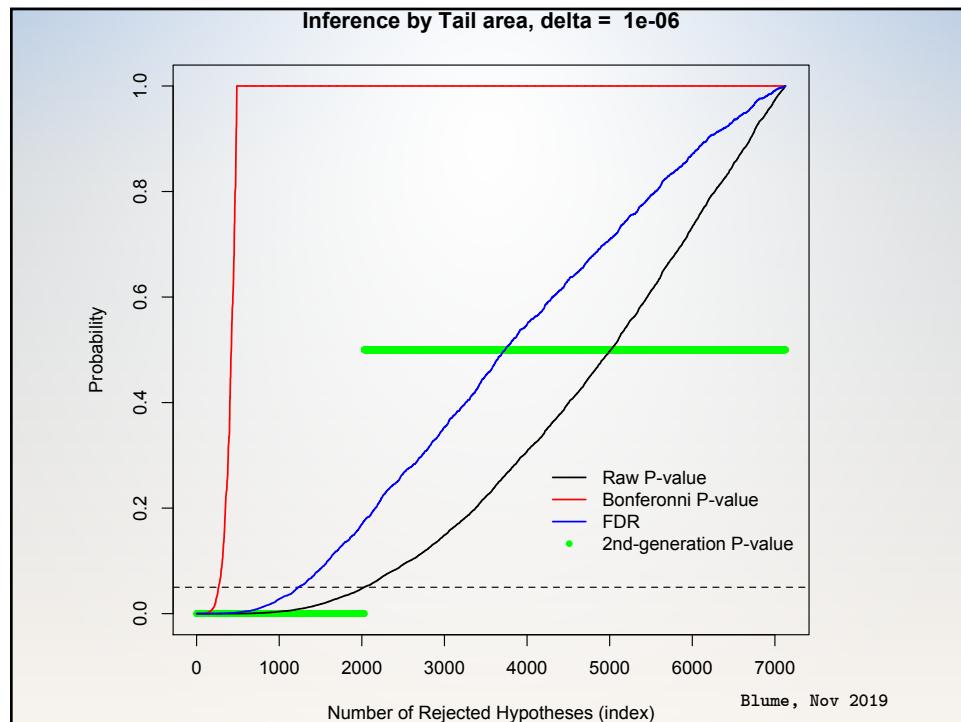
Blume, Nov 2019

Inference by Tail area, delta = 0.3



Blume, Nov 2019



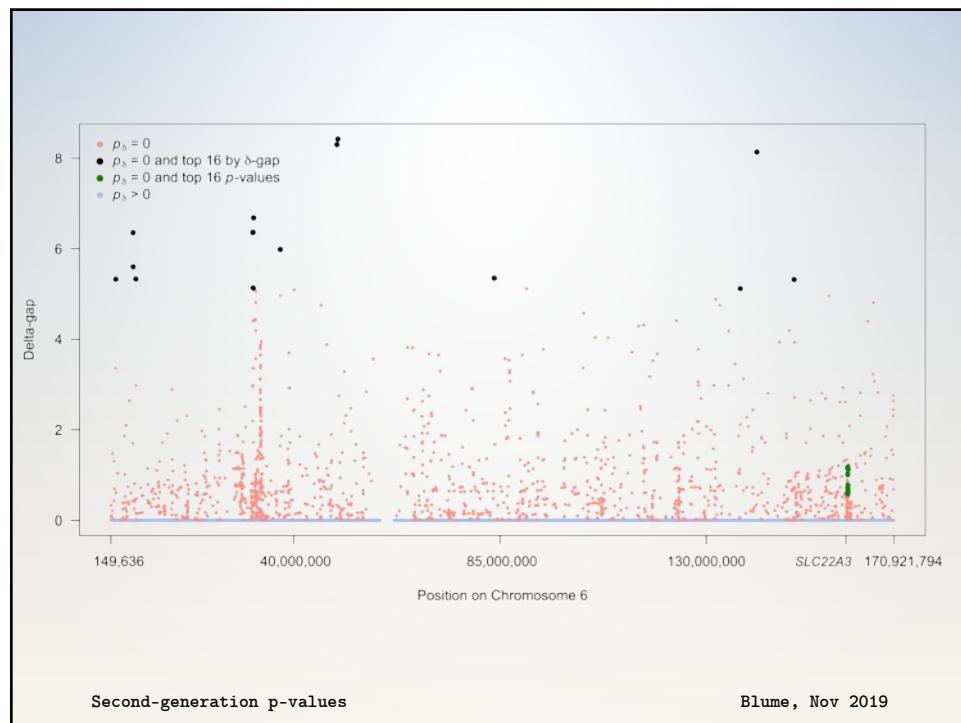
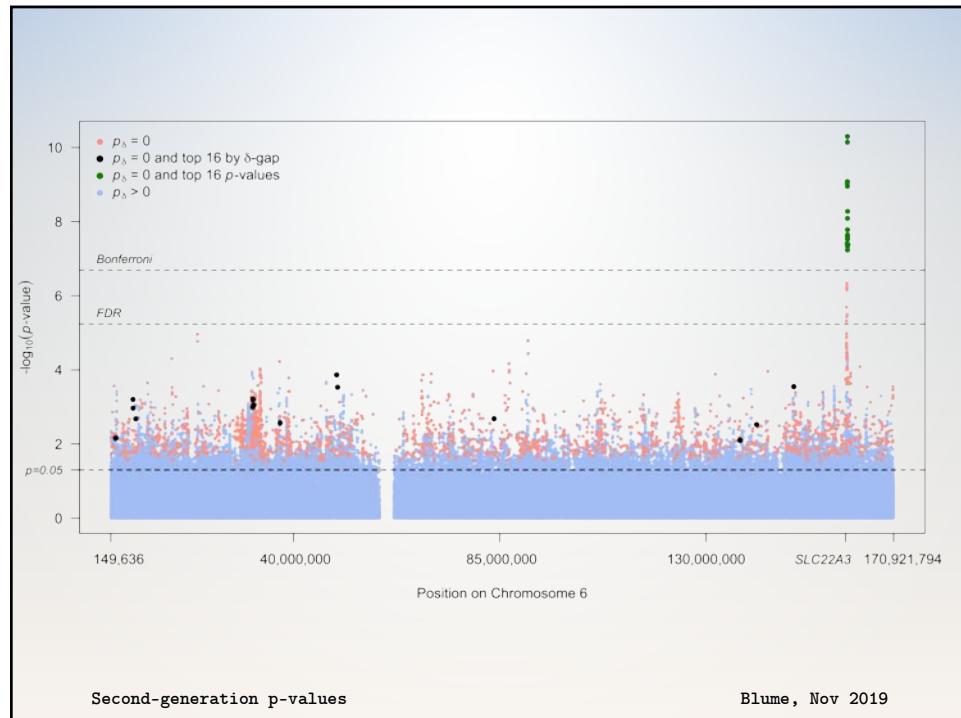


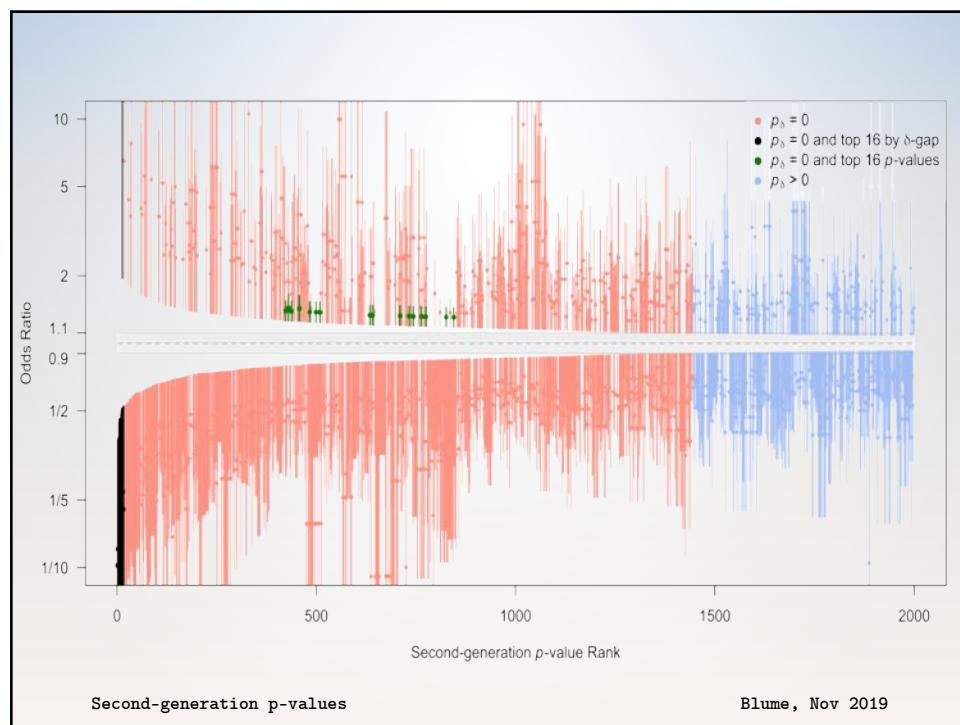
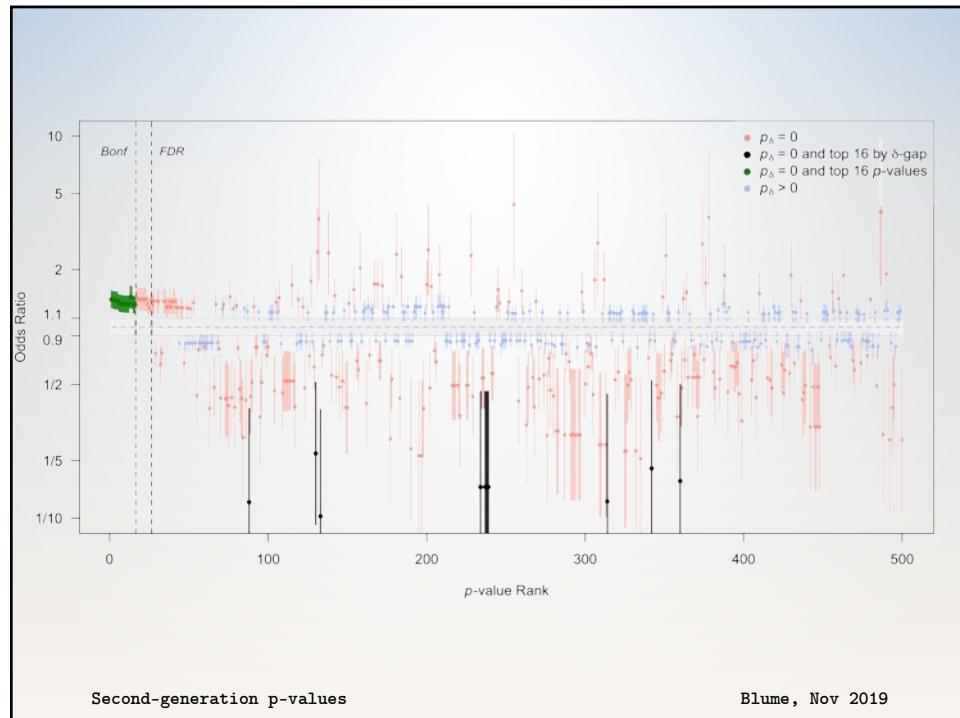
## Prostate Cancer SNPs

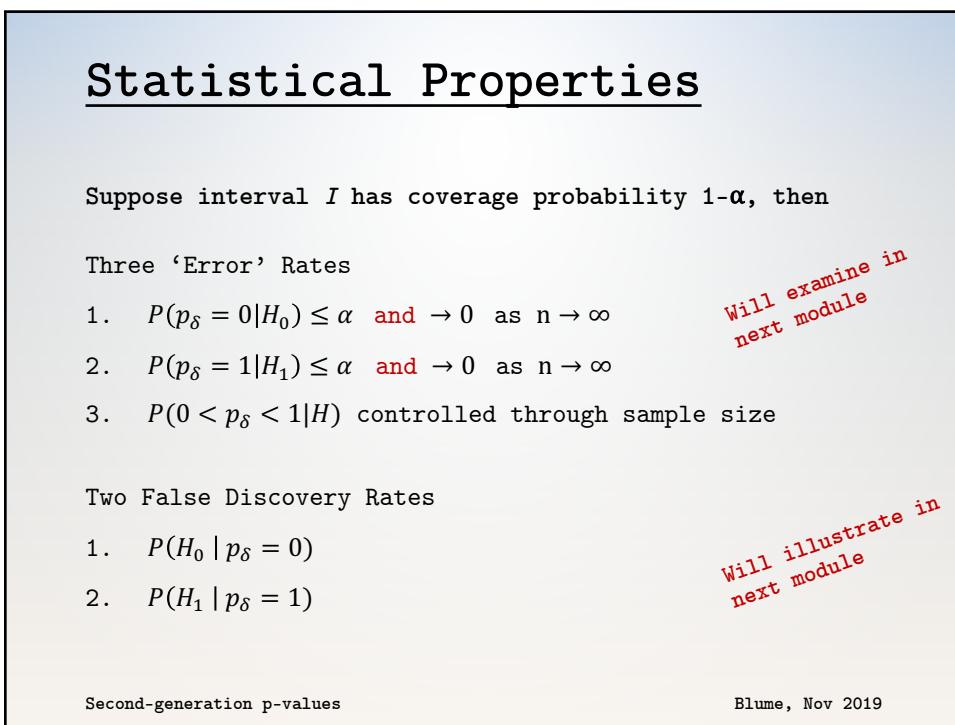
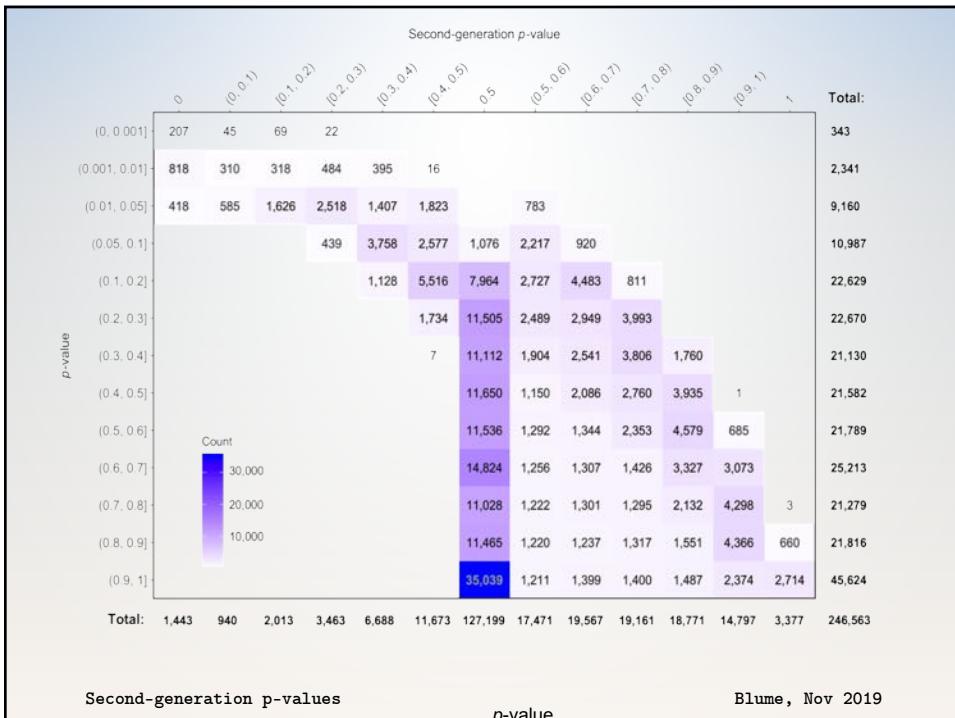
- International Consortium for Prostate Cancer Genetics (Schaid and Chang 2055; ICPCG 2018)
- 3,894 subjects: 2,511 cases & 1,383 controls
- 247,000 single-nucleotide polymorphisms (SNPs) from Chromosome 6
- Goal: Identify 'interesting' SNPs potentially associated with prostate cancer
- Looking for odds ratios of  $<0.9$  or  $>1.11$

Second-generation p-values

Blume, Nov 2019







## False Discovery Rates

- FDR for 5 SGPV=0 findings; computed under various null and alternative configurations (w/ flat prior).

SNP ID	SGPV rank	p-value rank	OR	1/8 SI lower limit	1/8 SI upper limit	FDR <sub>1</sub>	FDR <sub>2</sub>	FDR <sub>3</sub>
kgp4568244_C	1	133	0.10	0.03	0.37	2.9%	17.1%	3.3%
kgp8051290_G	13	2002	15.58	1.95	124.68	4.3%	30.3%	4.9%
kgp4497498_A	28	255	4.37	1.80	10.64	2.5%	8.6%	3.1%
rs3123636_G	423	1	1.39	1.26	1.55	0.01%	0.1%	0.4%
kgp7460928_G	1443	3310	1.78	1.11	2.87	2.4%	2.0%	3.0%

Second-generation p-values

Blume, Nov 2019

## Remarks

- Second-generation *p*-values...
  - Indicate compatibility with null or alternative
  - Indicate when the data are inconclusive (!)
  - Straightforward to compute and interpret
  - Controls error rate using *science*
  - Reduces the false discovery rate
- Anchoring the scale of the effect size...
  - Eliminates most Type I Errors
  - Improves scientific translation of statistical model

Second-generation p-values

Blume, Nov 2019

## Acknowledgements

- Collaborators
  - William D. Dupont
  - Robert A. Greevy
  - Lucy D'Agostino McGowan
  - Valerie Welty
  - Jeffrey R. Smith
  
- Website / Papers / Code
  - statisticalevidence.com
  - PLOS One ; TAS (In Press)
  - Google “Second-Generation *p*-value”
  - devtools::install\_github("weltybiostat/sgpv")

### Outrageous Claim (!?)

The SGPV achieves the inferential properties that many scientists hope, or believe, are attributes of the classic *p*-value.

Second-generation p-values

Blume, Nov 2019

Thank you for your attention.

Questions?

Second-generation p-values

Blume, Nov 2019