

TrueAllele® for Justice: Unbiased DNA Mixtures

Criminal Forensics Seminar
Louisiana Association of Criminal Defense Lawyers
August, 2018
Baton Rouge, LA

Jennifer M. Hornyak, MS
Cybergenetics, Pittsburgh, PA



Cybergenetics

Cybergenetics © 2003-2018

Pinkins confined

1989 – 5 men raped an Indiana woman
Darryl Pinkins and 2 others misidentified
1991 – wrongfully convicted, 65 year sentence

Pinkins guilty in bump-rape
NWI Times, May 4, 1991

Pinkins DNA evidence

2001 – DNA mixture evidence
2 contributors found, not the accused
but 5 were needed, post-conviction relief denied



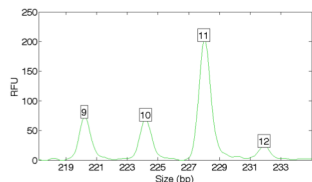
Jacket



Sweater

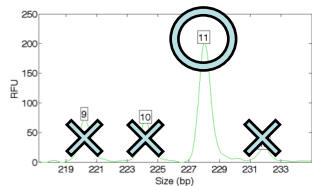
Good DNA mixture data

- two or more people
- small amounts of DNA
- degraded molecules



Bad DNA data interpretation

- biased exam
- wrong answer
- confusing result



Probabilistic genotyping



Options		Probability
A	→	A 20%
B		B 30%
C		C 50%

People choose their data

(1)
Simplify data



(2)
Peek at answer



(3)
Calculate statistic



- Put people in the process
- To overcome software failure
- And introduce human bias

Adjust data for bad software



Methods of misinterpretation

- **threshold** – method discards data
- **drop out** – method conjures data
- **wrong data** – relies on calibration
- **incomplete** – model missing variables
- **overconfident** – misses own uncertainty
- **human control** – introduces bias
- **not validated** – insufficient testing
- **undervalidated** – not fit for purpose

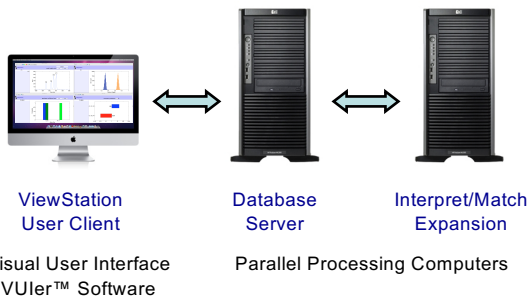
DNA injustice

Bad science leads to bad justice

- **inclusion**, biased wrong statistic
- **exclusion**, biased without statistic
- **inconclusive**, discard exculpatory evidence



TrueAllele® Casework



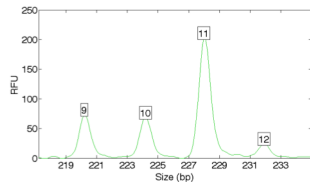
TrueAllele Pinkins findings

1. compared *evidence with evidence*
2. calculated *exclusionary match statistics*
3. revealed 5% *minor mixture contributor*
4. *jointly analyzed* DNA mixture data
5. showed three perpetrators were *brothers*

**found 5 unidentified genotypes,
defendants not linked to the crime**

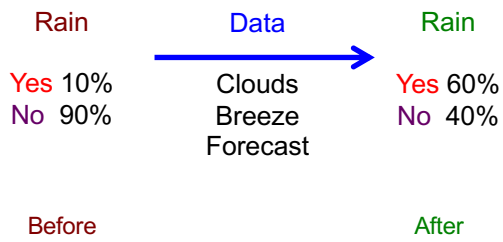
Good data interpretation

- objective
- accurate
- understandable



Bayes update

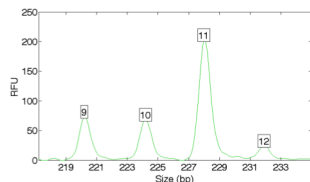
Assessing data changes our belief



All the data, all the time

Bayes: consider all data for valid answer

- all locus tests
- all data peaks
- no thresholds
- no dropout



Accurate unbiased method

Bayes: consider all variables for right answer

- all variables
- all possibilities
- no choices
- thorough testing



PAPER

CRIMINALISTICS

31 validation studies, 7 published

Mark W. Perlin,¹ Ph.D., M.D.; Jennifer M. Hornyak,¹ M.S.; Garrett Sugimoto,² M.S.; and Kevin W.P. Miller,² Ph.D.

TrueAllele[®] Genotype Identification on DNA Mixtures Containing up to Five Unknown Contributors*

J. Forensic Sci., July 2015, Vol. 60, No. 4
doi: 10.1111/1556-4029.12788
Available online at: onlinelibrary.wiley.com

TrueAllele justice

- **inclusion**, objective & accurate statistic
- **exclusion**, based on math, not opinion
- **inconclusive**, truly uninformative data

found 5 unidentified genotypes,
defendants not linked to the crime

Pinkins released



Crime labs on notice

MIX05 (2) – inconclusive, 4-14 zeros
MIX13 (3) – 70 of 100 labs falsely include
CPI statistic – random number, shutters labs



6 o'clock
or
nothing at all

Mixture interpretation failure

FORENSIC SCIENCES
PAPER
CRIMINALISTICS

Given lab result, what is real answer?

Mark W. Perlin,¹ M.D., Ph.D.; Matthew M. Legler,¹ B.S.; Cara E. Spencer,¹ M.S.; Jessica L. Smith,¹ M.S.; William P. Allan,¹ M.S.; Jamie L. Belrose,² M.S.; and Barry W. Duceman,³ Ph.D.

Validating TrueAllele® DNA Mixture Interpretation*†

FORENSIC SCIENCES
PAPER
CRIMINALISTICS

Given real answer, what is lab result?

Mark W. Perlin,¹ M.D., Ph.D.; Jamie L. Belrose,² M.S.; and Barry W. Duceman,³ Ph.D.

New York State TrueAllele® Casework
Validation Study*

Reliability of interpretation

Rule 702. Testimony by Expert Witnesses

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

- (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- (b) the testimony is based on sufficient facts or **data**;
- (c) the testimony is the product of reliable principles and **methods**; and
- (d) the expert has **reliably applied** the principles and methods to the facts of the case.

Unreliable data



- *choosing* introduces bias
- *discarding* violates Bayes
- *adding more* makes no sense
- wrong answers guaranteed

Unreliable method



- invalid use of DNA data
- calibrations are extraneous
- model leaves out variables
- unrealistic validation testing

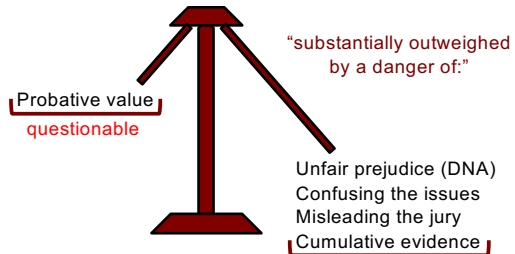
Unreliable result



- testing on limited samples
- not validated for actual use
- not applicable to case data
- report language confusing

Relevance of interpretation

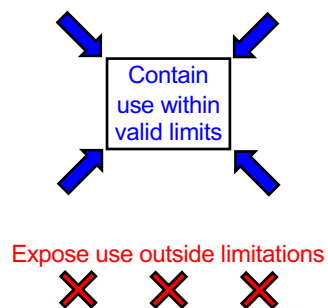
Rule 403. Excluding Relevant Evidence for Prejudice, Confusion, Waste of Time, or Other Reasons



Software summary

Majority	TrueAllele
Wrong	Accurate
Biased	Objective
Confusing	Understandable
Limited	Universal

Defense vigilance required



Recommendations

Most DNA mixture statistics
past, present and future
are wrong, biased and confusing

- educate defenders on DNA
- verify results (automation)
- cross exam to elicit truth
- expose the sins of the past

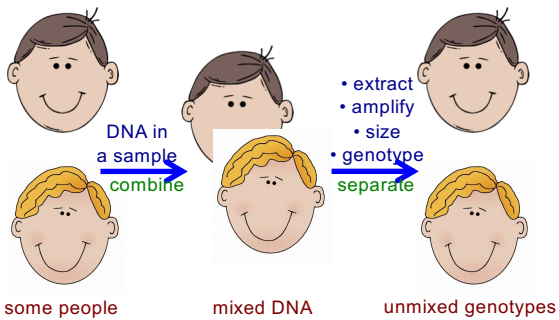
On the Threshold of Injustice: Manipulating DNA Evidence

American Academy of Forensic Sciences
Jurisprudence Section
February, 2017
New Orleans, LA



Cybergentics

DNA mixing and unmixing



Probabilistic genotyping

Bayesian

Noise	Model baseline
Much data	Use all peaks
Low data	All genotypes
Stutter	Mine the data
Variation	More variables

Objective computer operation

Less math, less capability

Bayesian

Noise	Model baseline
Much data	Use all peaks
Low data	All genotypes
Stutter	Mine the data
Variation	More variables

Incomplete

Thresholds
Discard peaks
Set dropout
Calibrate lab
Give up

Subjective human operator

Choices and consequences

Objective probability can be fair



Gambling with justice

Human choices
introduce bias

Subjectivity leads to
unfair outcomes

DNA evidence and results

Death by strangulation of 12 year old boy

150 biological **evidence** items,
with focus on DNA under victim's **fingernails**

Software finds match statistic of **ten million**,
connecting fingernails to defendant DNA

Unknown **minor** contributor is **0.4%** or **1:250**

Rule 702

- a. Sufficient data
- b. Reliable method
- c. Reliably apply method to data

Sufficient data

- Mixture
- ratio is 1:250
 - less than 1 cell

- Peak height
- 30 to 70 rfu

Fingernail data show
low mixture amount
& low peak heights
for minor contributor

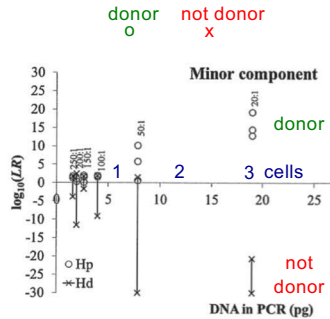
Reliable method

Mixture (validation)

- ratio of 1:25
- many cells

Peak threshold

- 30 rfu in study



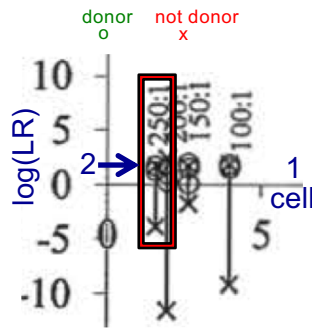
Reliably apply method to data

Mixture (case)

- ratio is 1:250
- less than 1 cell

Peak threshold

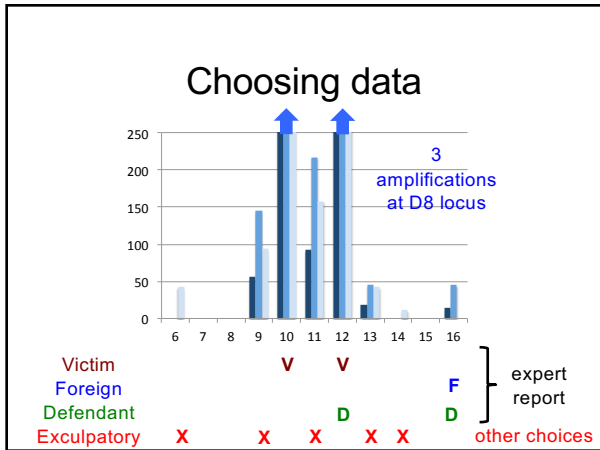
- 30 rfu in study
- 50 rfu in case



Applying thresholds

RFU	Data choices	All stutters	
80	0	9	
70	30	51	
60	250	1,660	
50	15,500,000	69,200	include
40	0	0	exclude
30	0	0	

Different choices, different answers
Software does not agree with itself



Double dropout

Allele	Amp 1	Amp 2	Amp 3
9	55	80	97
10	1,315	2,009	2,653
11		95	121
12	969	1,757	2,368

"Q" means **not** 9, 10, 11 or 12 data allele
 Defendant's **17,17** is **not in the data**

Hp weight for Q,Q genotype = 15%
 Hd weight for Q,Q genotype = 14%
 Likelihood ratio is 15%/14% = **1.05 > 1**
 Non-data Q,Q matches defendant 17,17
Excluded from data, but inclusionary LR

Judge's ruling

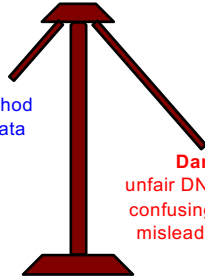
The Expert conceded at the hearing that **no internal validation studies were performed** by the State crime lab for the use of the Software on casework samples developed at the lab.

As a result the Expert was forced to **pick and choose data** from different "reliable sources" and **input parameters** into the program in such a way that he believed the system would tolerate.

ORDERED that the **defendant's motion to preclude** the prosecution from calling an expert witness to testify on their direct case regarding any conclusion reached by the use of the Software is **granted** as the prosecution cannot lay a foundation for the introduction of evidence that had not been internally validated.

Rule 403

Probative value
unreliably applied method
on insufficient DNA data



Danger of
unfair DNA prejudice,
confusing the issues,
misleading the jury

Recommendations

Discovery for Software

- validation studies (internal & other)
- user, procedure and training manuals
 - papers, reports, math description
 - data choices, parameter settings
- all electronic DNA Data in the case
- demand working Software program
- run Software on Data to replicate results
- run different Software on Data to confirm

Overcoming Bias in DNA Mixture Interpretation

American Academy of Forensic Sciences

February, 2016

Las Vegas, NV



Cybergenetics

DNA



Does
Not
Advocate

Gold standard of forensic evidence



However, ... there may be
problems ... with how the DNA
was ... interpreted, such as
when there are mixed samples

Law, Probability and Risk (2009) 8, 257–276
Advance Access publication on July 28, 2009

doi:10.1093/lpr/mgp013

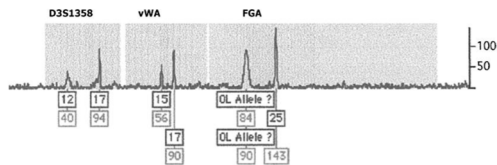
Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation†

WILLIAM C. THOMPSON*

Department of Criminology, Law and Society,
University of California, Irvine, CA 92697, USA

[Received on 17 November 2008; revised on 19 April 2009; accepted on 24 April 2009]

Forensic DNA analysts tend to underestimate the frequency of matching profiles (and overestimate likelihood ratios) by shifting the purported criteria for a ‘match’ or ‘inclusion’ after the profile of a suspect becomes known—a process analogous to the well-known Texas sharpshooter fallacy. Using examples from casework, informal and naturalistic experiments, and analysts’ own testimony, this article demonstrates how *post hoc* target shifting occurs and how it can distort the frequency and likelihood ratio statistics used to characterize DNA matches, making matches appear more probative than they actually are. It concludes by calling for broader adoption of more rigorous analytical procedures, such as sequential unmasking, that can reduce the sharpshooter fallacy by fixing the target before the shots are taken.



Defendant	D3S1358	vWA	FGA
Tom	17,17	15,17	25,25
Dick	12,17	15,17	20,25
Harry	14,17	15,17	20,25
Sally	12,17	15,15	20,22

FIG. 2. Electropherogram of a saliva sample and four suspect profiles.

Science and Justice 51 (2011) 204–208

Contents lists available at SciVerse ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

Subjectivity and bias in forensic DNA mixture interpretation[☆]

Itiel E. Dror^{a,b,*}, Greg Hampikian^c

^a Institute of Cognitive Neuroscience, University College London (UCL), London, UK
^b Cognitive Consultants International (CCI), London, UK
^c Departments of Biology and Criminal Justice, Boise State University, USA

ARTICLE INFO

Article history:
 Received 30 July 2010
 Received in revised form 30 July 2011
 Accepted 4 August 2011

Keywords:
 Human cognition
 Bias
 Forensic decision making
 Contextual influences
 DNA interpretation

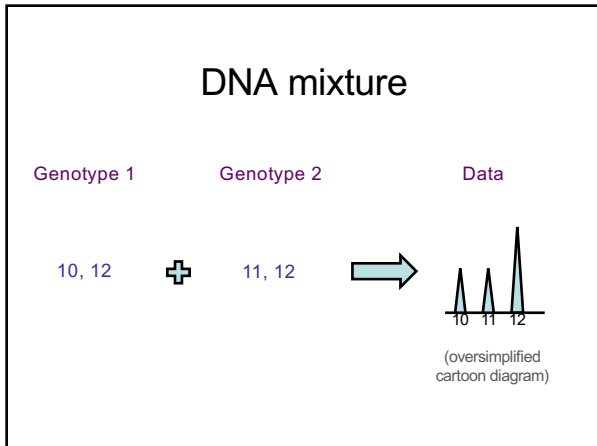
ABSTRACT

The objectivity of forensic science decision making has received increased attention and scrutiny. However, there are only a few published studies experimentally addressing the potential for contextual bias. Because of the esteem of DNA evidence, it is important to study and assess the impact of subjectivity and bias on DNA mixture interpretation. The study reported here presents empirical data suggesting that DNA mixture interpretation is subjective. When 17 North American expert DNA examiners were asked for their interpretation of data from an adjudicated criminal case in that jurisdiction, they produced inconsistent interpretations. Furthermore, the majority of 'context free' experts disagreed with the laboratory's pre-trial conclusions, suggesting that the extraneous context of the criminal case may have influenced the interpretation of the DNA evidence, thereby showing a biasing effect of contextual information in DNA mixture interpretation.

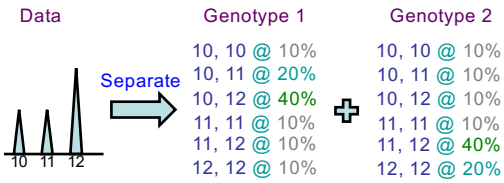
© 2011 Forensic Science Society. Published by Elsevier Ireland Ltd. All rights reserved.

Case context impact

	With context	Without context
Include	2	1
Exclude		12
Inconclusive		4

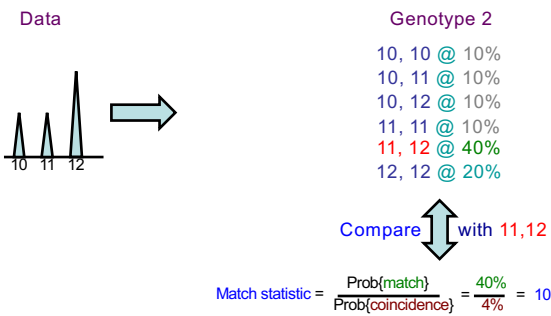


Interpret #1: separate



Unmix the mixture

Interpret #2: compare



Cognitive bias

Illogical thinking affects decisions

- **Anchoring** – rely on first information
- **Apophenia** – perceive meaningful patterns
- **Attribution bias** – find causal explanations
- **Confirmation bias** – interpretation confirms belief
- **Framing** – social construction of reality
- **Halo effect** – sentiments affect evaluation
- **Oversimplification** – simplicity trumps accuracy
- **Self-serving bias** – distort to maintain self-esteem

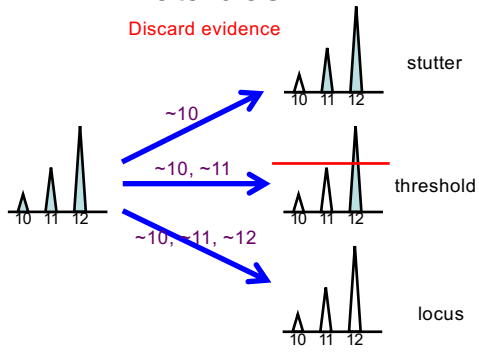
Contextual bias

Background information affects decisions

- **Academic bias** – beliefs shape research
- **Educational bias** – whitewash damaging evidence
- **Experimenter bias** – expectations affect outcomes
- **Inductive bias** – tilt toward training examples
- **Media bias** – selecting mass media stories
- **Motivational bias** – reaching desired outcome
- **Reporting bias** – under-report undesirable results
- **Social desirability bias** – want to be seen positively

Data bias

Discard evidence



Genotype bias

Actual		Desired
10, 10 @ 5%		10, 10 @ 0%
10, 11 @ 5%		10, 11 @ 0%
10, 12 @ 75%	→	10, 12 @ 100%
11, 11 @ 5%		11, 11 @ 0%
11, 12 @ 5%		11, 12 @ 0%
12, 12 @ 5%		12, 12 @ 0%

RMP – random match probability
analyst chooses only one genotype
inflates DNA match statistic

Perlin, M.W. "Inclusion probability for DNA mixtures is a subjective one-sided match statistic unrelated to identification information." *Journal of Pathology Informatics*, 6(1):59, 2015.

Match bias

CPI – combined probability of inclusion
analyst begins by including the suspect
unrealistic, unproven model
random number generator
lacks probative value

LR – likelihood ratio
analyst ignores much of the data
calculation requires suspect genotype
introduces "phantom" peaks (drop out)
considers few genotype possibilities

Process bias

(1)
Choose, alter, discard,
edit, and manipulate
the DNA **data signals**



Hidden cognitive and contextual bias
largely determine the outcome

(2)
Compare defendant's
genotype to edited
data & decide if he is
in the DNA evidence



(3)
If he is "included",
then **calculate** a
DNA mixture statistic



Presented as
unbiased science

Software bias

Why labs choose mixture software

- Puts analyst in charge
 - Results confirm belief
 - Simplifies the problem
 - Gets desired answer
 - The FBI uses it
 - Familiar process
- Confirmation bias**
Confirmation bias
Oversimplification
Motivational bias
Social desirability bias
Social desirability bias

Relevance (FRE 403)

Admissibility of **biased DNA** evidence

Rule 401
"evidence makes a fact more or less probable"

Probative value
inflated

Rule 403
"substantially outweighed by a danger of:"

Unfair prejudice
Confusing the issues
Misleading the jury
Wasting time
Cumulative evidence
"DNA"

Cross examination

Hundreds of effective questions can elicit bias

"Did you know the defendant's genotype during your analysis of the evidence?"

"Doesn't knowing your customer's desired answer bias your decisions?"

"Have any scientific studies demonstrated otherwise?"

Sequential unmasking

Human DNA review proposal (reduce bias):

1. **First analyze** the crime scene **data**, without knowing context or references
2. **Then compare** with reference samples

But there is potential **bias** in choosing **data**, conducting **analysis**, and making **comparisons**.

Human analysts can always introduce bias.
Why is a human even involved in this process?
Why not use an unbiased computer instead?

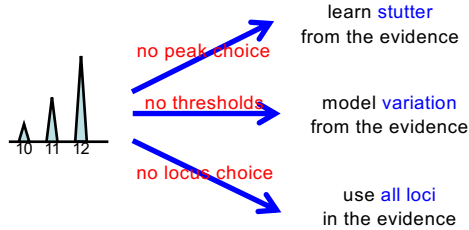
Unbiased interpretation

Use an objective **computer** to:

1. **Examine** *all* DNA data, without having suspect's genotype
2. **Separate** genotypes of each DNA mixture contributor, considering *all* possible solutions
3. **Compare** genotypes only *afterwards* to calculate match statistics

Eliminate all human involvement
to overcome cognitive & contextual bias
in DNA mixture interpretation

No data bias – use all evidence



No genotype bias – objective

Actual		Desired
10, 10 @ 5%	↔	Use the actual genotype probability
10, 11 @ 5%		
10, 12 @ 75%		
11, 11 @ 5%		
11, 12 @ 5%		
12, 12 @ 5%		

Do not change probability

No match bias – accurate

CPI – combined probability of inclusion
random number generator
bad forensic science
review all past cases

LR – likelihood ratio
don't ignore any data
don't use suspect genotype
don't concoct "phantom" peaks
use all genotype possibilities

No process bias – remove analyst

(1)
Do not change
data signals



Eliminate cognitive and contextual
bias from the process

(2)
Do not use
defendant genotype



(3)
Calculate accurate
DNA match statistic

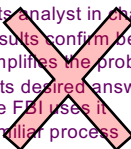


Present
unbiased science

No software bias – true stats

Accurate, objective, thorough, validated

- Puts analyst in charge
- Results confirm belief
- Simplifies the problem
- Gets desired answer
- The FBI uses it
- Familiar process



Examine all the data
without human choice



Separate genotypes
consider all solutions



Compare genotypes
stats decide outcome

Resources

<http://www.cybgen.com/information>



- Courses
- Newsletters
- Newsroom
- Presentations
- Publications
- Webinars

<http://www.youtube.com/user/TrueAllele>
TrueAllele YouTube channel



Cybergenetics



perlin@cybgen.com
