

# Gepaarter und ungepaarter t-Test

für D-UWIS, D-ERDW, D-USYS und D-HEST – SS15



# Repetition: t-Test für eine Stichprobe

1. **Modell:**  $X_i$  kontinuierliche Messgrösse;  
 $X_1, X_2, \dots, X_n$  i. i. d.,  $\mathcal{N}(\mu, \sigma_X^2)$ ,  $\sigma_X$  wird mit  $\hat{\sigma}_X$  geschätzt

2. **Nullhypothese:**  $\mathcal{H}_0: \mu = \mu_0$   
**Alternative:**  $\mathcal{H}_A: \mu \neq \mu_0$  (oder  $<$  oder  $>$ )

3. **Teststatistik:**  

$$T = \frac{(\bar{X}_n - \mu_0)}{\hat{\sigma}_{\bar{X}_n}} = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_X} = \frac{\text{beobachtet} - \text{erwartet}}{\text{geschätzter Standardfehler}}$$
 Verteilung unter  $\mathcal{H}_0$ :  $T \sim t_{n-1}$

4. **Signifikanzniveau:**  $\alpha$

5. **Verwerfungsbereich** für die Teststatistik:

$$K = (-\infty, -t_{n-1; 1-\frac{\alpha}{2}}] \cup [t_{n-1; 1-\frac{\alpha}{2}}, \infty)$$

$$K = (-\infty, -t_{n-1; 1-\alpha}] \text{ bei } \mathcal{H}_A: \mu < \mu_0$$

$$K = [t_{n-1; 1-\alpha}, \infty) \text{ bei } \mathcal{H}_A: \mu > \mu_0$$

6. **Testentscheid:**  
Liegt beobachteter Wert  $t$  der Teststatistik in  $K$

# Lernziele heute

- ungepaarter t-Test
- ungepaarter Wilcoxon-Test (MWU Test)
- multiples Testen

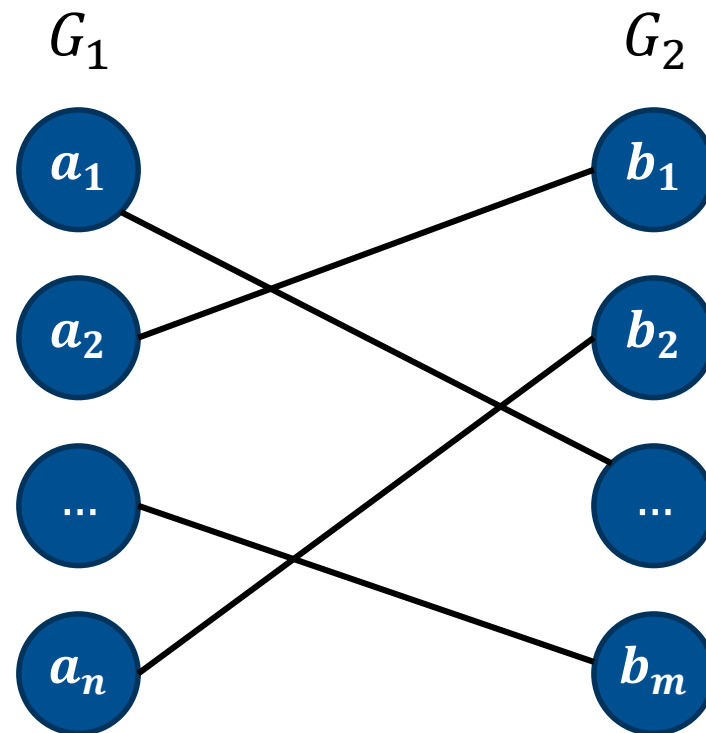
## Hausaufgaben

- Skript: Kapitel 4.8 lesen
- Serie 10 lösen
- Quiz 10 bearbeiten



## 4.8 Tests bei zwei Stichproben

- Zwei gepaarte Stichproben



$$m = n$$

Jeder Beobachtung in  $G_1$  kann eine Beobachtung in  $G_2$  zugeordnet werden.

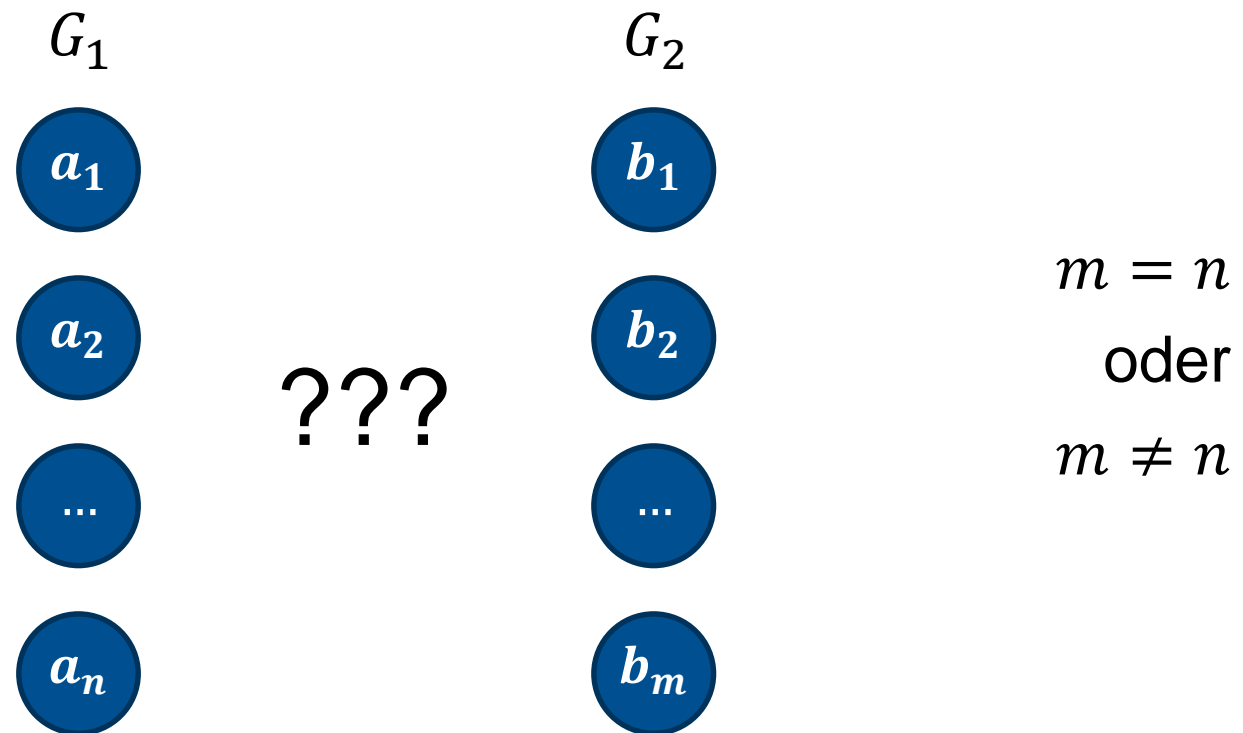
# Zwei gepaarte Stichproben

- Situationen:
  - Vorher/nachher
  - Links/rechts
  - Zwillinge
  - ....
- Überlegung:
  - $n$  Personen  $\rightarrow a_1, a_2, \dots, a_n$  in  $G_1$  und  $b_1, b_2, \dots, b_n$  in  $G_2$
  - Betrachte die Differenzen der Paare:  
 $a_i - b_i = x_i \Rightarrow x_1, x_2, \dots, x_n \rightarrow$  t-Test für eine Stichprobe
- Beispiel Reaktionszeit aus letzter VL; HH - NH



## 4.8 Tests bei zwei Stichproben

- Zwei ungepaarte Stichproben



Eine Beobachtung in  $G_1$  kann keiner Beobachtung in  $G_2$  zugeordnet werden.



# Einfluss von Öl auf aquatische Lebewesen



© Thomas Seilnacht



- Können wir feststellen, ob und ab welcher Konzentration Öl einen Einfluss – und welchen – auf Fische im Wasser hat?

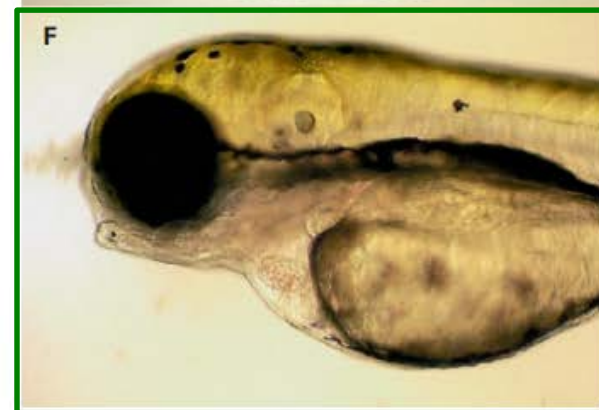
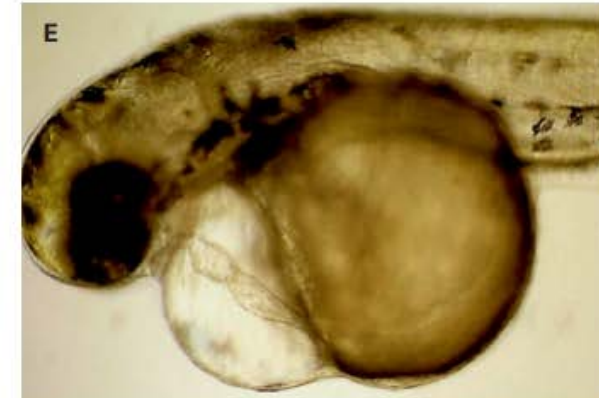
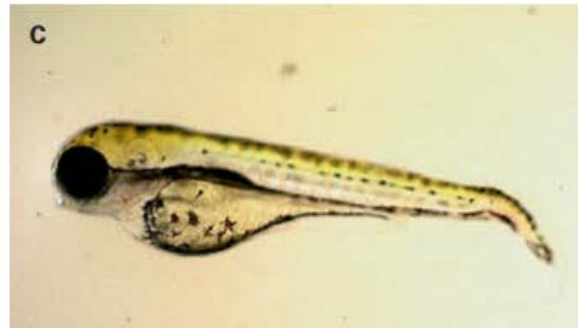


# Was für Schadstoffkonzentrationen?

- Experiment:
  - Embryonen von Zebrafärblingen
  - Unterschiedliche Zeitpunkte der Exposition
    - nach 4h, nach 24h und nach 96h (für jeweils 24h)
  - Unterschiedliche Konzentrationen von Rohöl
    - Verdünnungen von 0.5 – 1000 ppm («parts per million»)
    - Konzentrationen im subakuten Bereich (nicht letal,  $\geq 40\%$  morphologische Veränderungen)
- Auswertung:
  - Zebrafärblinge unter Mikroskop nach Anomalien untersuchen
- Resultat:
  - Morphologische Veränderungen deutlich weniger ausgeprägt bei niedrigen Konzentrationen

# «Sichtbare» morphologische Veränderungen

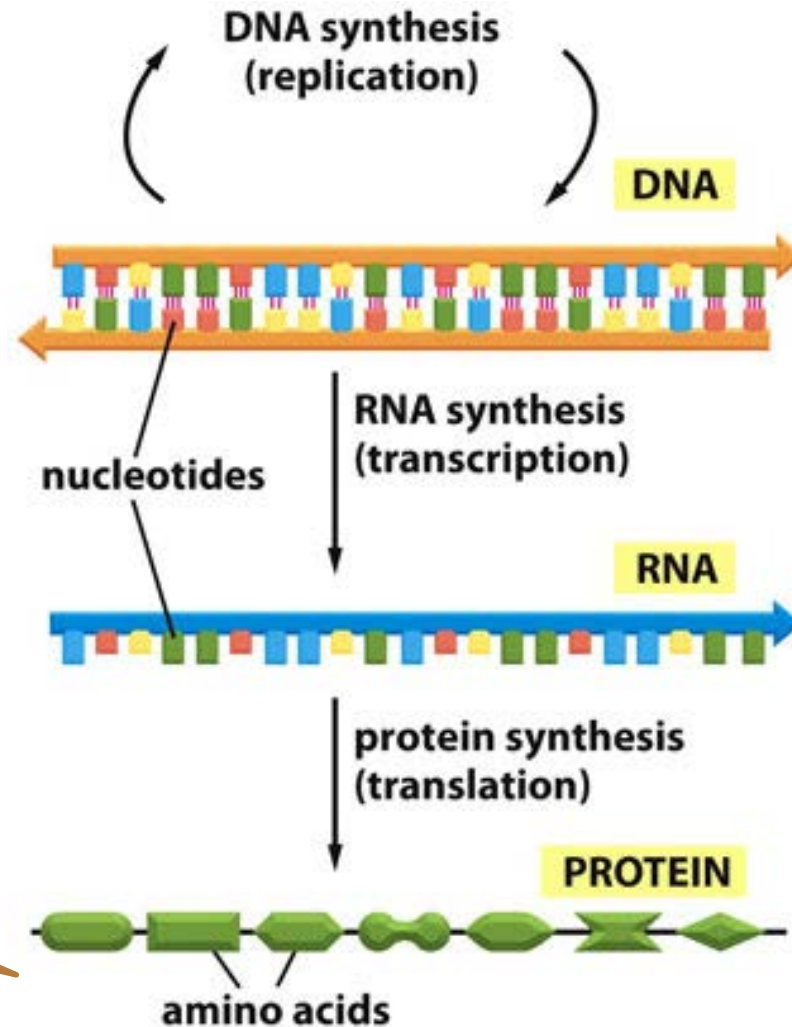
Abb. 1: Morphologische Veränderungen nach Exposition mit Rohöl oder dem wässrigen Extrakt.  
A: unbehandelte Kontrolle, B: gekrümmter Schwanz, C: abgeknickter Schwanz,  
D: Ödem und abgeknickter Schwanz, E: vergrößertes Herz, F: unbehandelte Kontrolle, normales Herz



## ...und «nicht sichtbare» Veränderungen?

- Eine Konzentration von 100ppm oder weniger scheint nicht so einen grossen Einfluss auf die Morphologie der Fische zu haben
- Wie sieht das auf genetischer Ebene aus?
- Wissenschaftliche Fragestellung:
  - *Haben Zebraabärblinge, welche kurz nach ihrer Befruchtung Rohöl ausgesetzt werden, eine **veränderte Genexpression** im Vergleich zu denjenigen, welche keinem Rohöl ausgesetzt werden (Kontrolle)?*
  - *Und bei welchen Konzentrationen können wir das nachweisen?*
- Was sind die Daten, welche wir für diese Fragestellung brauchen?

# Zentrales Dogma der Molekularbiologie



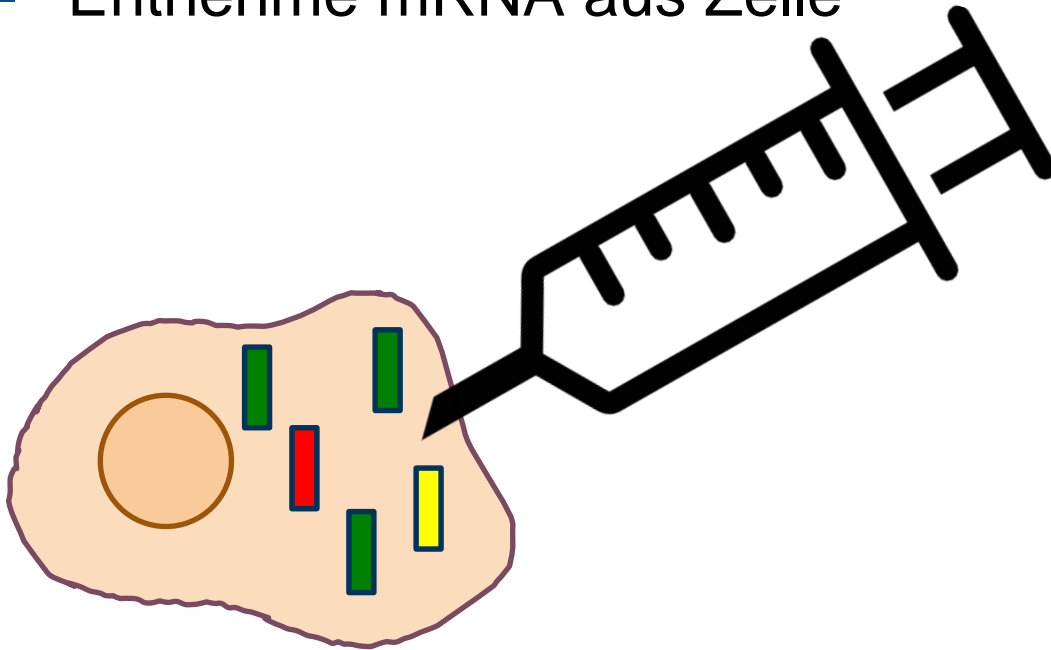
Proteine zu messen ist ungleich schwieriger

misst man die Menge an mRNA, weiss man wie aktiv ein Gen ist

Figure 1-2 Essential Cell Biology 3/e (© Garland Science 2010)

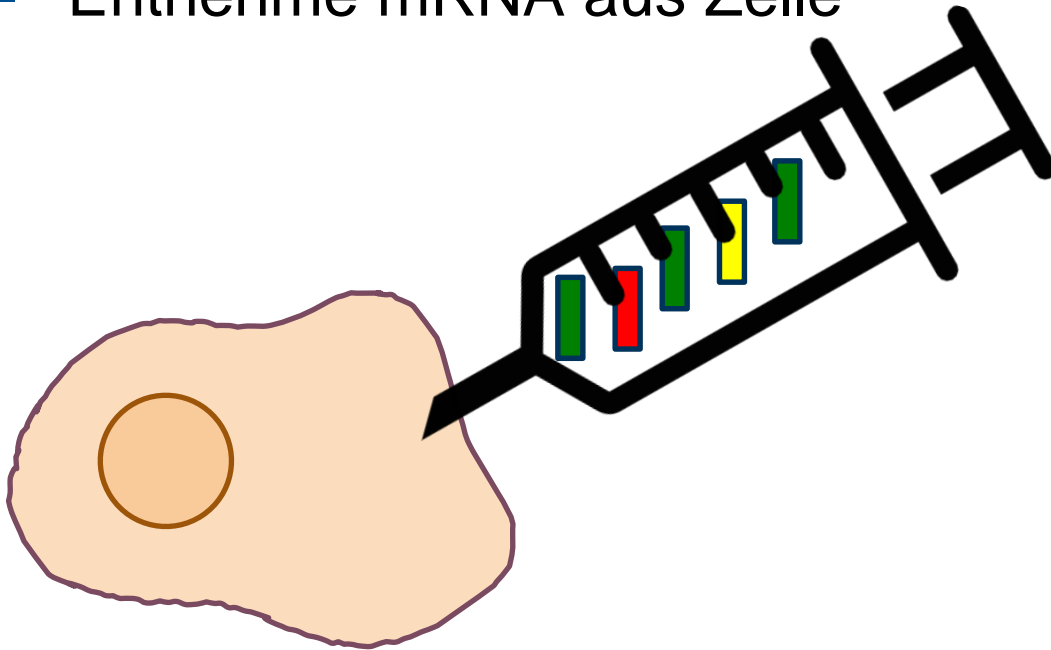
# Wie messen wir die Genexpression?

- Entnahme mRNA aus Zelle



# Wie messen wir die Genexpression?

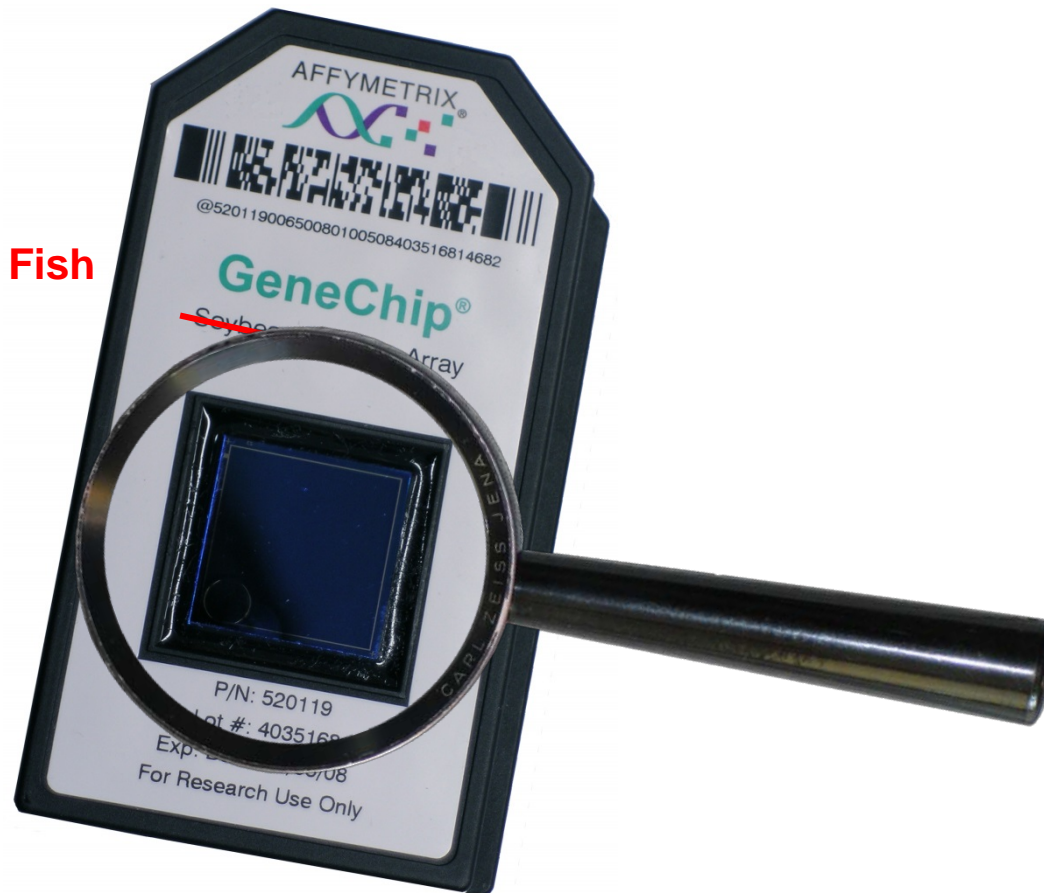
- Entnahme mRNA aus Zelle



# Wie messen wir die Genexpression?

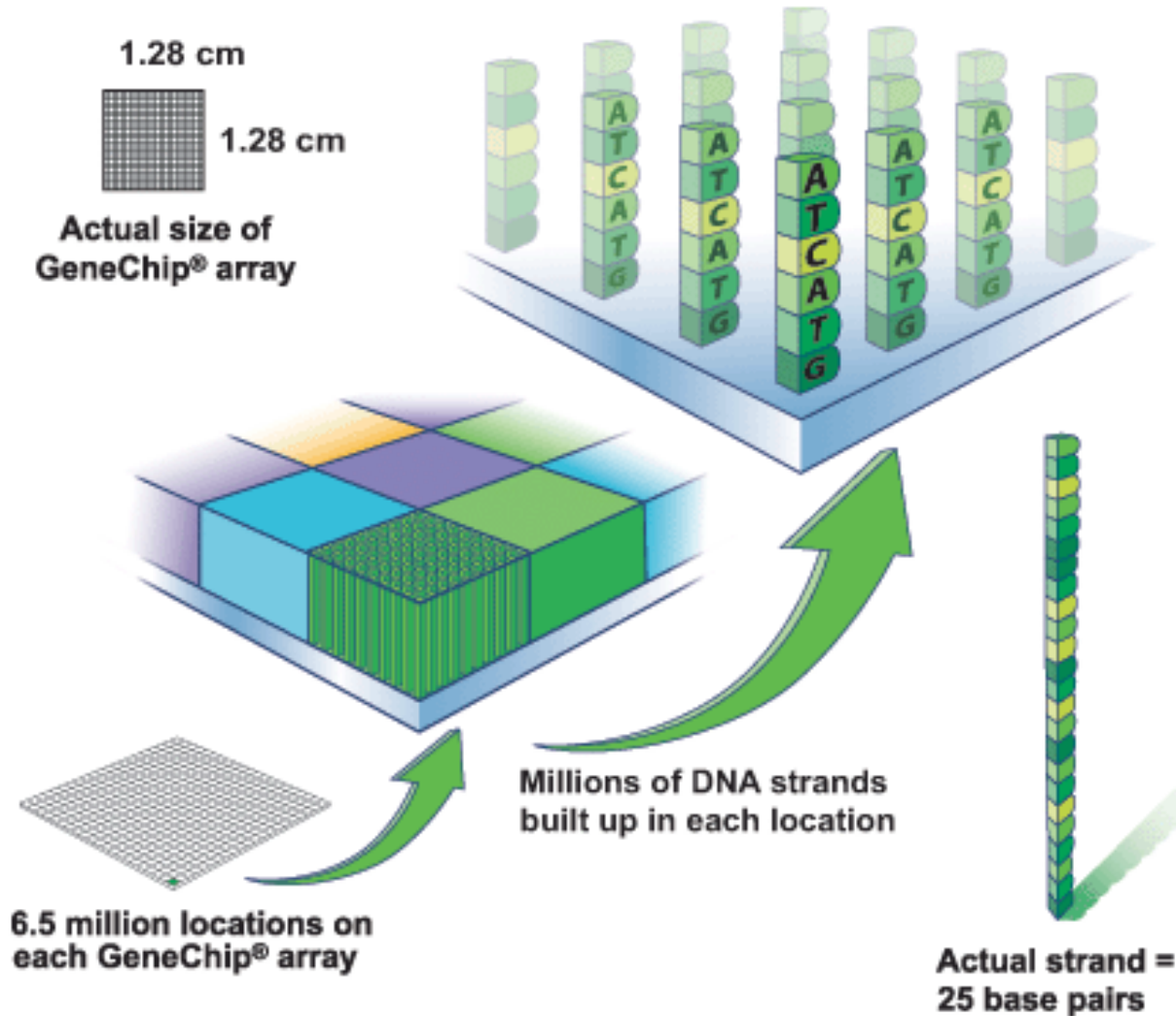
- Nehme einen «Microarray»

Zebra Fish



# Wie sieht ein Microarray aus?

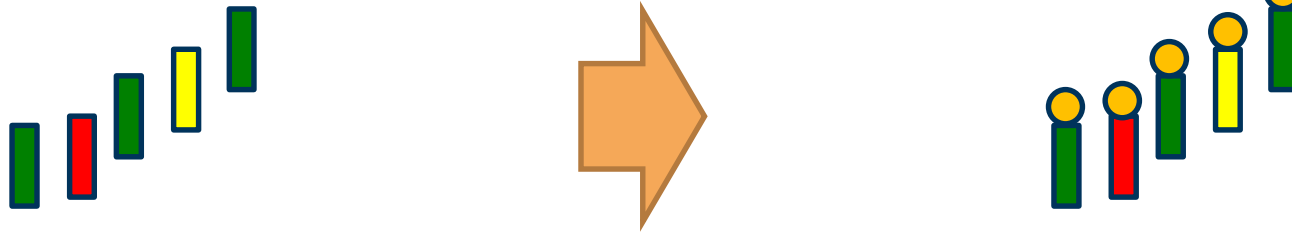
From Computer Desktop Encyclopedia  
Reproduced with permission.  
© 2007 Affymetrix





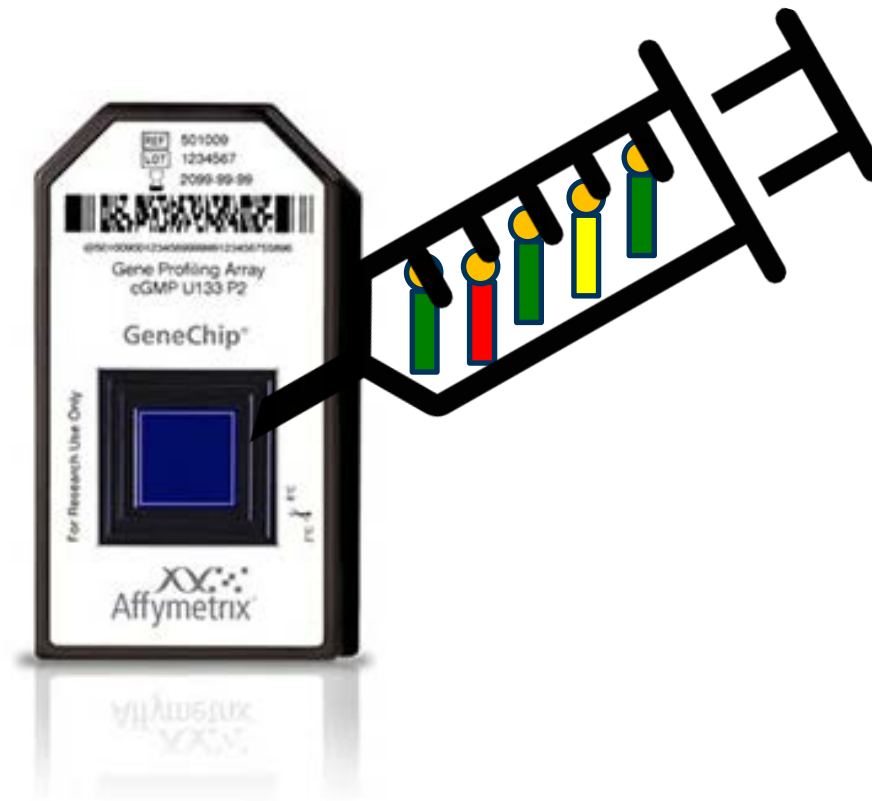
# Wie messen wir die Genexpression?

- Klebe an die mRNA Schnipsel ein «Glow-In-The-Dark» Protein an



# Wie messen wir die Genexpression?

- mRNA auf Microarray



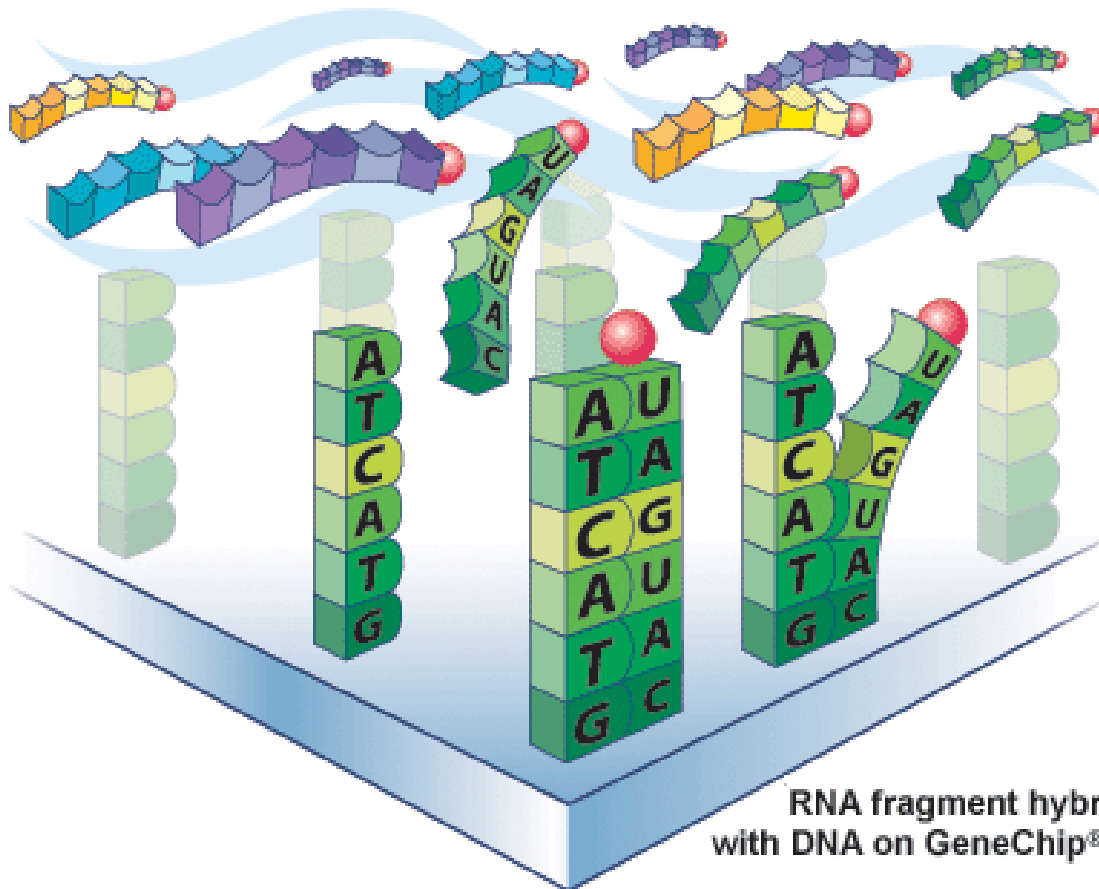
# Wie messen wir die Genexpression?

- mRNA auf Microarray



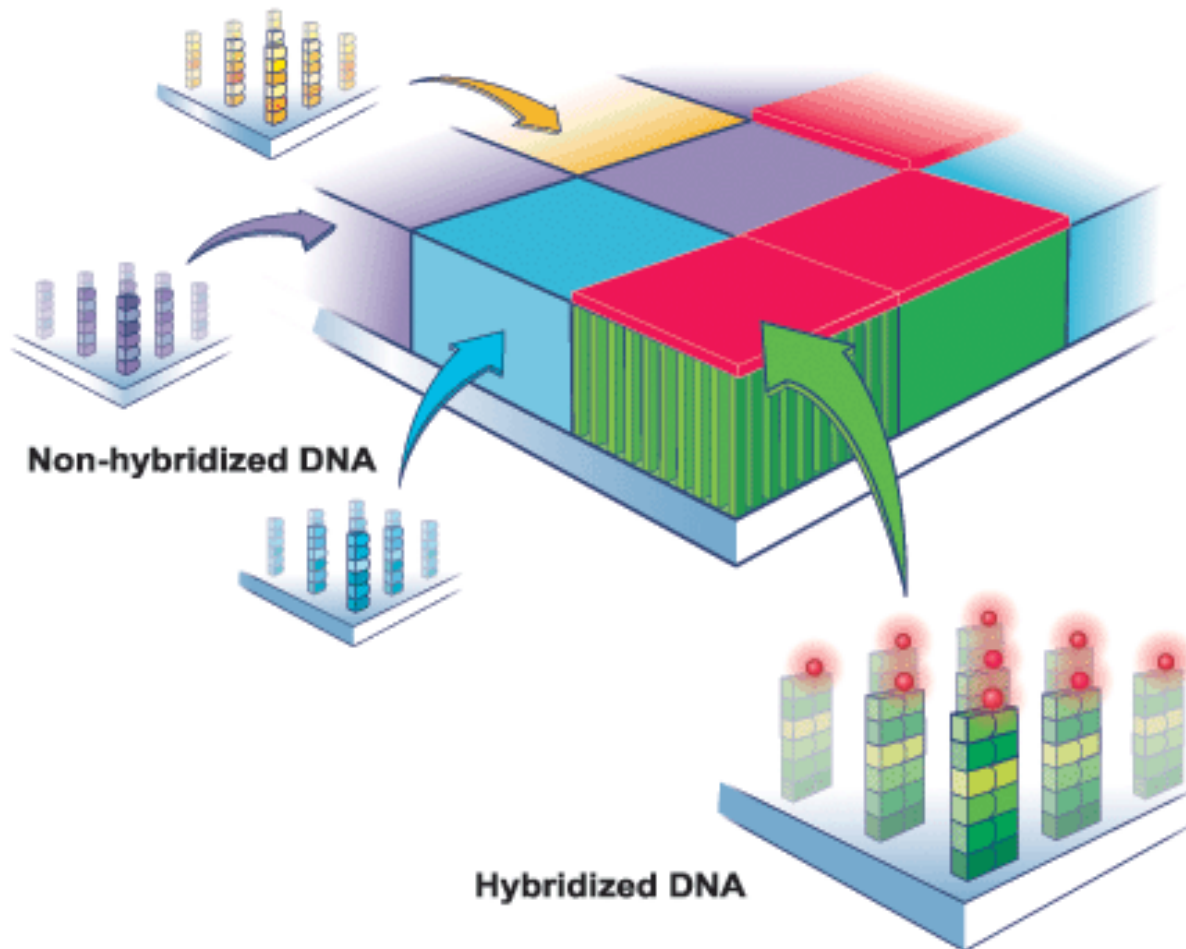
# Was passiert auf dem Microarray?

RNA fragments with fluorescent tags from sample to be tested



# Voilà: Ein Feuerwerk!

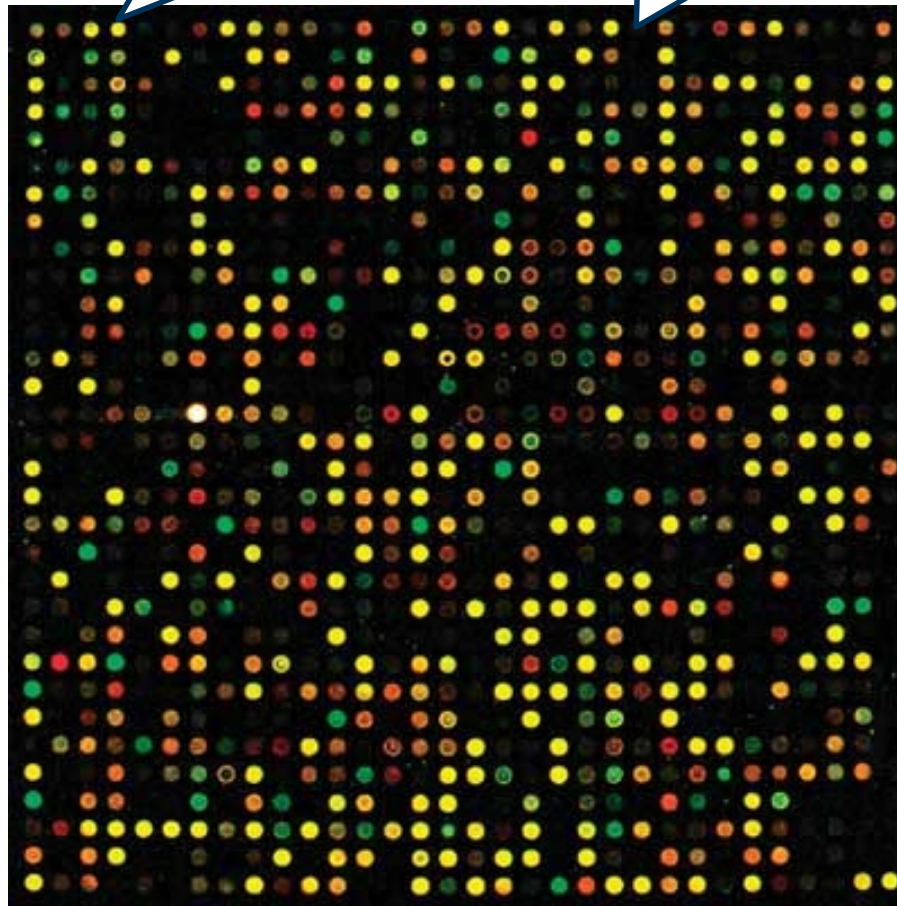
Shining a laser light at GeneChip® array causes tagged DNA fragments that hybridized to glow



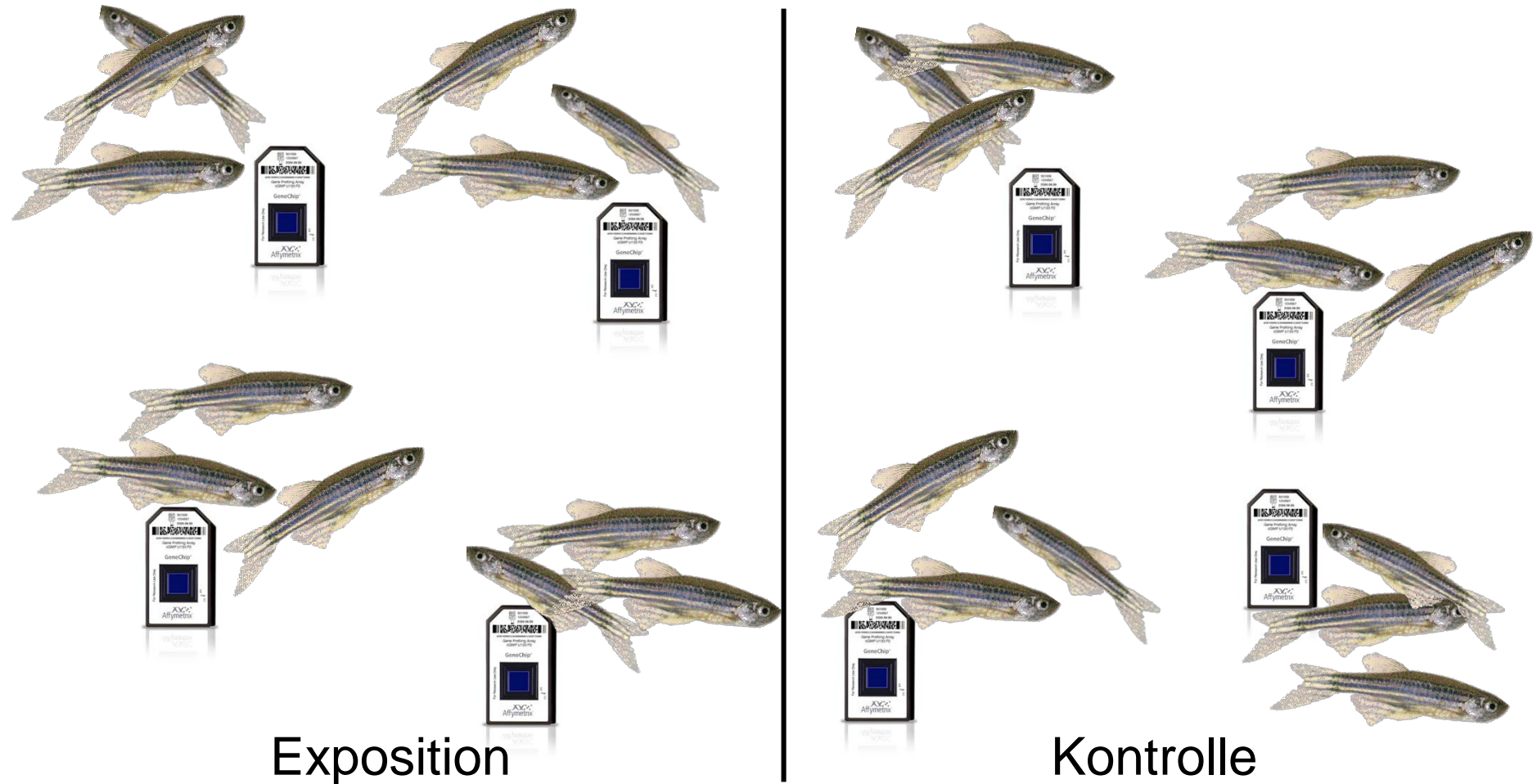
# Helligkeit = Aktivität eines Gens

Gen 4 sehr aktiv

Gen 23 nicht aktiv



# Man misst die Genexpression in den Fischen



# Aktivität aller Gene in den Fischen

Kontrolle				
Gen	Rep. 1	Rep. 2	...	Rep. n
1	2.1	1.9	...	2.3
2	2.4	2.0	...	2.2
...	...	...	...	...
50'000	1.9	1.8	...	2.1

Exposition				
Gen	Rep. 1	Rep. 2	...	Rep. m
1	1.8	2.2	...	2.0
2	2.7	2.9	...	3.0
...	...	...	...	...
50'000	1.7	2.0	...	1.9



# Aktivität aller Gene in den Fischen

Kontrolle				
Gen	Rep. 1	Rep. 2	...	Rep. n
1	2.1	1.9	...	2.3
2	2.4	2.0	...	2.2
...	...	...	...	...
50'000	1.9	1.8	...	2.1

Exposition				
Gen	Rep. 1	Rep. 2	...	Rep. m
1	1.8	2.2	...	2.0
2	2.7	2.9	...	3.0
...	...	...	...	...
50'000	1.7	2.0	...	1.9

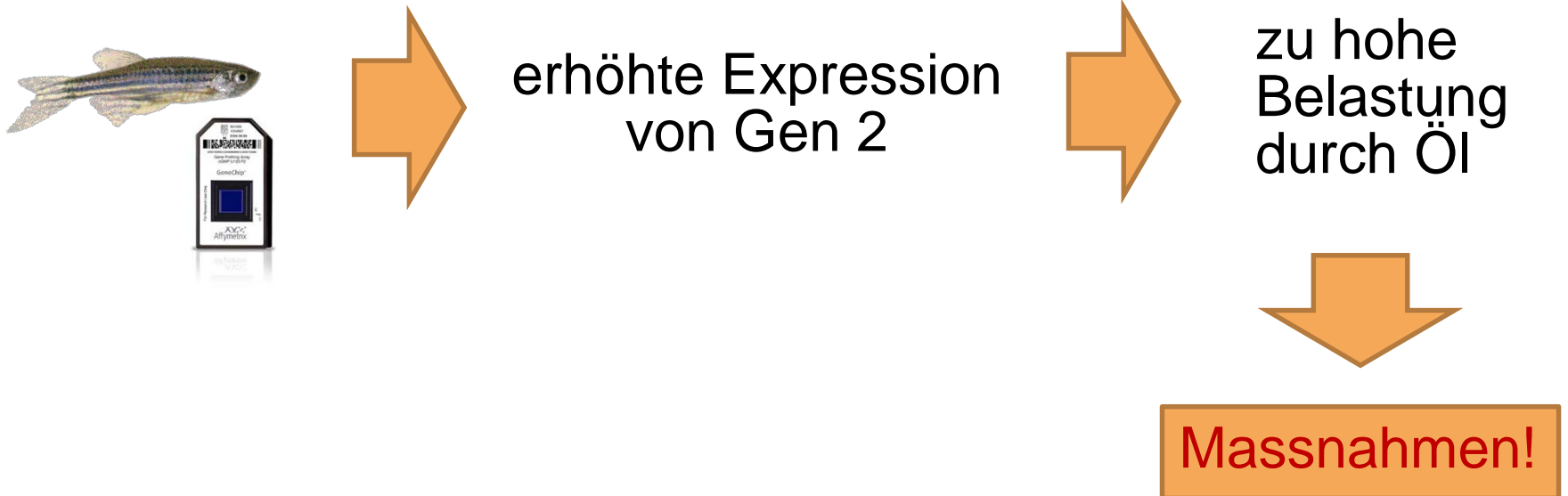
# Ist die Aktivität von Gen 2 **signifikant** höher?

Kontrolle				
Gen	Rep. 1	Rep. 2	...	Rep. n
1	2.1	1.9	...	2.3
2	2.4	2.0	...	2.2
...	...	...	...	...
50'000	1.9	1.8	...	2.1

Exposition				
Gen	Rep. 1	Rep. 2	...	Rep. m
1	1.8	2.2	...	2.0
2	2.7	2.9	...	3.0
...	...	...	...	...
50'000	1.7	2.0	...	1.9

## Falls ja:

- Verwende Gen 2 als *Biomarker* für zukünftige Kontrollen



# Ist die Aktivität von Gen 2 **signifikant** höher?

Kontrolle				
Gen	Rep. 1	Rep. 2	...	Rep. n
1	2.1	1.9	...	2.3
2	2.4	2.0	...	2.2
...	...	...	...	...
50'000	...	...	...	...

**ungepaarter t-Test**

Exp				
Gen	Rep. 1	Rep. 2	...	Rep. n
1	1.8	2.2	...	2.0
2	2.7	2.9	...	3.0
...	...	...	...	...
50'000	1.7	2.0	...	1.9

# Ungepaarter t-Test (1/3)

## 1. Modell:

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim \mathcal{N}(\mu_X, \sigma^2)$$

$$Y_1, Y_2, \dots, Y_m \text{ i. i. d. } \sim \mathcal{N}(\mu_Y, \sigma^2)$$

## 2. Nullhypothese:

$$\mathcal{H}_0: \mu_X = \mu_Y$$

### Alternative:

$$\mathcal{H}_A: \mu_X \neq \mu_Y$$

$$\mathcal{H}_A: \mu_X > \mu_Y$$

$$\mathcal{H}_A: \mu_X < \mu_Y$$

## Ungepaarter t-Test (2/3)

### 3. Teststatistik:

$$\bar{x}_n = \frac{1}{n} \sum x_i$$

$$T = \frac{\bar{X}_n - \bar{Y}_m}{S_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$\hat{\sigma}_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x}_n)^2$$

wobei

$$\begin{aligned} S_{pool}^2 &= \frac{1}{n+m-2} \left( \sum_{i=1}^n (X_i - \bar{X}_n)^2 + \sum_{i=1}^m (Y_i - \bar{Y}_m)^2 \right) = \\ &= \frac{1}{n+m-2} \left( (n-1)\hat{\sigma}_x^2 + (m-1)\hat{\sigma}_y^2 \right) \end{aligned}$$

**Verteilung der Teststatistik unter  $\mathcal{H}_0$ :  $T \sim t_{n+m-2}$ .**

## Ungepaarter t-Test (3/3)

4. Signifikanzniveau:  $\alpha$

5. Verwerfungsbereich der Teststatistik:

$$K = \left(-\infty, -t_{n+m-2; 1-\alpha/2}\right] \cup \left[t_{n+m-2; 1-\alpha/2}, \infty\right) \text{ bei } \mathcal{H}_A: \mu_X \neq \mu_Y$$

$$K = \left[t_{n+m-2; 1-\alpha}, \infty\right) \text{ bei } \mathcal{H}_A: \mu_X > \mu_Y$$

$$K = \left(-\infty, -t_{m+n-2; 1-\alpha}\right] \text{ bei } \mathcal{H}_A: \mu_X < \mu_Y$$

6. Testentscheid: Liegt der beobachtete Wert  $t$  von  $T$  in  $K$

## Beispiel: Microarray, ungepaarter t-Test bei Gen 2

- $n = 5, m = 4$
- $\bar{x} = 1.58, \bar{y} = 2.43$
- $\hat{\sigma}_x = 0.40, \hat{\sigma}_y = 0.41$

### 1. Modell:

$$X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$$

$$Y_1, Y_2, \dots, Y_m \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

### 2. $\mathcal{H}_0: \mu_x = \mu_y, \mathcal{H}_A: \mu_x \neq \mu_y$

### 3. Teststatistik:

$$T = \frac{\sqrt{n+m}(\bar{x}_n - \bar{y}_m)}{S_{pool}^2}$$

$$S_{pool}^2 = \frac{1}{7} (4 \cdot 0.40^2 + 3 \cdot 0.41^2) \approx 0.16$$

$$\Rightarrow S_{pool} = \sqrt{0.16} = 0.40$$

$$\text{falls } \mathcal{H}_0: T \sim t_{n+m-2} = t_7$$

### 4. Signifikanz: $\alpha = 0.05$

### 5. Verwerfungsbereich:

$$K = (-\infty, -t_{7;0.975}] \cup [t_{7;0.975}, \infty) =$$

$$= (-\infty, -2.36] \cup [2.36, \infty)$$

### 6. Testentscheid:

$$t = \frac{\sqrt{5+4}(1.58 - 2.43)}{0.40}$$

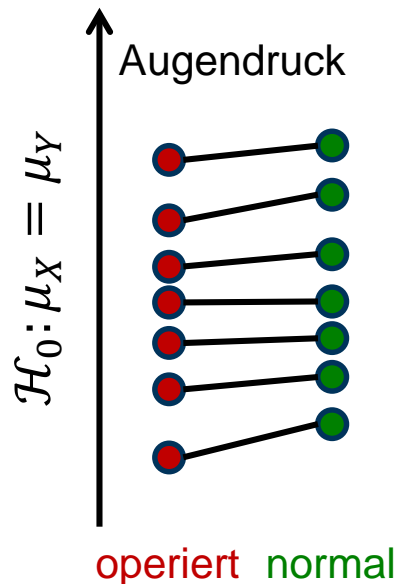
$$\approx -3.17 \Rightarrow t \in K$$

$\mathcal{H}_0$  wird verworfen



# Gepaart versus ungepaart

- Bsp.:
  - Augeninnendruck (ein Auge operiert, das andere nicht), gepaarter Test ist angebracht
  - Gemäss Voraussetzungen könnte auch ein ungepaarter Test angewendet werden



**Ungepaart**

Teststatistik:

$$T = \frac{\bar{X} - \bar{Y}}{\hat{\sigma}_{\bar{X}}}$$

**Gepaart**

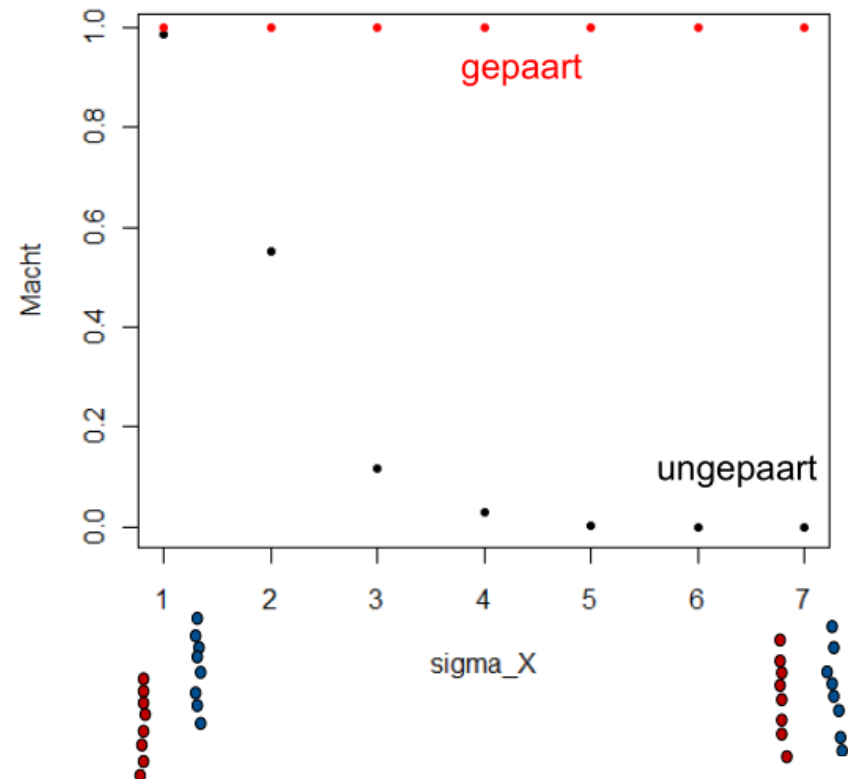
$$D_i = X_i - Y_i$$

Teststatistik:

$$T = \frac{\bar{D}}{\hat{\sigma}_{\bar{D}}}$$

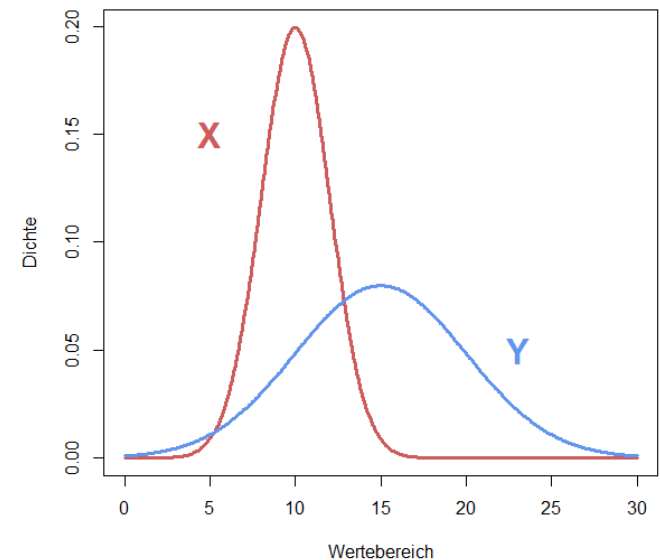
# Gepaart versus ungepaart: Simulationsstudie

- $\mathcal{H}_0: \mu_D = 0$  bzw.  $\mathcal{H}_0: \mu_X = \mu_Y$ ;  $n = m = 10$
- $X \sim \mathcal{N}(100, \sigma_X^2)$ ,  $D \sim \mathcal{N}(2, 1)$ ,  $Y = X + D$  : gepaartes Setup
- Der **gepaarte** t-Test hat **mehr Macht**, wenn die Daten verrauscht sind



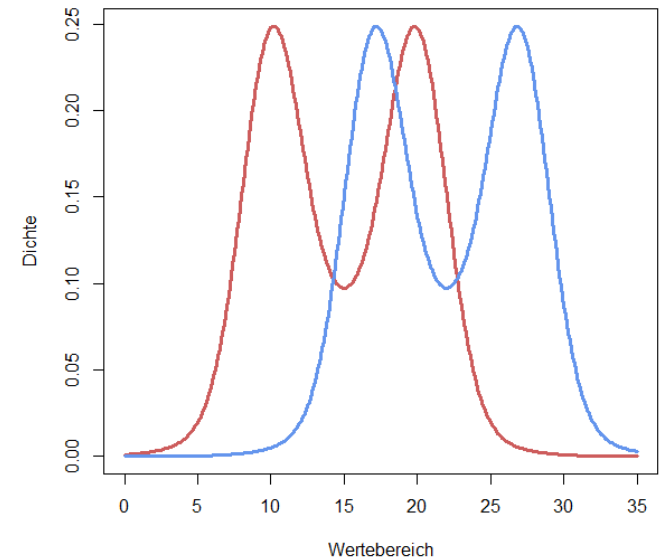
# t-Test falls Varianz in den Gruppen verschieden

- heisst auch: Welsh-Test
- Grundidee identisch
- Teststatistik und Verteilung, falls  $\mathcal{H}_0$  stimmt, ist komplizierter
- Computer: Meist der default t-Test
- Praxis: Man sollte immer annehmend, dass die Varianz in den Gruppen unterschiedlich ist → Welsh Test
- Prüfung: Wir nehmen der Einfachheit halber an, dass die Varianz jeweils gleich ist → t-Test



## Two-sample Wilcoxon Test (a.k.a. Mann-Whitney U-Test)

- Falls Daten **nicht** normalverteilt
- $X_i \sim F, i = 1, 2, \dots, n;$
- $Y_j \sim G, i = 1, 2, \dots, m$
- $\mathcal{H}_0: F = G$
- $\mathcal{H}_A: F = G + \delta$ , mit  $\delta \neq 0$  (oder einseitig)  
d.h. Verteilungen sind verschoben, haben aber gleiche Form
- Teststatistik:
  - Bilde Ränge über beide Gruppen hinweg
  - Falls Gruppen gleich, sollten Rangsummen etwa gleich sein
  - Falls Gruppen ungleich, sollten die Rangsummen in einem gewissen Verhältnis stehen ( $\neq 1$ )



## Beispiel: Two-sample Wilcoxon Test

- Behandlung (Trt) und Kontrolle (Contr) je 2 Patienten
- Beobachtung: Trt: 1.2, 3.1; Contr: 5.9, 4.4
- Ränge: Trt: 1, 2; Contr: 4, 3
- Rangsumme R in Contr:  $4 + 3 = 7$
- Falls  $\mathcal{H}_0$  stimmt, sind alle Ränge in Contr gleich wahrscheinlich

Ränge	1, 2	1, 3	1, 4	2, 3	2, 4	3, 4
R	3	4	5	5	6	7

- z.B. für einseitigen Test:

$$P[R \geq 7] = P[R = 7] = \frac{1}{6} \approx 0.167$$

- $\mathcal{H}_0$  kann auf dem 5% Niveau nicht verworfen werden

# Übersicht der Tests für ungepaarte Stichproben

	Annahmen				$n_{min}$ (falls $n=m$ ) bei $\alpha = 0.05$	Macht für Beispiel
	$\sigma_X = \sigma_Y$	$X_i \sim \mathcal{N}$ $Y_j \sim \mathcal{N}$	$F, G$ haben gleiche Form	i.i.d.		
<b>t-Test</b>	●	●	●	●	2	57%
<b>Welsh-Test</b>		●		●	2	56%
<b>Wilcoxon</b>	●		●	●	4	53%

Verwendetes Beispiel:

- $X_i \sim \mathcal{N}(\mu_X, \sigma^2), n = 10$
- $Y_j \sim \mathcal{N}(\mu_Y, \sigma^2), m = 10$
- $\mathcal{H}_0: \mu_X = \mu_Y; \mathcal{H}_A: \mu_X \neq \mu_Y; \alpha = 0.05$
- Macht berechnet mit konkreter Alternative:  $X_i \sim \mathcal{N}(0,1), Y_j \sim \mathcal{N}(1,1)$

# Multiples Testen

- Microarray mit  $m = 1000$  Genen
- D.h. wir müssen 1000 t-Test auf dem 5% Niveau machen
- Angenommen, kein Gen hat einen Effekt
- ca. 50 Tests (5% von 1000) werden trotzdem ein signifikantes Ergebnis liefern ( $\mathcal{H}_0$  verwerfen)
- D.h. 50 Gene werden als «wichtig» angegeben, obwohl sie gar nicht «wichtig» sind
- Wie bekommt man eine Liste mit Genen, die «wirklich wichtig» sind?

# Multiple Testen: Bonferroni Korrektur

- Wollen eine konservative Liste mit der Eigenschaft

$$P[\text{mind. ein Fehler 1. Art}] \leq \alpha$$

- Bonferroni Korrektur: Teste jedes Gen mit Signifikanzniveau  $\frac{\alpha}{m}$  statt  $\alpha$ ; wobei  $m$  die Anzahl Gene

- Begründung:

$F_i$ : Fehler 1. Art bei Gen  $i$

$$P\left(\bigcup_{i=1}^m F_i\right) \leq \sum_{i=1}^m P[F_i] = \sum_{i=1}^m \frac{\alpha}{m} = \alpha$$

- Nachteil: Liste evtl. extrem konservativ (z.B. gar kein Gen enthalten)



# Zusammenfassung

- ungepaarter t-Test – differentielle Genexpression bei Zebrafischen
- ungepaarter Wilcoxon-Test (MWU Test) – Nicht ganz Normal?
- multiples Testen – **CAVEAT!!**

## Hausaufgaben

- Skript: Kapitel 4.8 lesen
- Serie 10 lösen
- Quiz 10 bearbeiten

