

Gepaarter und ungepaarter t-Test

Statistik (Biol./Pharm./HST) – FS 2014



Wdh: t-Test für eine Stichprobe

1. **Modell:** X_i ist eine kontinuierliche Messgröße;
 $X_1, \dots, X_n \text{ iid } \mathcal{N}(\mu, \sigma_X^2)$, σ_X wird durch $\widehat{\sigma}_X$ geschätzt
2. **Nullhypothese:** $H_0 : \mu = \mu_0$,
Alternative: $H_A : \mu \neq \mu_0$ (oder “<” oder “>”)
3. **Teststatistik:**

$$T = \frac{(\bar{X}_n - \mu_0)}{\widehat{\sigma}_{\bar{X}_n}} = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\widehat{\sigma}_X} = \frac{\text{beobachtet} - \text{erwartet}}{\text{geschätzter Standardfehler}}.$$

Verteilung der Teststatistik unter H_0 : $T \sim t_{n-1}$

4. **Signifikanzniveau:** α
5. **Verwerfungsbereich für die Teststatistik:**

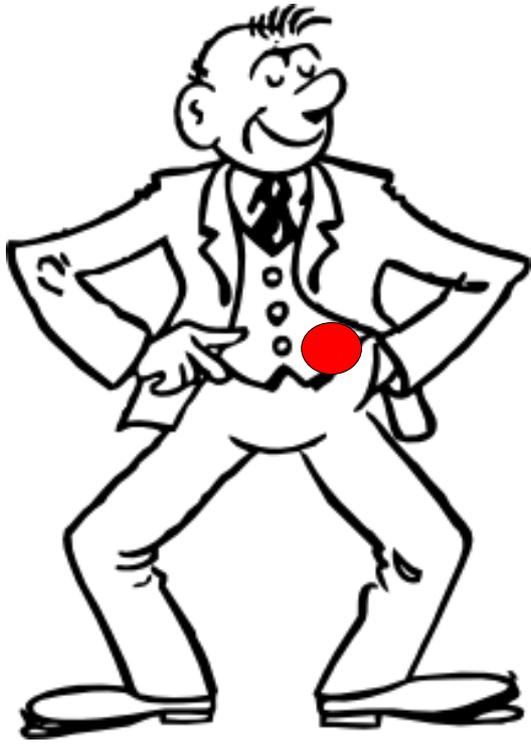
$$K = (-\infty, -t_{n-1; 1-\frac{\alpha}{2}}] \cup [t_{n-1; 1-\frac{\alpha}{2}}, \infty) \text{ bei } H_A : \mu \neq \mu_0,$$

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

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

6. **Testentscheid:** Überprüfe, ob der beobachtete Wert der Teststatistik im Verwerfungsbereich liegt.

Mr. X



Krebs

Zwei Krebstypen

1

Typ 1: Mild

Chemotherapie nicht nötig

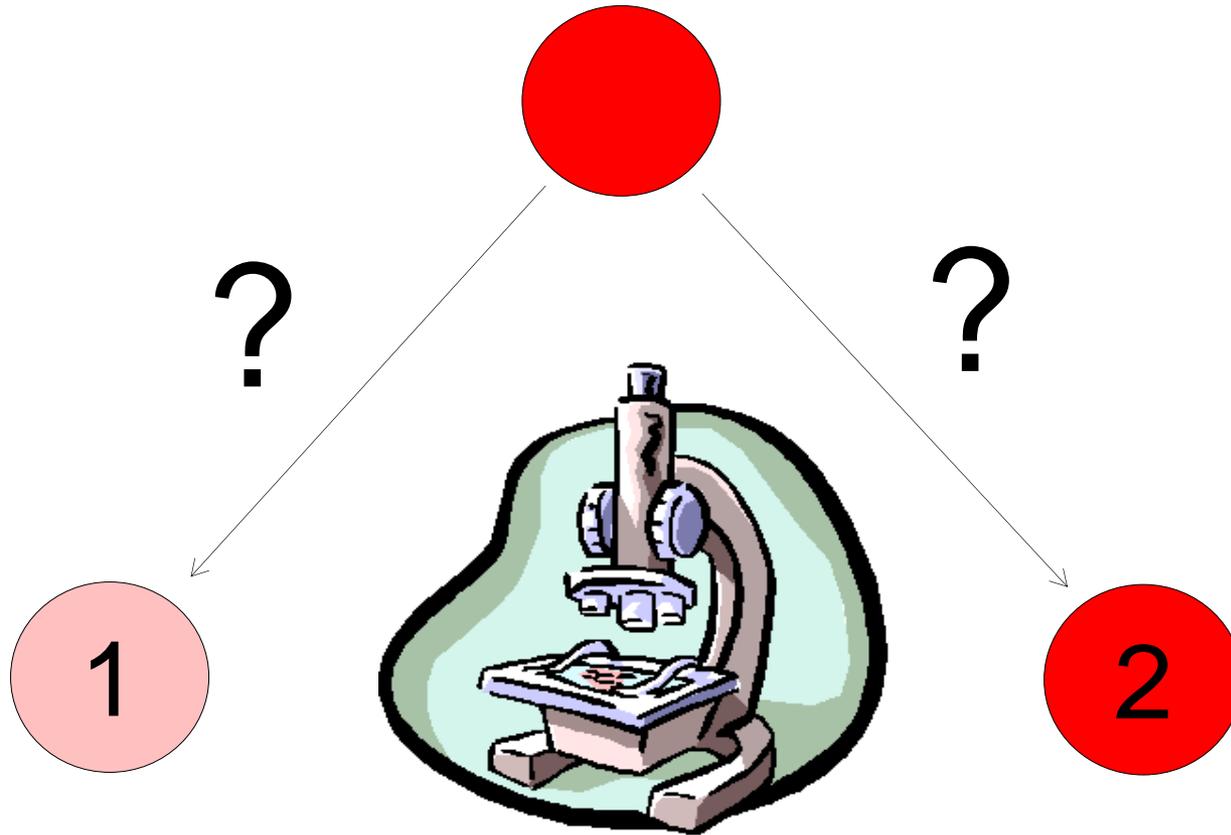
2

Typ 2: **Schwer**

Chemotherapie **nötig**

Problem:

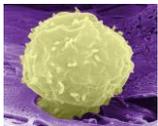
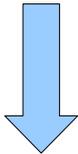
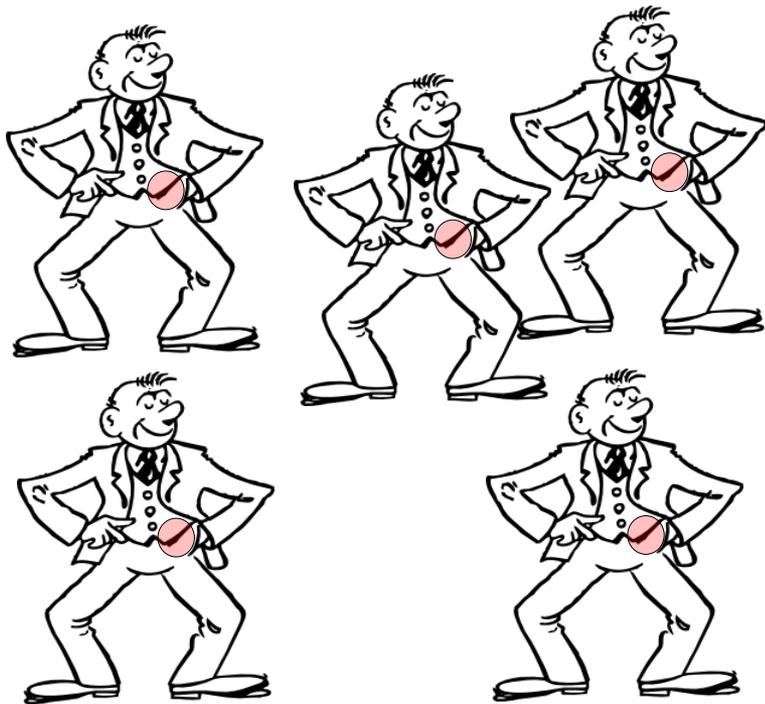
Typ erst nach langer Zeit erkennbar



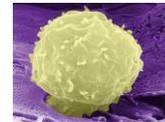
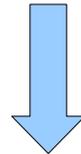
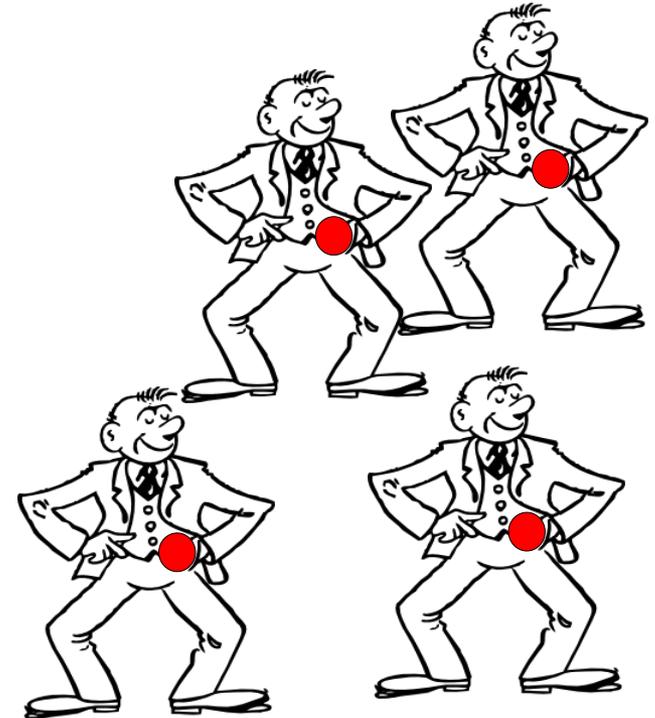
Jetzt **Chemotherapie** oder nicht?

Wie kann man
verschiedene Arten von Krebs
frühzeitig
unterscheiden?

Typ 1



Typ 2

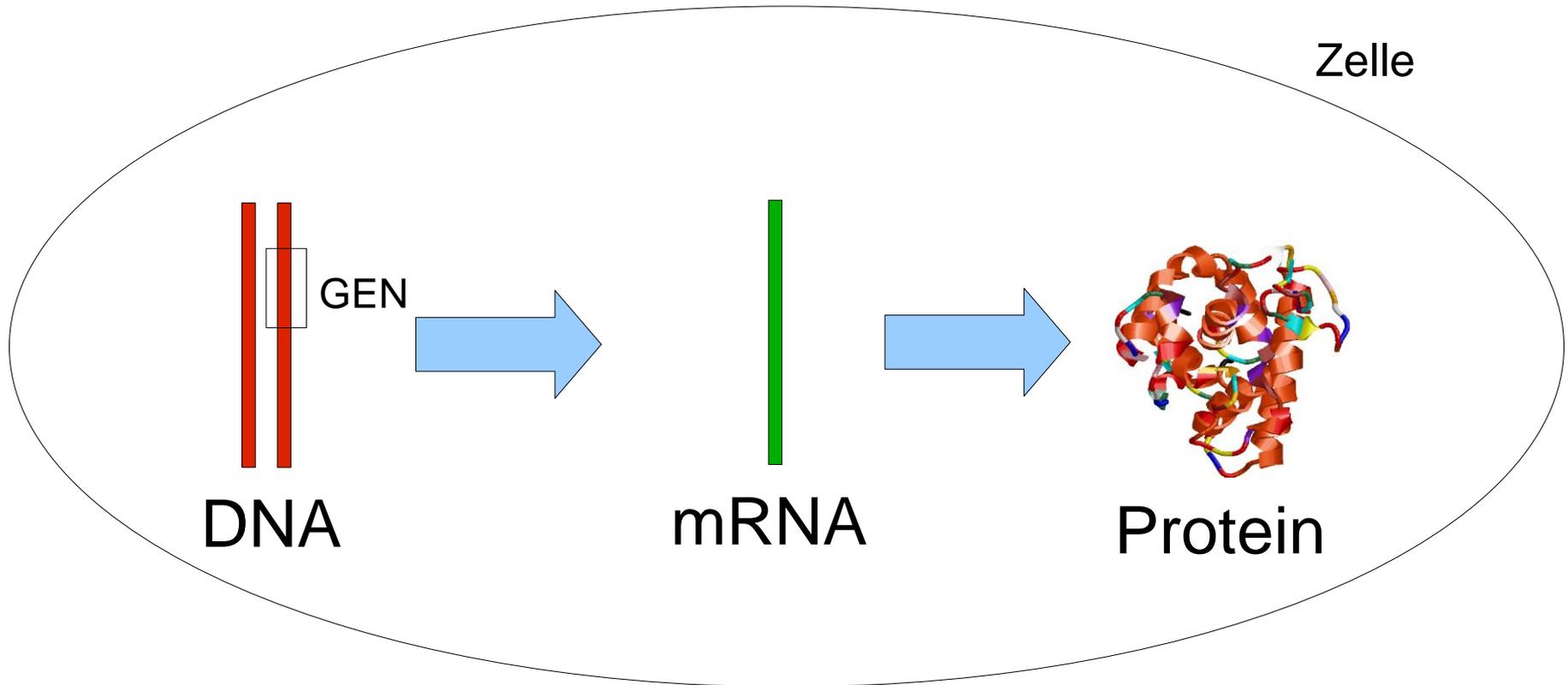


Vergleiche Krebszellen

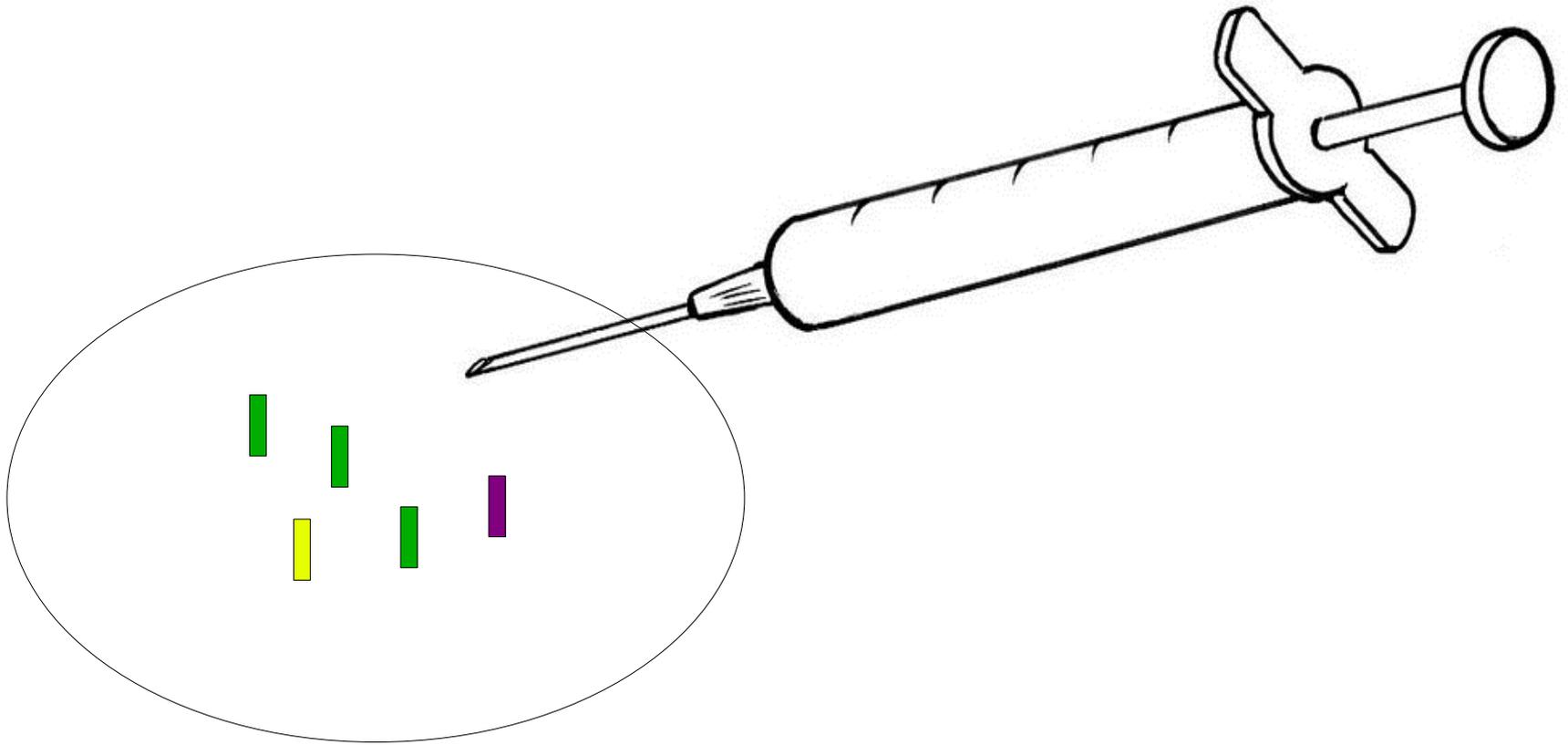
Idee:

Vergleiche Aktivität
innerhalb der Zelle

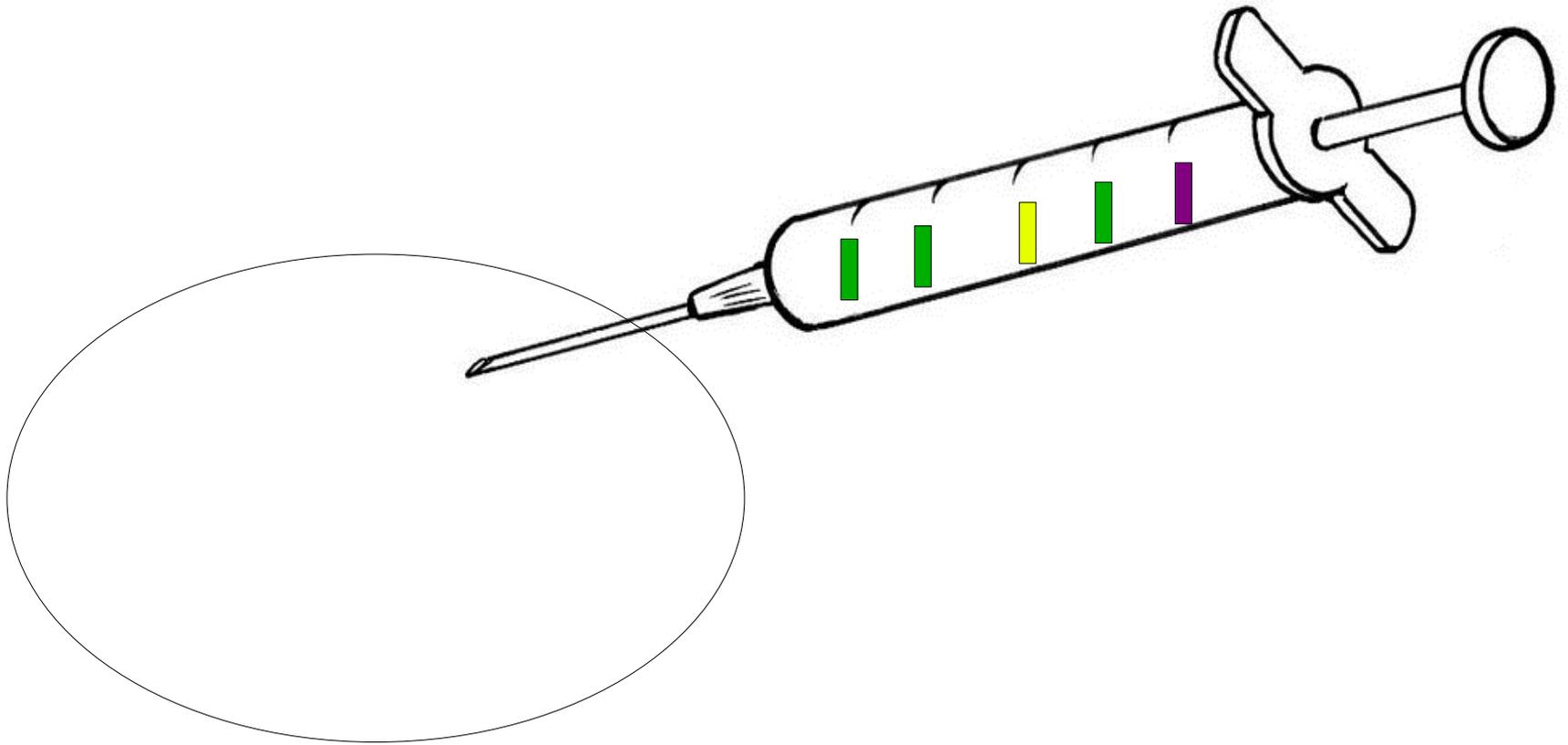
Zentrales Dogma der Molekularbiologie



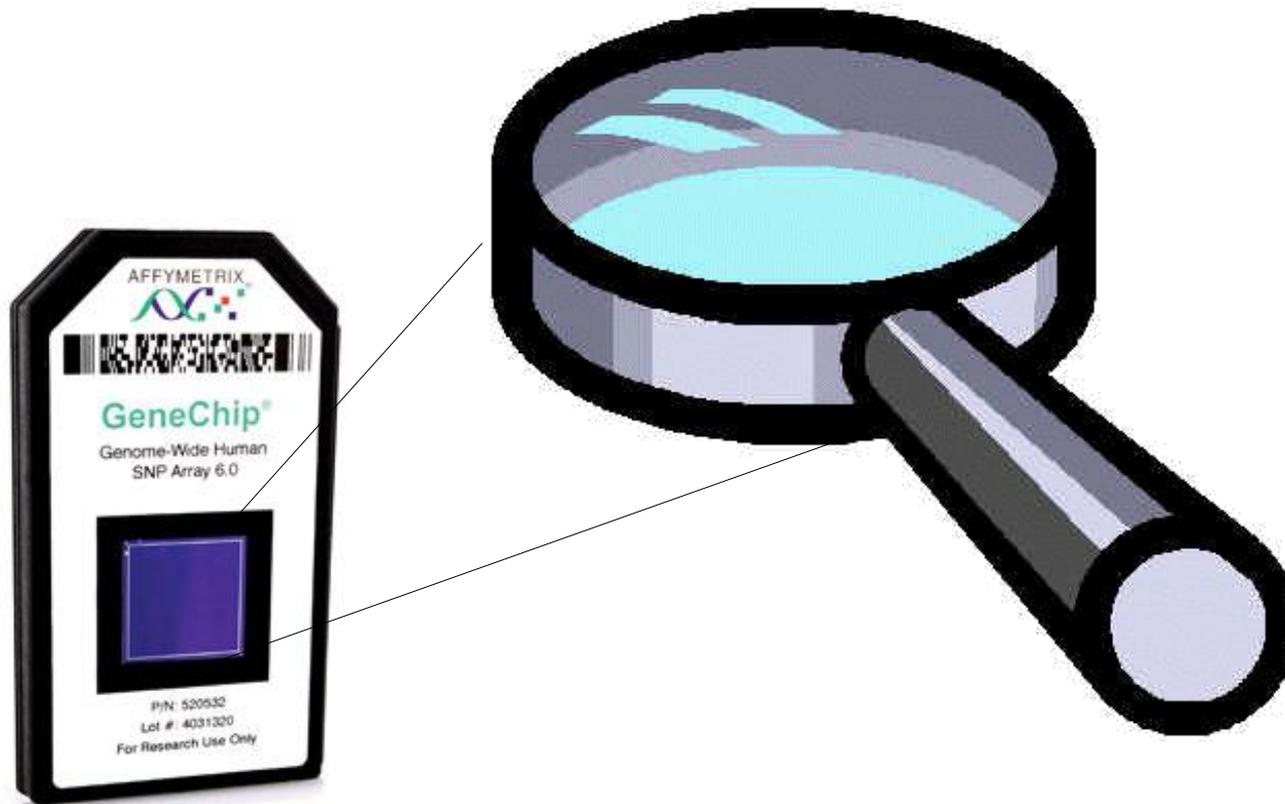
Entnahme mRNA



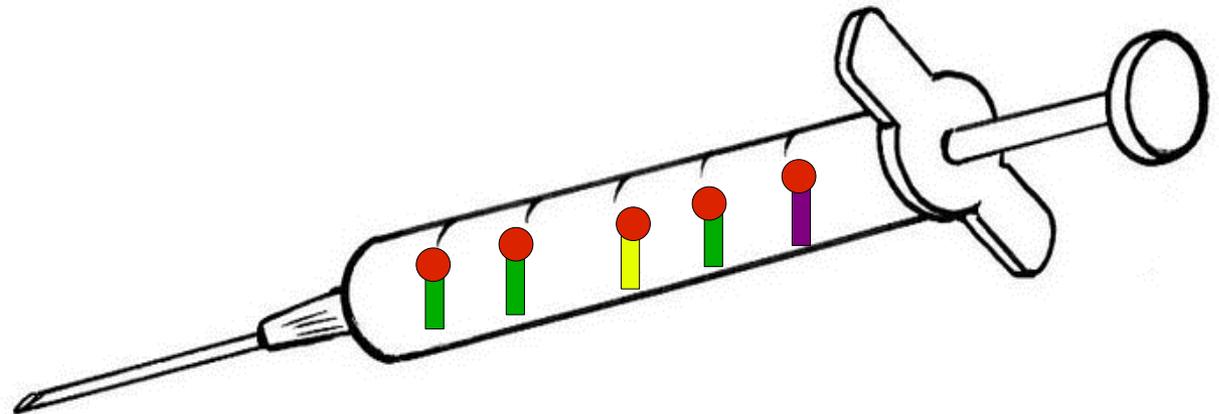
Entnahme mRNA



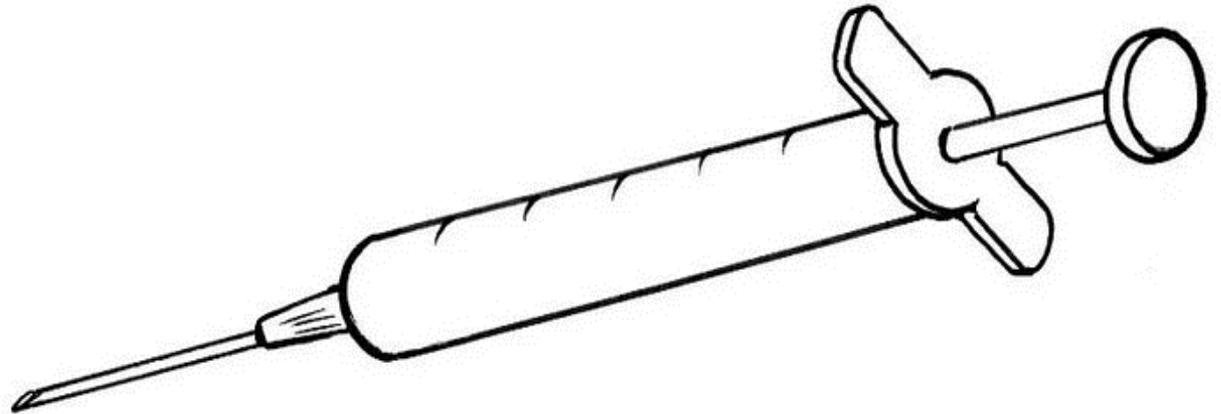
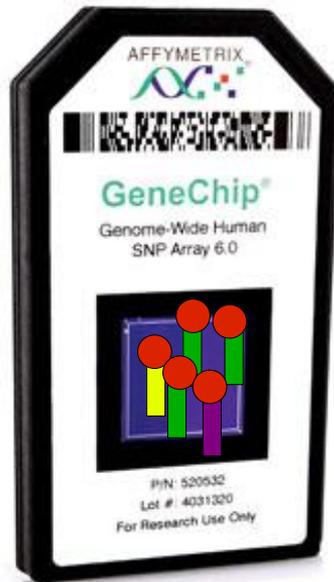
Nehme einen Microarray



mRNA auf Microarray



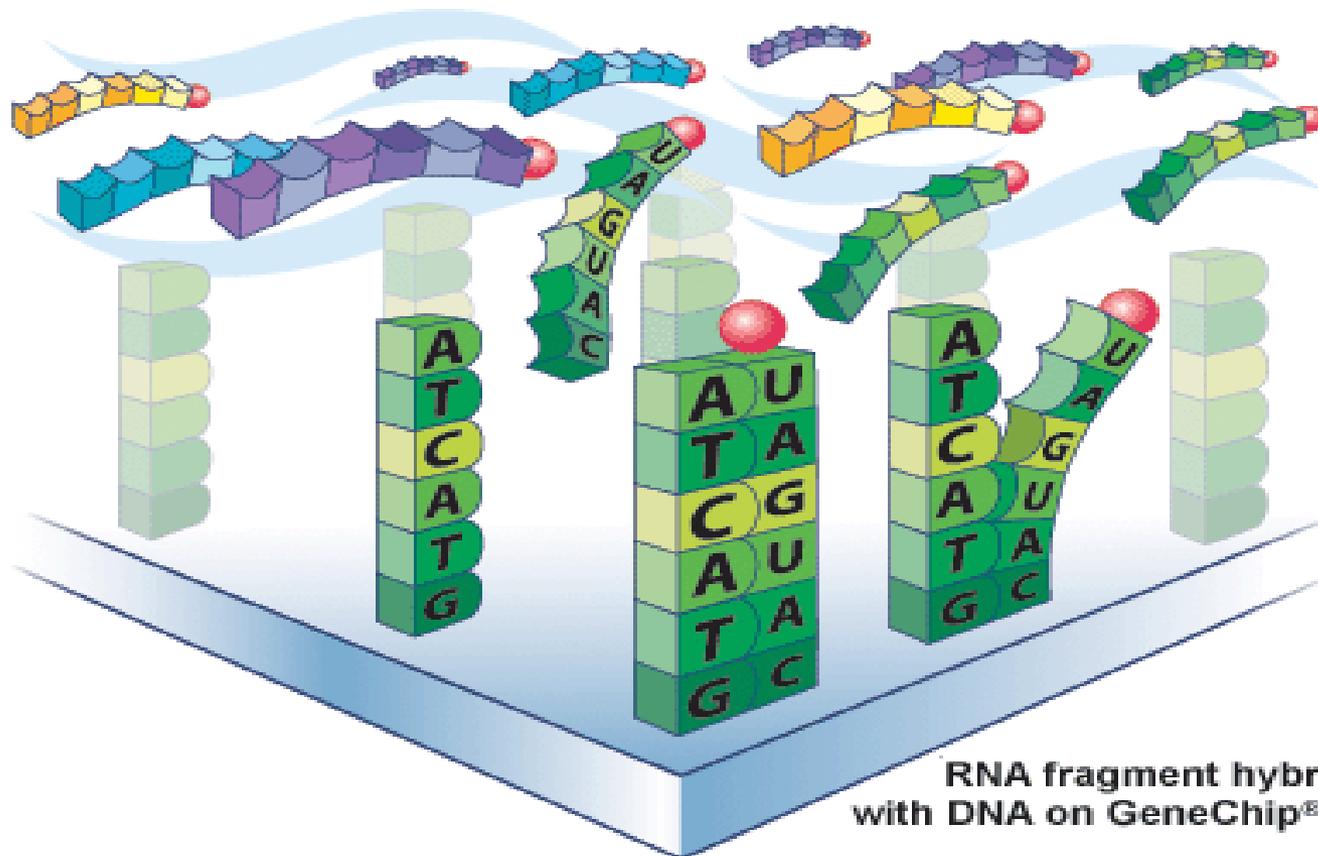
mRNA auf Microarray



Auf dem Microarray

From Computer Desktop Encyclopedia
Reproduced with permission.
© 2007 Affymetrix

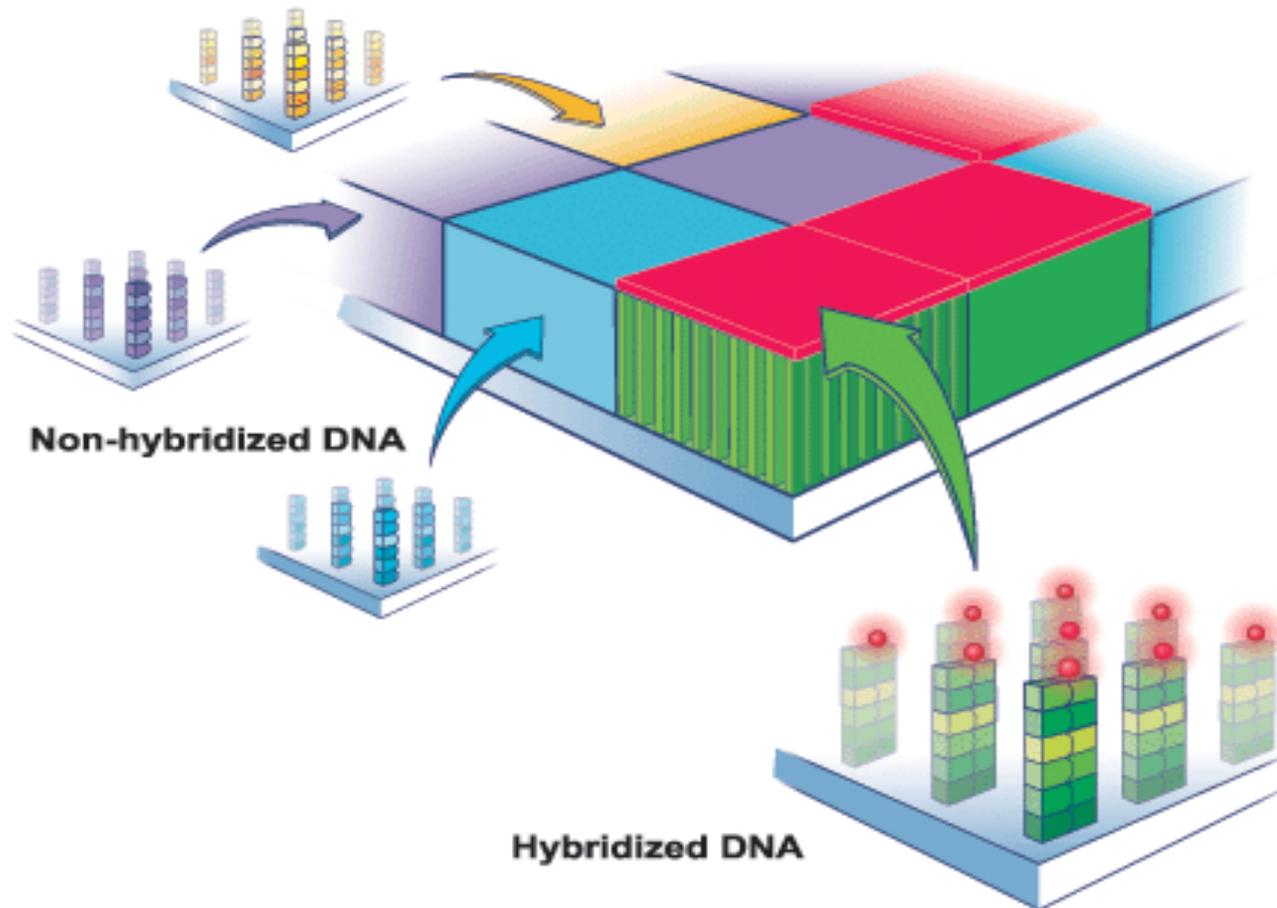
RNA fragments with fluorescent tags from sample to be tested



Voila: Ein Feuerwerk!

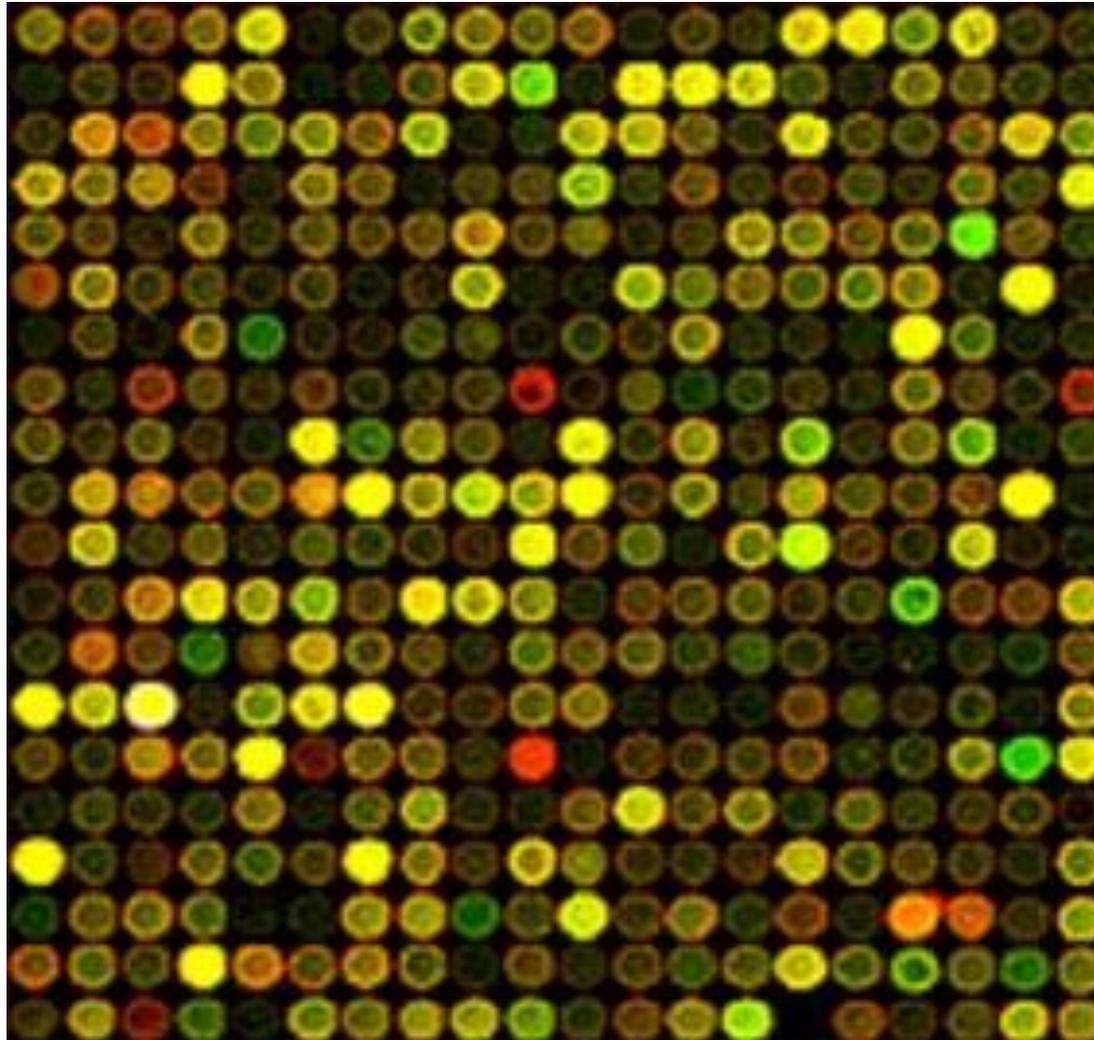
From Computer Desktop Encyclopedia
Reproduced with permission.
© 2007 Affymetrix

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



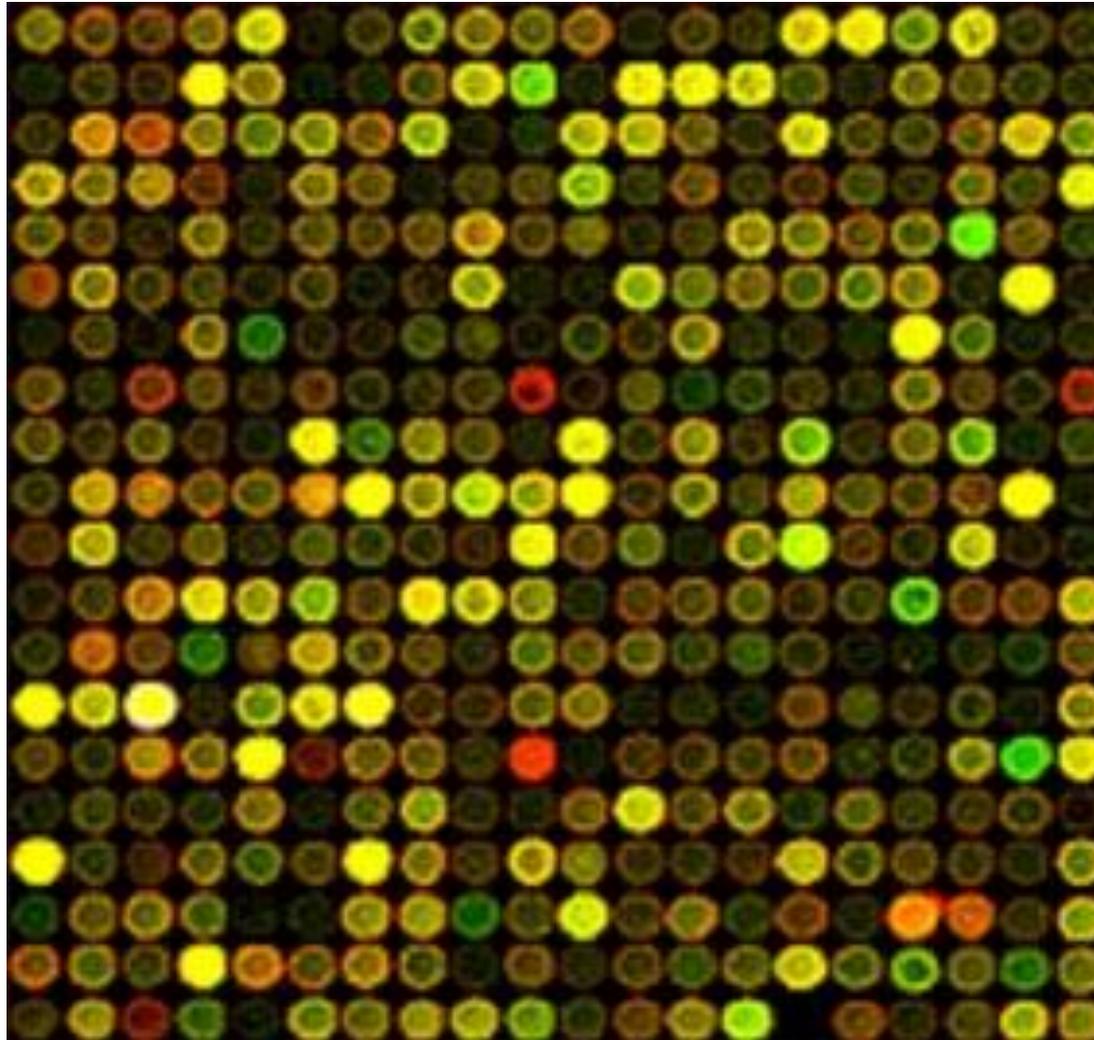
Helligkeit = Aktivität des Gens

Gen 5 sehr aktiv



Helligkeit = Aktivität des Gens

Gen 6 nicht aktiv

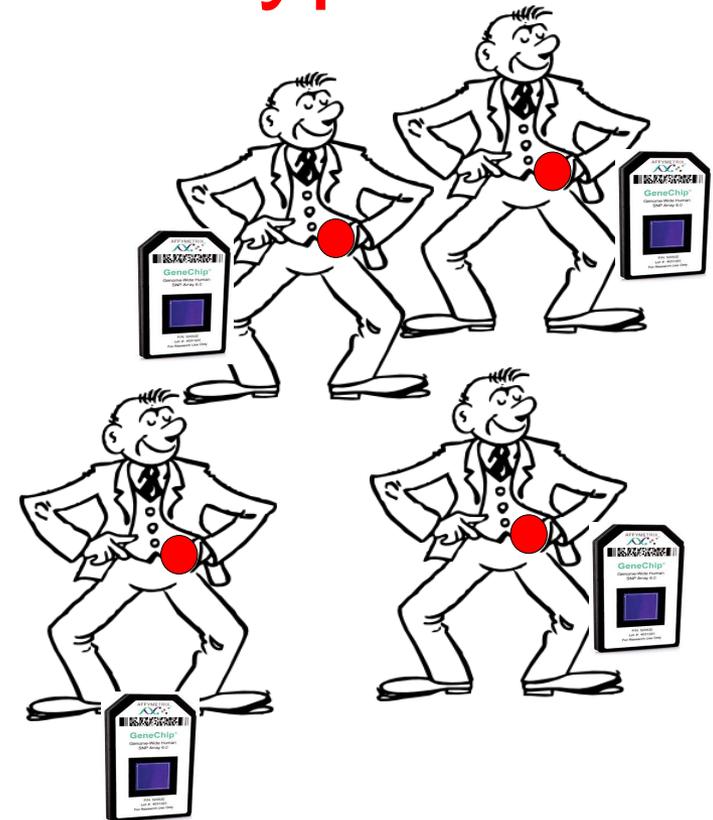


Für jeden Patienten ein Microarray

Typ 1



Typ 2



Microarray: Aktivität aller Gene in der Zelle

Typ 1

Typ 2

Gen	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5
1	2.1	1.3	1.9	1.2	1.4
2	2.4	2.3	2.5	2.1	2.0
...					
50000					

Gen	Pat. 1	Pat. 2	Pat. 3	Pat. 4
1	1.9	2.5	2.4	2.9
2	2.3	2.2	2.4	2.1
...				
50000				

Microarray: Aktivität aller Gene in der Zelle

Typ 1

Typ 2

Gen	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5
1	2.1	1.3	1.9	1.2	1.4
2	2.4	2.3	2.5	2.1	2.0
...					
50000					

Gen	Pat. 1	Pat. 2	Pat. 3	Pat. 4
1	1.9	2.5	2.4	2.9
2	2.3	2.2	2.4	2.1
...				
50000				

Microarray: Aktivität aller Gene in der Zelle

Typ 1

Typ 2

Gen	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5
1	2.1	1.3	1.9	1.2	1.4
2	2.4	2.3	2.5	2.1	2.0
...					
50000					

Gen	Pat. 1	Pat. 2	Pat. 3	Pat. 4
1	1.9	2.5	2.4	2.9
2	2.3	2.2	2.4	2.1
...				
50000				

Ist Gen 1 bei **Typ 2-Tumorzellen**
signifikant aktiver?

Falls ja: Gen 1 kann **Typ1-Tumor** und **Typ-2 Tumor**
unterscheiden!

Falls ja:



Gen 1 nicht aktiv

Gen 1 aktiv

Typ 1

Typ 2

~~Chemotherapie~~

Chemotherapie

Microarray: Aktivität aller Gene in der Zelle

Typ 1

Typ 2

Gen	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5
1	2.1	1.3	1.9	1.2	1.4
2	2.4	2.3	2.5	2.1	2.0
...					
50000					

Gen	Pat. 1	Pat. 2	Pat. 3	Pat. 4
1	1.9	2.5	2.4	2.9
2	2.3	2.2	2.4	2.1
...				
50000				

Ist Gen 1 bei **Typ 2-Tumorzellen**
signifikant aktiver?

Ungepaarter t-Test

Ungepaarter t-Test: 1/3

1. Modell:

$$\begin{aligned} X_1, \dots, X_n & \text{ iid } \sim \mathcal{N}(\mu_X, \sigma^2), \\ Y_1, \dots, Y_m & \text{ iid } \sim \mathcal{N}(\mu_Y, \sigma^2). \end{aligned}$$

2. Nullhypothese:

$$H_0 : \mu_X = \mu_Y.$$

Alternative:

$$\begin{aligned} & H_A : \mu_X \neq \mu_Y \text{ (zweiseitig)} \\ \text{oder } & H_A : \mu_X > \mu_Y \text{ (einseitig)} \\ \text{oder } & H_A : \mu_X < \mu_Y \text{ (einseitig)} \end{aligned}$$

Ungepaarter t-Test: 2/3

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

3. Teststatistik:

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

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 H_0 : $T \sim t_{n+m-2}$.

$$\frac{1}{n-1} \sum (x_i - \bar{x})^2$$

Ungepaarter t-Test: 3/3

4. **Signifikanzniveau:** α

5. **Verwerfungsbereich für die Teststatistik:**

$$\begin{array}{ll} (-\infty, -t_{n+m-2, 1-\alpha/2}] \cup [t_{n+m-2, 1-\alpha/2}, \infty) & \text{bei Alternative } H_A : \mu_X \neq \mu_Y, \\ [t_{n+m-2, 1-\alpha}, \infty) & \text{bei Alternative } H_A : \mu_X > \mu_Y, \\ (-\infty, -t_{n+m-2, 1-\alpha}] & \text{bei Alternative } H_A : \mu_X < \mu_Y. \end{array}$$

6. **Testentscheid:** Entscheide, ob der beobachtete Wert der Teststatistik im Verwerfungsbereich der Teststatistik liegt.

Microarray: Aktivität aller Gene in der Zelle

Typ 1

Typ 2

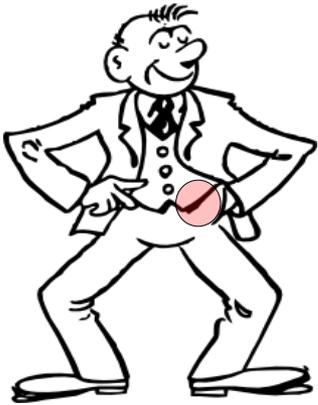
Gen	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5
1	2.1	1.3	1.9	1.2	1.4
2	2.4	2.3	2.5	2.1	2.0
...					
50000					

Gen	Pat. 1	Pat. 2	Pat. 3	Pat. 4
1	1.9	2.5	2.4	2.9
2	2.3	2.2	2.4	2.1
...				
50000				

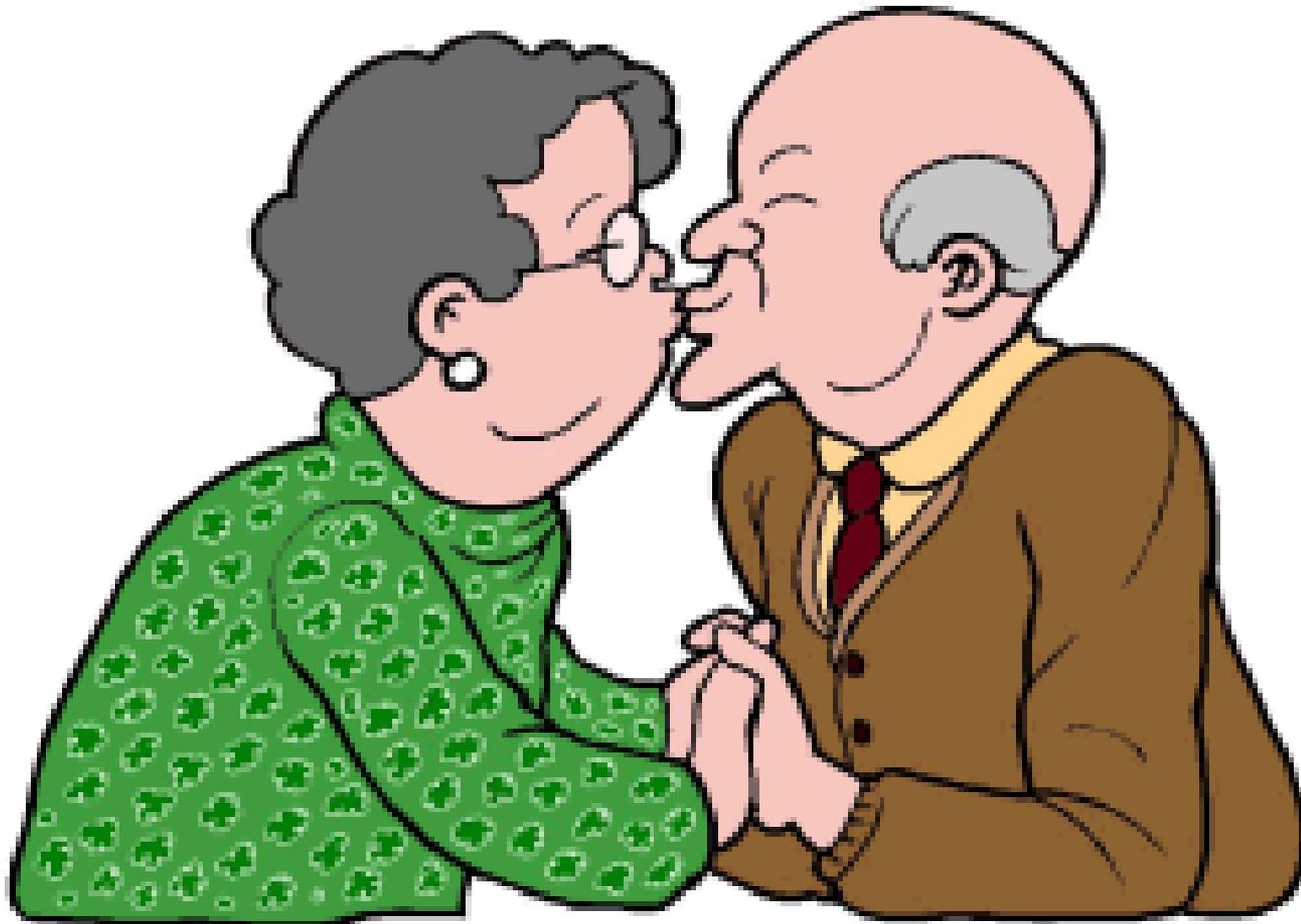
$$n = 5, \quad m = 4, \quad \bar{x} = 1.58, \quad \bar{y} = 2.43, \quad \hat{\sigma}_x = 0.40, \quad \hat{\sigma}_y = 0.41$$

Ungepaarter t-Test

Happy End !

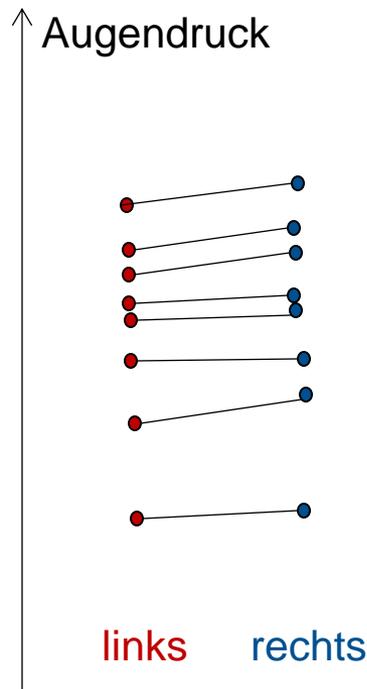


Happy End !



Gepaart vs. Ungepaart

- Bsp: Augeninnendruck; ein Auge behandelt, das andere nicht (gepaarter Test ist angebracht)
- Gemäss Vorraussetzungen dürfte auch ein ungepaarter Test angewendet werden



$$H_0: \mu_X = \mu_Y$$

Ungepaart:

Intuition Teststatistik: $T = \frac{\bar{X} - \bar{Y}}{\widehat{\sigma}_{\bar{X}}}$

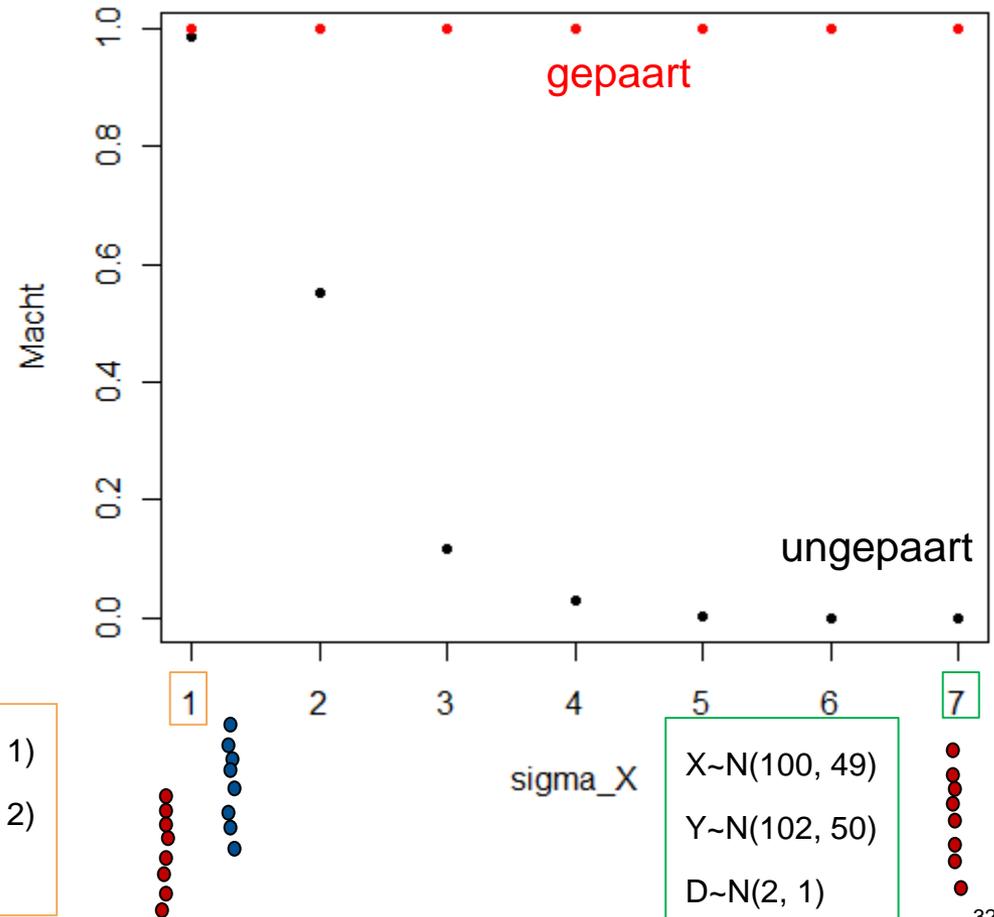
Gepaart:

Differenz $D_i = X_i - Y_i$

Teststatistik $T = \frac{\bar{D}}{\widehat{\sigma}_{\bar{D}}}$

Gepaart vs. Ungepaart: Simulationsstudie

- $H_0: \mu_D = 0$ bzw. $H_0: \mu_X = \mu_Y$; $n=m=10$
- $X \sim N(100, \sigma_X^2)$, $D \sim N(2, 1)$, $Y = X + D \sim N(102, 1 + \sigma_X^2)$
gepaarte Situation
- Der **gepaarte t-Test** hat **mehr Macht**, wenn die Daten verrauscht sind.



t-Test falls Varianz in Gruppen unterschiedlich (aka Welch-Test)

- Grundidee identisch
- Teststatistik und Verteilung falls H_0 stimmt ist komplizierter
- Computer: Dieser Test ist meist der “default” t-Test
- Praxis: Man sollte immer annehmen, dass die Varianz der Gruppen unterschiedlich ist; d.h., Welch-Test verwenden
- Prüfung: Um einfacher rechnen zu können, werden wir beim t-Test gleiche Varianzen in den beiden Gruppen annehmen

Mann-Whitney U-Test (aka Wilcoxon Rank-sum Test)

- Falls Daten nicht normalverteilt

- $X_i \sim F, i = 1, \dots, n; Y_j \sim G, j = 1, \dots, m$

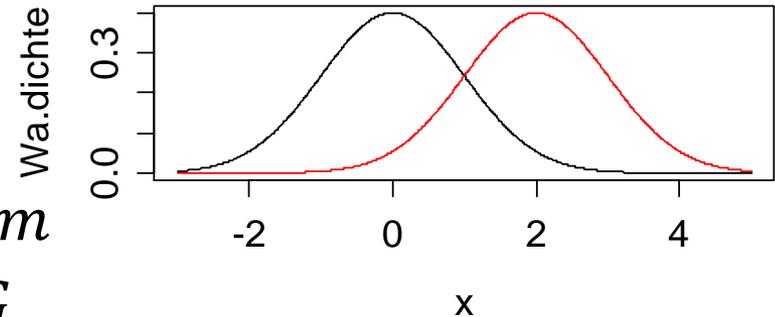
$$H_0: F = G$$

$$H_A: F = G + \delta \ (\delta \neq 0) \text{ (oder einseitig)}$$

(d.h., Verteilungen sind verschoben, haben aber gleiche Form)

- Teststatistik:

- Bilde Ränge über beide Gruppen hinweg
- Falls Gruppen gleich gross sind, sollten Rangsummen etwa gleich sein
- Falls Gruppen ungleich, sollten Rangsummen in einem gewissen Verhältnis stehen



Bsp: Mann-Whitney U-Test

- Behandlung (B) und Kontrolle (K) je 2 Patienten
- Beobachtung: B: 1.2, 3.1; K: 5.9, 4.4
- Gesamtrang: B: 1, 2; K: 4, 3
- Rangsumme R in K: 4 + 3 = 7
- Falls H_0 stimmt sind alle Ränge in K 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:

$$P(R \geq 7) = P(R = 7) = \frac{1}{6} \approx 0.167$$

P-Wert ←

- H_0 kann auf dem 5% Niveau nicht verworfen werden
- Praxis: Computer verwenden

Übersicht: Tests für ungepaarte Stichproben

Test	Annahmen				n_{min} (falls $n = m$) bei $\alpha = 0.05$	Macht für ein Beispiel (1)
	$\sigma_X = \sigma_Y$	$X_i \sim N$ $Y_i \sim N$	F, G haben gleiche Form	iid pro Gruppe		
t ($\sigma_X = \sigma_Y$)	x	x	x	x	2	57 %
t ($\sigma_X \neq \sigma_Y$)		x		x	2	56 %
MW U-Test	x		x	x	4	53 %

(1): $X_i \sim N(\mu_X, \sigma^2), Y_i \sim N(\mu_Y, \sigma^2), n = m = 10; H_0: \mu_X = \mu_Y; H_A: \mu_X \neq \mu_Y; \alpha = 0.05$

Macht berechnet für konkrete Alternative: $X_i \sim N(0,1), Y_i \sim N(1,1)$

Multiple Testen

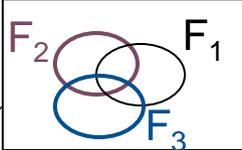
- Microarray Test mit $m=1000$ Genen
- 1000 t-Tests auf dem 5% Sign.niveau
- Angenommen, kein Gen hat einen Effekt:

Ca. 50 Tests (5% von 1000) werden trotzdem ein signifikantes Ergebnis liefern !

50 Gene werden als “wichtig” angegeben, obwohl sie gar nicht “wichtig” sind.

- Wie erzeugt man eine Liste von “wirklich wichtigen” Genen?

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 Sign.niveau $\frac{\alpha}{m}$,
statt α ; m ist die Anzahl Gene
- Begründung:
 F_i : Fehler 1. Art bei Gen i


“union bound”: Gesamtfläche ist sicher kleiner als Summe der Einzelflächen

$$P\left(\bigcup_{i=1}^m F_i\right) \leq \sum_{i=1}^m P(F_i) = \sum_{i=1}^m \frac{\alpha}{m} = \alpha$$
- Nachteil: Liste ist evtl. “extrem” konservativ (z.B. gar kein Gen enthalten)

Wiederholung: Hypothesentests

- Für Prüfung
Eine Aufgabe: **Binomialtest**
Eine Aufgabe: **t-Test und Vorzeichentest**

Test	Was wird getestet	Bsp
Binomialtest	Ist Anteil in Gruppe gleich π_0 ?	Wirksamkeit von Medikament
<i>t-Test: Eine Stichprobe</i>	<i>Ist Erwartungswert in Gruppe gleich μ_0?</i>	<i>Füllmenge in Getränkeflaschen</i>
<i>t-Test: 2 gepaarte Stichproben</i>	<i>Ist Erwartungswert in beiden Gruppen gleich?</i>	<i>Reaktionszeit von Haupt- und Nebenhand</i>
<i>Vorzeichentest</i>	<i>Ist Median in Gruppe gleich μ_0?</i>	<i>Wie oben</i>
<i>t-Test: 2 ungepaarte Stichproben</i>	<i>Ist Erwartungswert in beiden Gruppen gleich?</i>	<i>Aktivität von Gen XY bei Gesunden und Kranken</i>