



האוניברסיטה
העברית
בירושלים
THE HEBREW
UNIVERSITY
OF JERUSALEM

אלגוריתמים בביו' חישובית

76558

איך לנסח שאלת מחקר

תומי קפלן
31/12/2023

דנ"א, הגנום, וחיית אחרות

- הדנ"א בנוי מארבע חומצות הגרעין

Purines

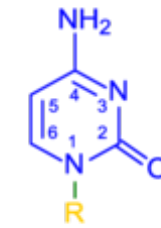


Adenine

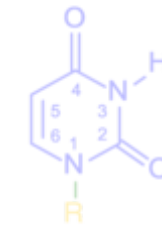


Guanine

Pyrimidines



Cytosine



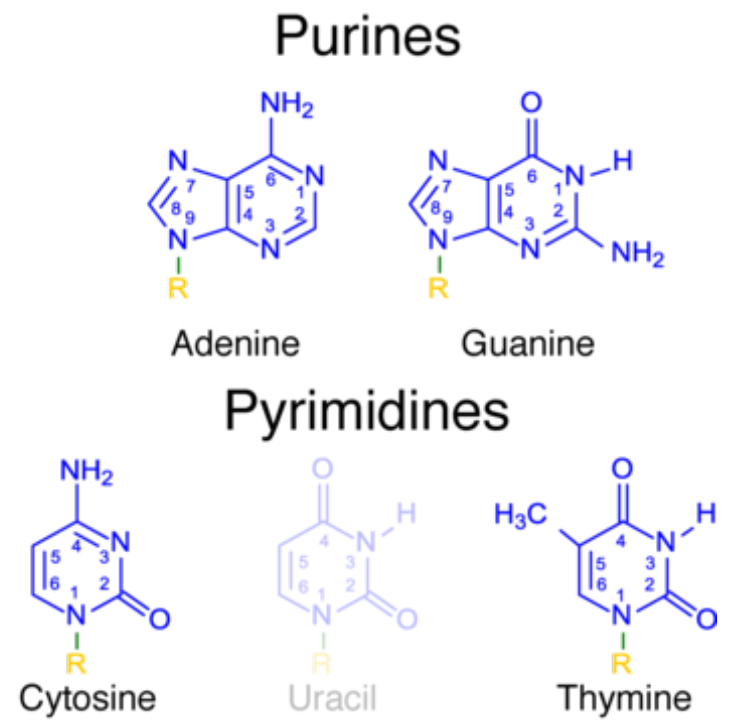
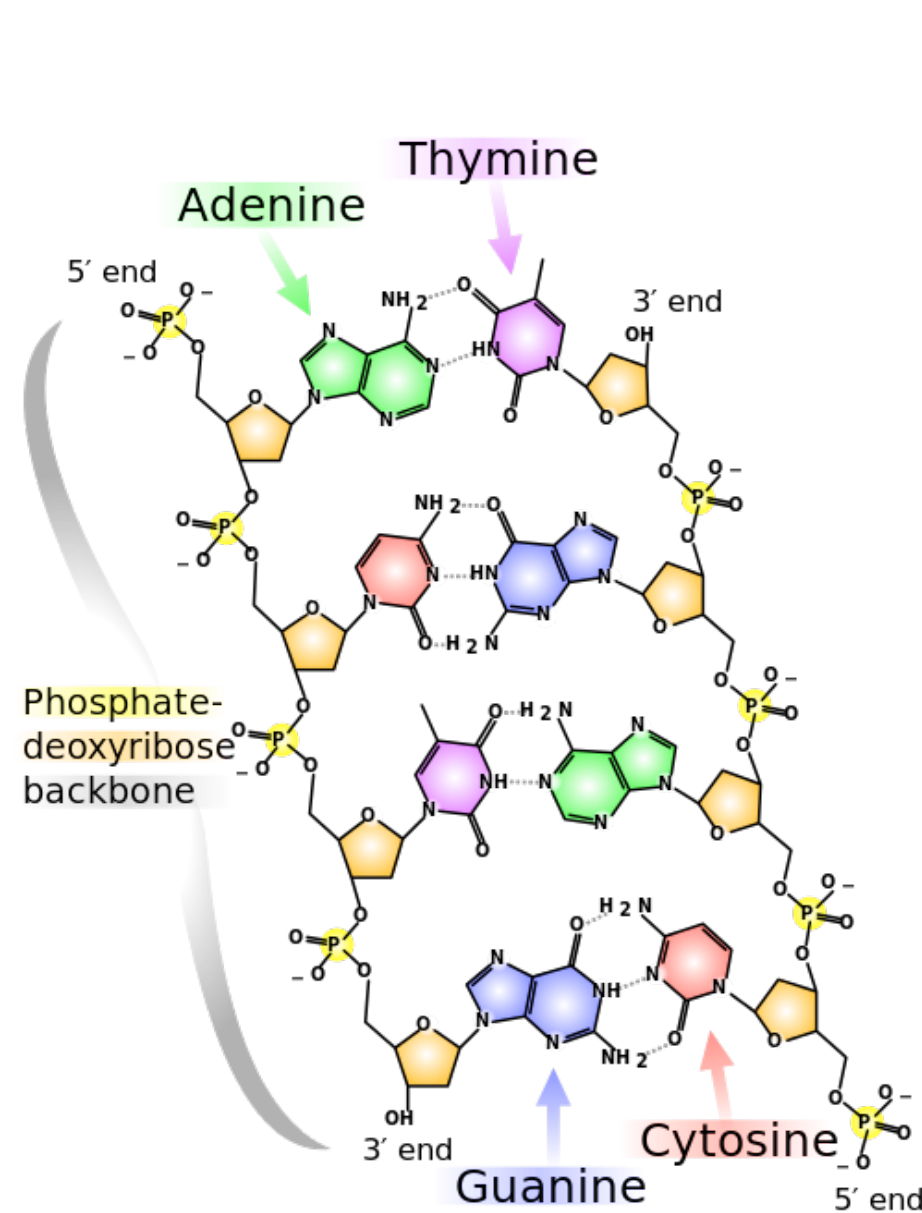
Uracil



Thymine

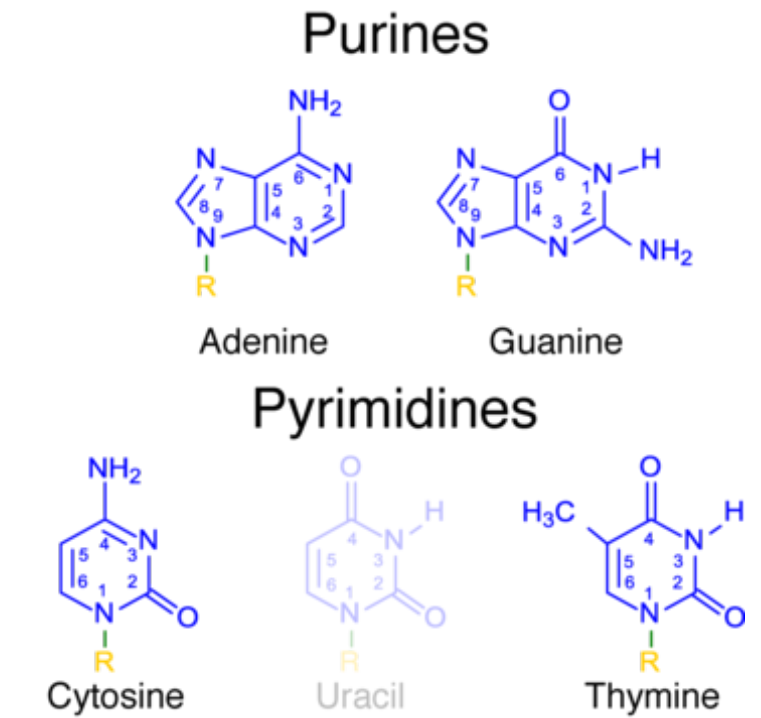
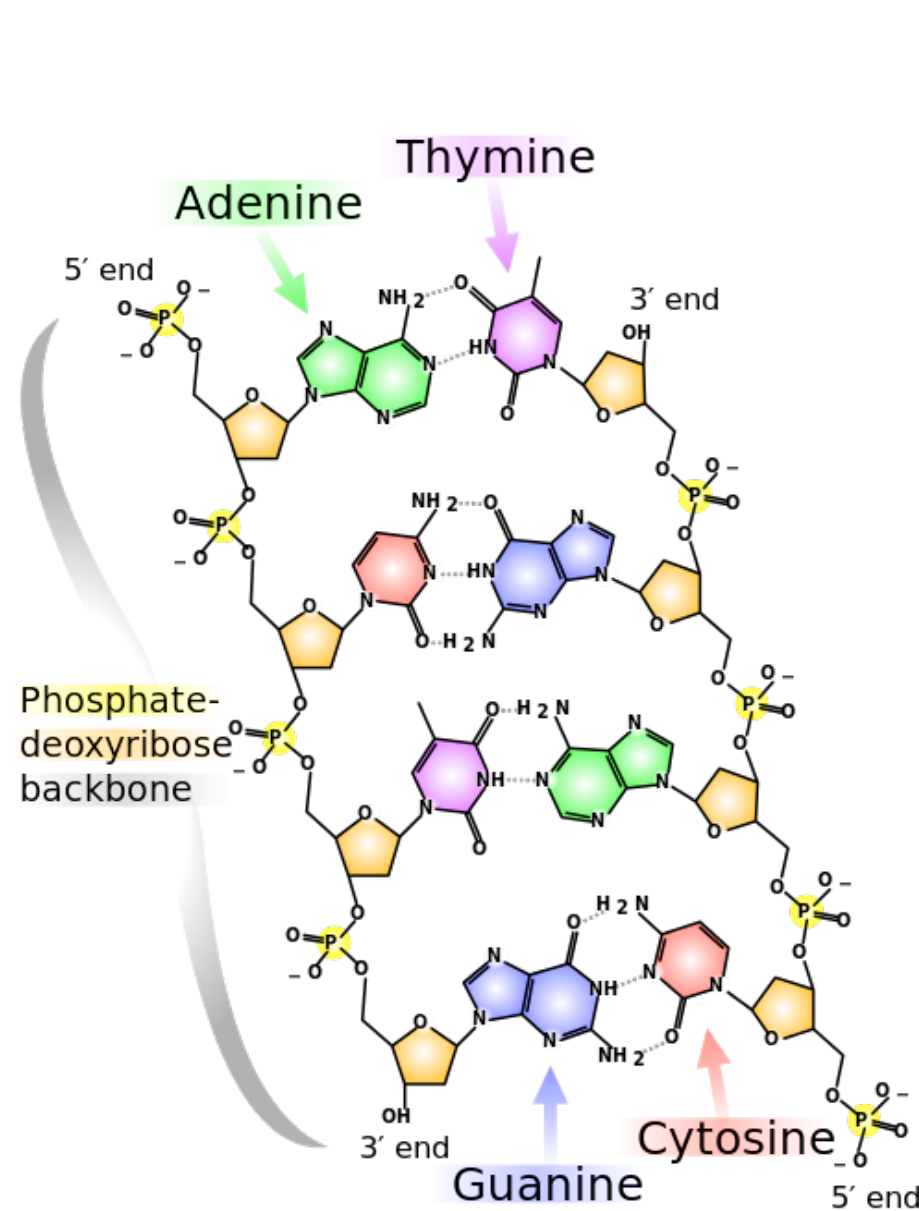
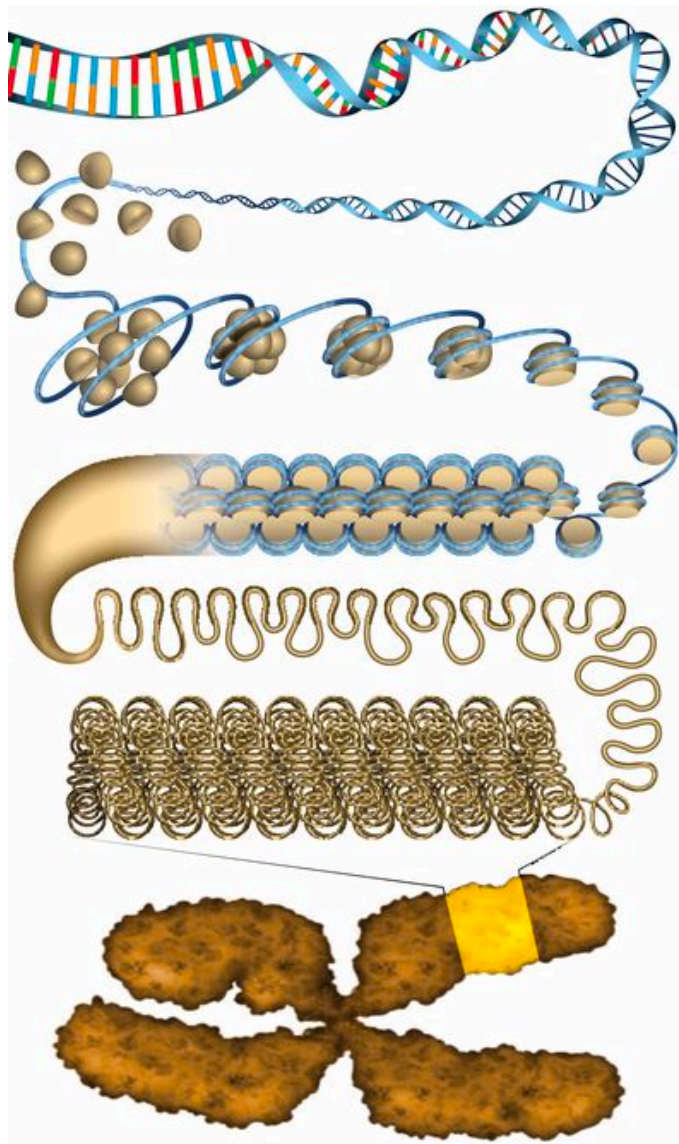
דנ"א, הגנום, וחיות אחרות

- הדנ"א בנוי מארבע חומצות הגרעין



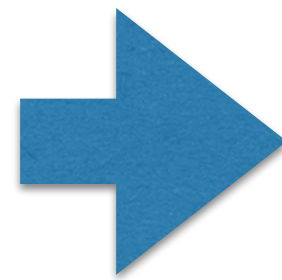
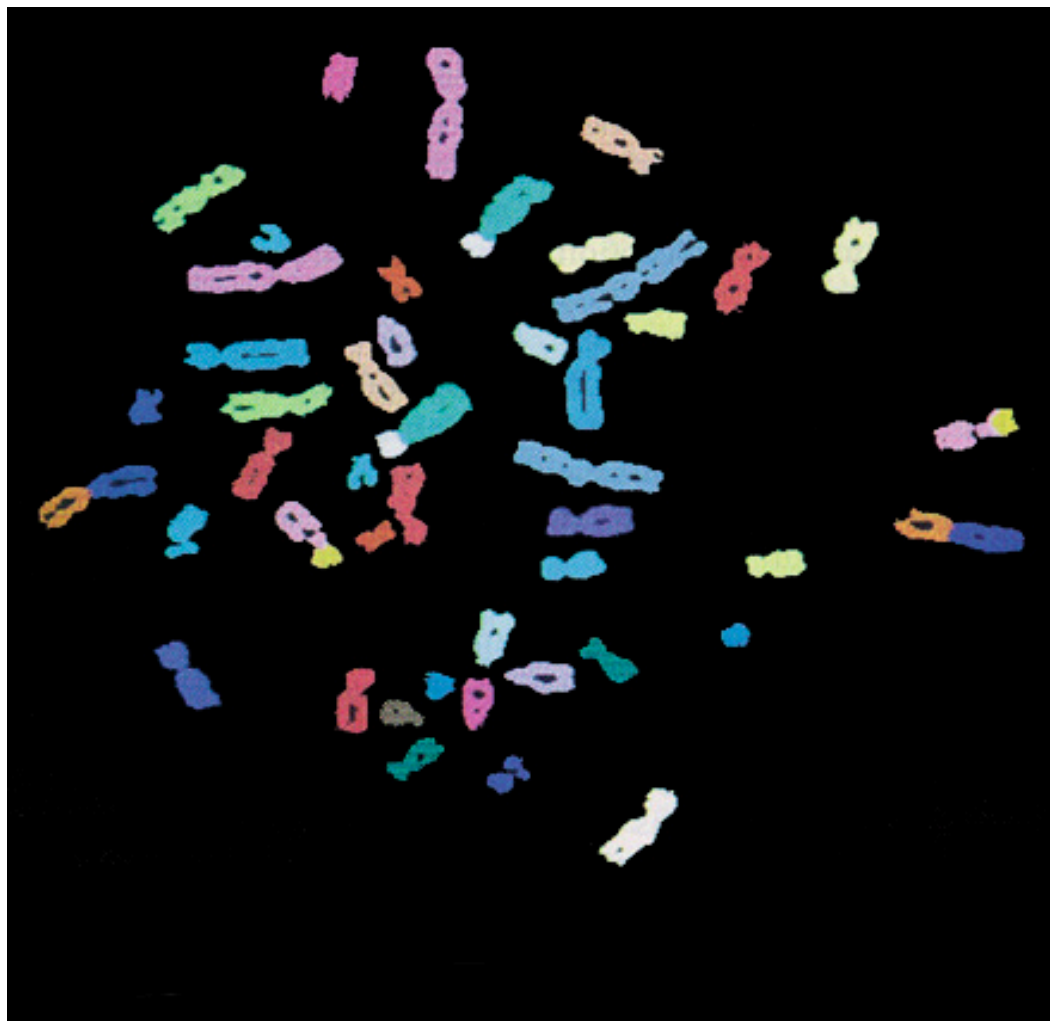
דנ"א, הגנום, וחיות אחרות

- הדנ"א בנוי מארבע חומצות הגרעין



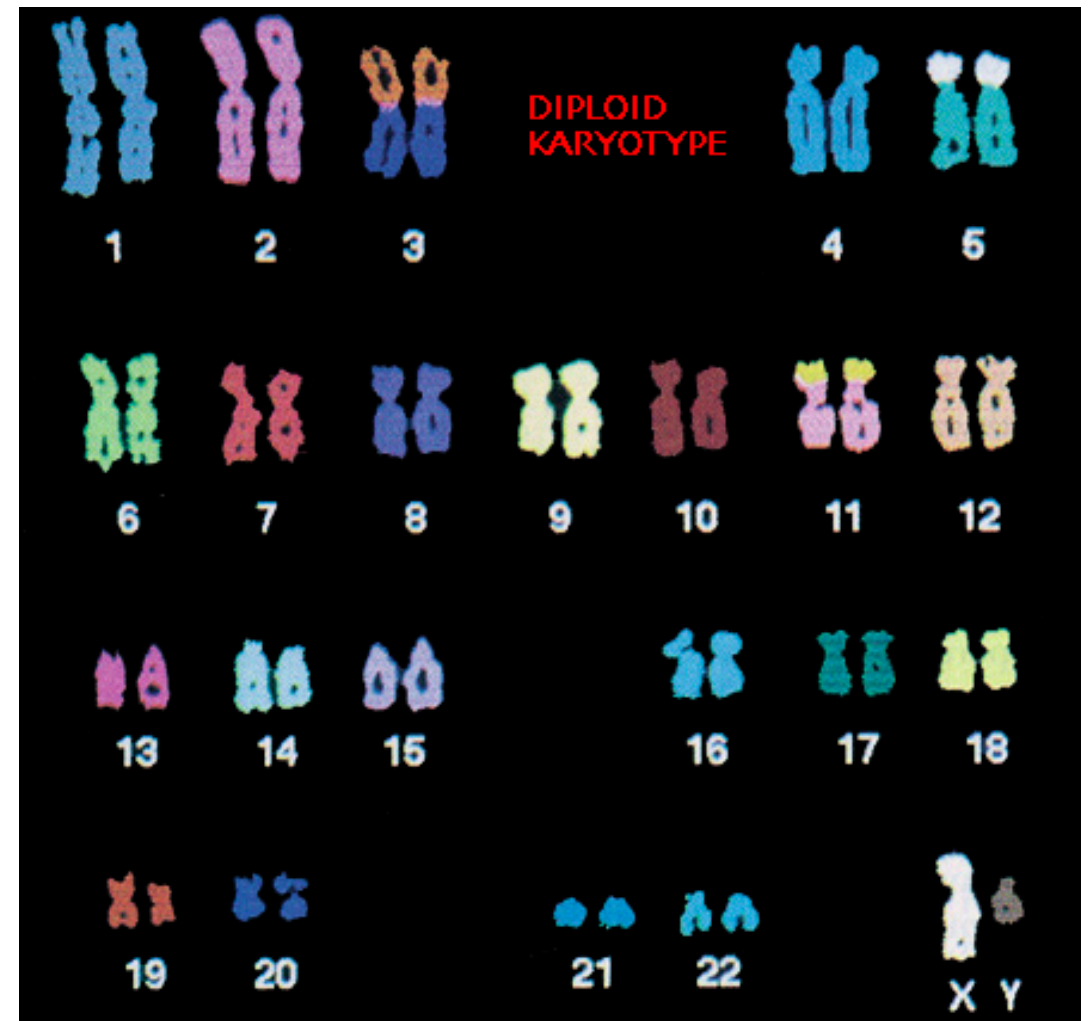
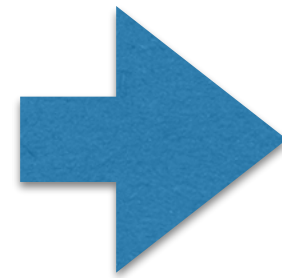
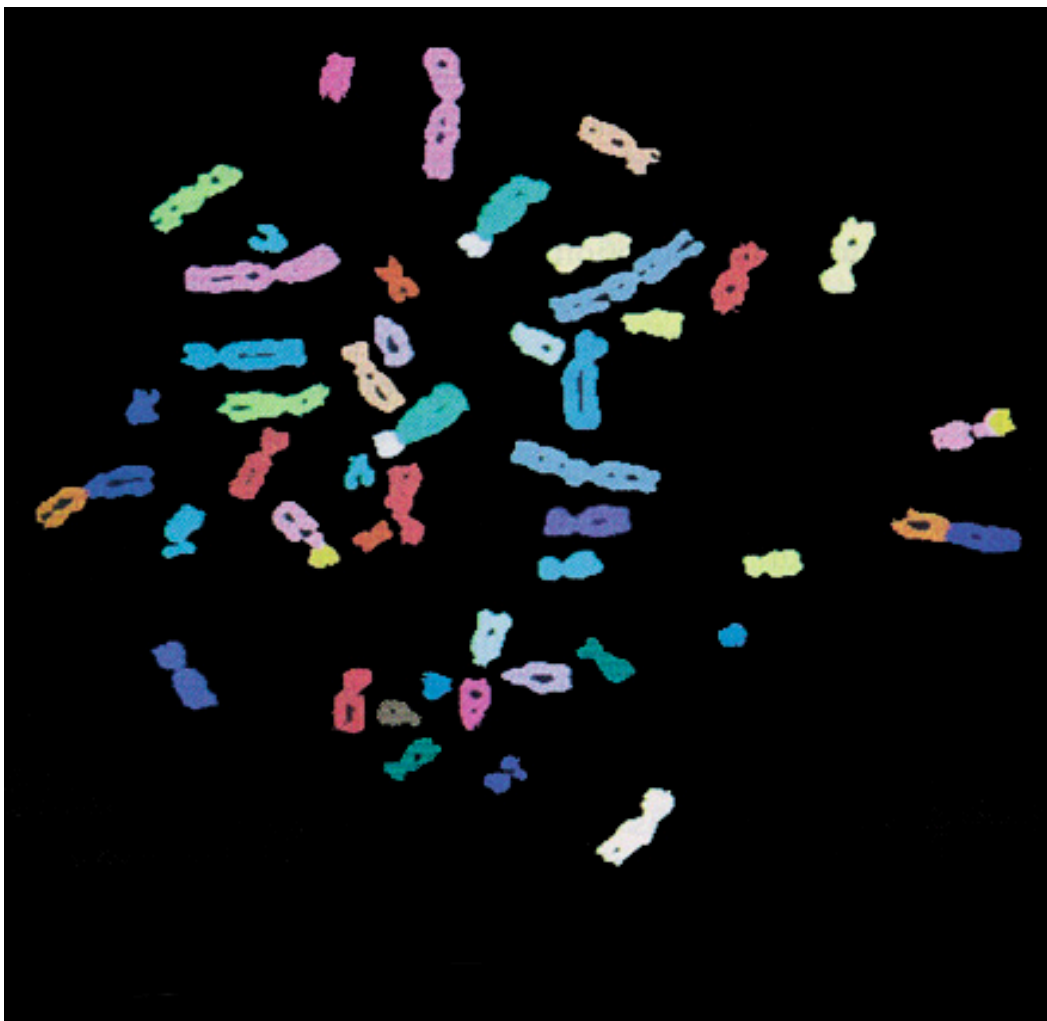
דנ"א, הגנום, וחיות אחרות

- הדנ"א בנוי מארבע חומצות הגרעין
- 23 זוגות כרומוזומים בכל תא [48-249Mb]



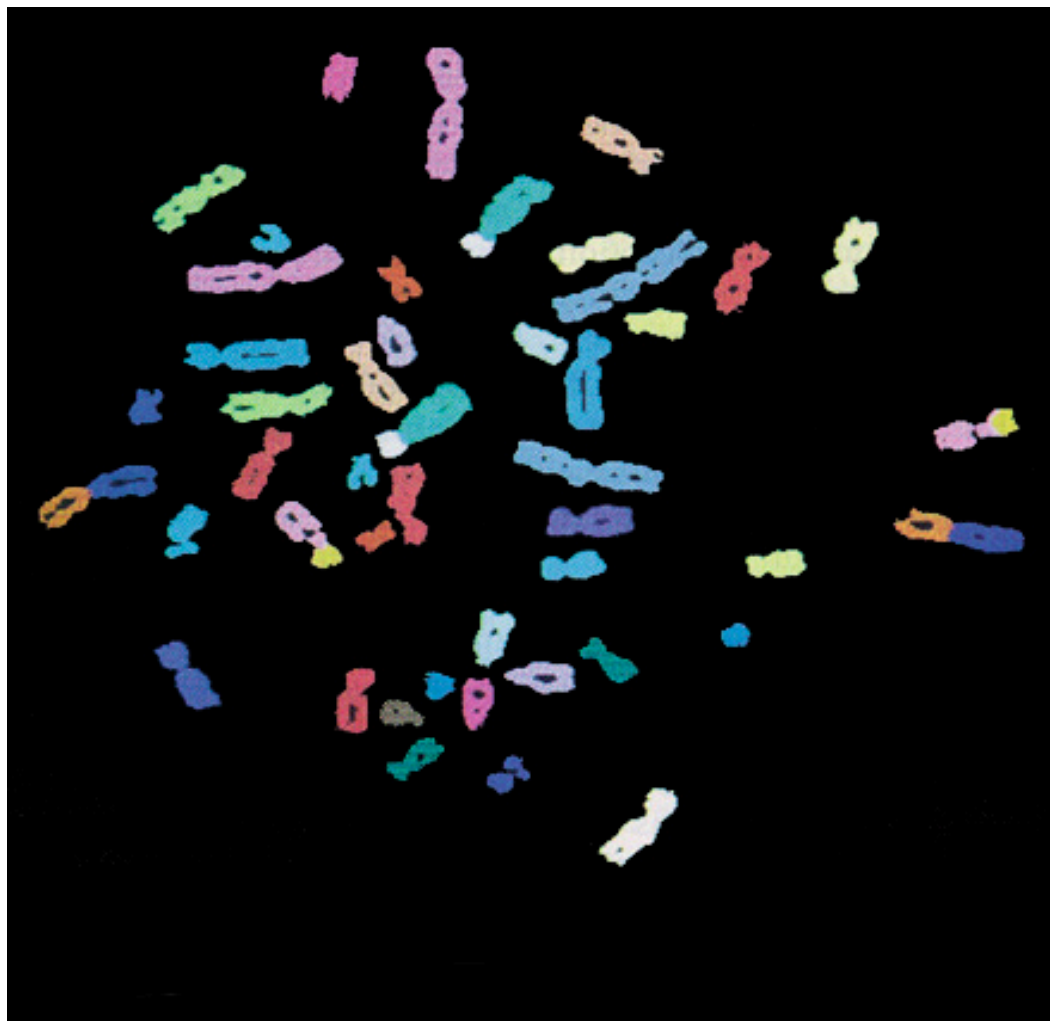
דנ"א, הגנום, וחיות אחרות

- הדנ"א בנוי מארבע חומצות הגרעין
- 23 זוגות כרומוזומים בכל תא [48-249Mb]



דנ"א, הגנום, וחיות אחרות

- הדנ"א בנוי מארבע חומצות הגרעין
- 23 זוגות כרומוזומים בכל תא [48-249Mb]



דנ"א, הגנום, וחיות אחרות

- הדנ"א בנוי מארבע חומצות הגרעין
- 23 זוגות כרומוזומים בכל תא [48-249Mb]



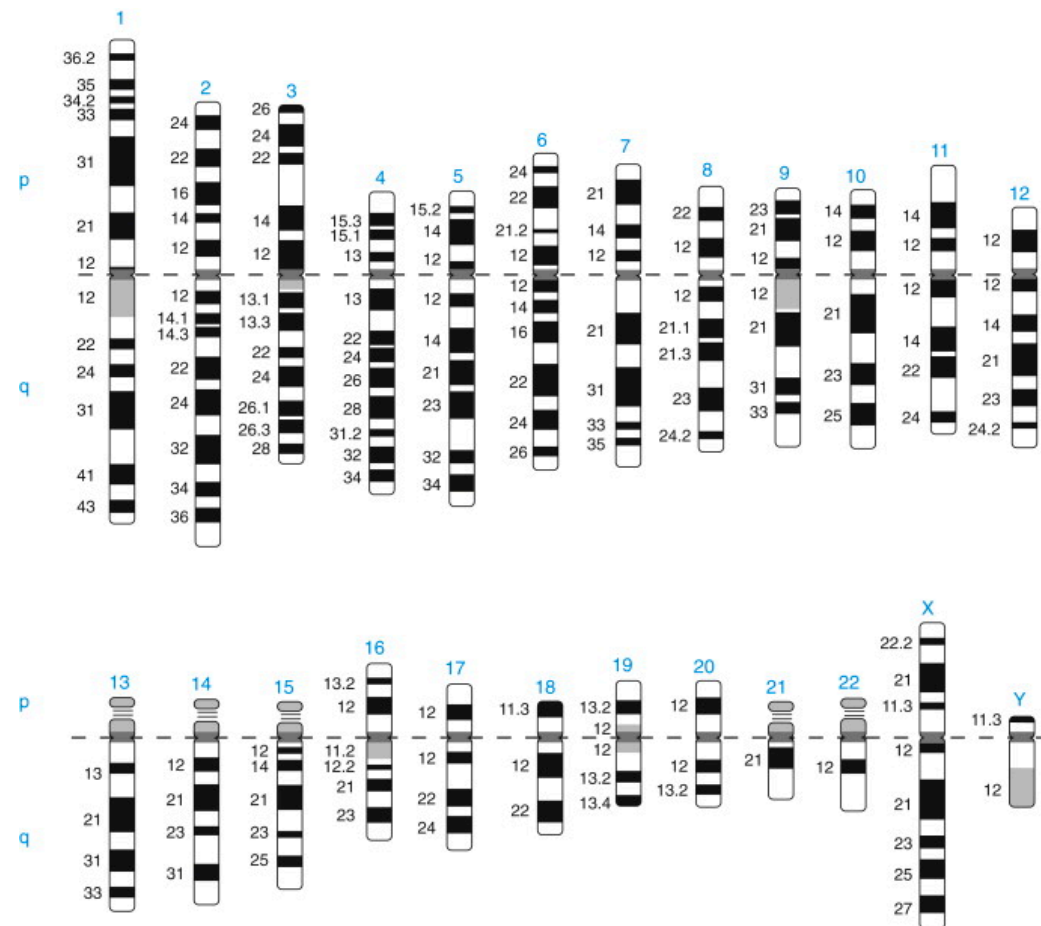
דנ"א, הגנום, וחיות אחרות

- הדנ"א בנוי מארבע חומצות הגרעין

- 23 זוגות כרומוזומים בכל תא [48-249Mb]



Giemsa staining



דנ"א, הגנום, וחיות אחרות

- הדנ"א בנוי מארבע חומצות הגרעין
- 23 זוגות כרומוזומים בכל תא [48-249Mb]
- רצף הגנום כולו 3Gb בסיסים
- 1.5% ממנו מקודד לחלבונים [קודון \Leftarrow חומצת אמינו]
- 23K חלבונים שונים
- רצף החלבון \Leftarrow מבנה החלבון \Leftarrow תפקיד החלבון

תחומים בביולוגיה חישובית

- דנ"א: מיפוי, גנטיקה, גנומיקה, מטאגנומיקה, מתילציה, אבולוציה, סרטן, דנ"א חופשי, פרגמנטומיקה

תחומים בביולוגיה חישובית

- דנ"א: מיפוי, גנטיקה, גנומיקה, מטאגנומיקה, מתילציה, אבולוציה, סרטן, דנ"א חופשי, פרגמנטומיקה
- דנ"א: ביטוי גנים, דנ"א של התא הבודד, דנ"א קצר

תחומים בביולוגיה חישובית

- דנ"א: מיפוי, גנטיקה, גנומיקה, מטאגנומיקה, מתילציה, אבולוציה, סרטן, דנ"א חופשי, פרגמנטומיקה
- רנ"א: ביטוי גנים, רנ"א של התא הבודד, רנ"א קצר
- חלבונים: ספקטרומטריה של מאסות, אפטמרים, ניבוי מבנים, ניבוי אינטראקציות, תכנון מבנים, נוגדנים

תחומים בביולוגיה חישובית

- דנ"א: מיפוי, גנטיקה, גנומיקה, מטאגנומיקה, מתילציה, אבולוציה, סרטן, דנ"א חופשי, פרגמנטומיקה
- רנ"א: ביטוי גנים, רנ"א של התא הבודד, רנ"א קצר
- חלבונים: ספקטרומטריה של מאסות, אפטמרים, ניבוי מבנים, ניבוי אינטראקציות, תכנון מבנים, נוגדנים
- רשתות, מסלולים, תהליכים

תחומים בביולוגיה חישובית

- דנ"א: מיפוי, גנטיקה, גנומיקה, מטאגנומיקה, מתילציה, אבולוציה, סרטן, דנ"א חופשי, פרגמנטומיקה
- רנ"א: ביטוי גנים, רנ"א של התא הבודד, רנ"א קצר
- חלבונים: ספקטרומטריה של מאסות, אפטמרים, ניבוי מבנים, ניבוי אינטראקציות, תכנון מבנים, נוגדנים
- רשתות, מסלולים, תהליכים
- אקולוגיה ואבולוציה

תחומים בביולוגיה חישובית

- דנ"א: מיפוי, גנטיקה, גנומיקה, מטאגנומיקה, מתילציה, אבולוציה, סרטן, דנ"א חופשי, פרגמנטומיקה
- רנ"א: ביטוי גנים, רנ"א של התא הבודד, רנ"א קצר
- חלבונים: ספקטרומטריה של מאסות, אפטמרים, ניבוי מבנים, ניבוי אינטראקציות, תכנון מבנים, נוגדנים
- רשתות, מסלולים, תהליכים
- אקולוגיה ואבולוציה
- ?

חלון למתודה המדעית

- איך עובדים עם דאטה?
- הבנה טכנולוגית כדי להבין בעיות אפשריות / רעש
- איך לייצג ולפרש את המידע הנאסף

חלון למתודה המדעית

- איך עובדים עם דאטה?
- הבנה טכנולוגית כדי להבין בעיות אפשריות / רעש
- איך לייצג ולפרש את המידע הנאסף
- איך לתרגם שאלת מחקר למודל מתמטי/חישובי?
- איך לשערך פרמטרים?

חלון למתודה המדעית

- איך עובדים עם דאטה?
- הבנה טכנולוגית כדי להבין בעיות אפשריות / רעש
- איך לייצג ולפרש את המידע הנאסף
- איך לתרגם שאלת מחקר למודל מתמטי/חישובי?
- איך לשערך פרמטרים?
- איך לבחון מובהקות סטטיסטית
- דוגמא: יצא לי 6/6 בקוביה. האם מפתיע? מיסטי?

דוגמא 1

- ריצפנו חלבון חדש. בדצוננו לדעת מה הוא עושה.

דוגמא 1

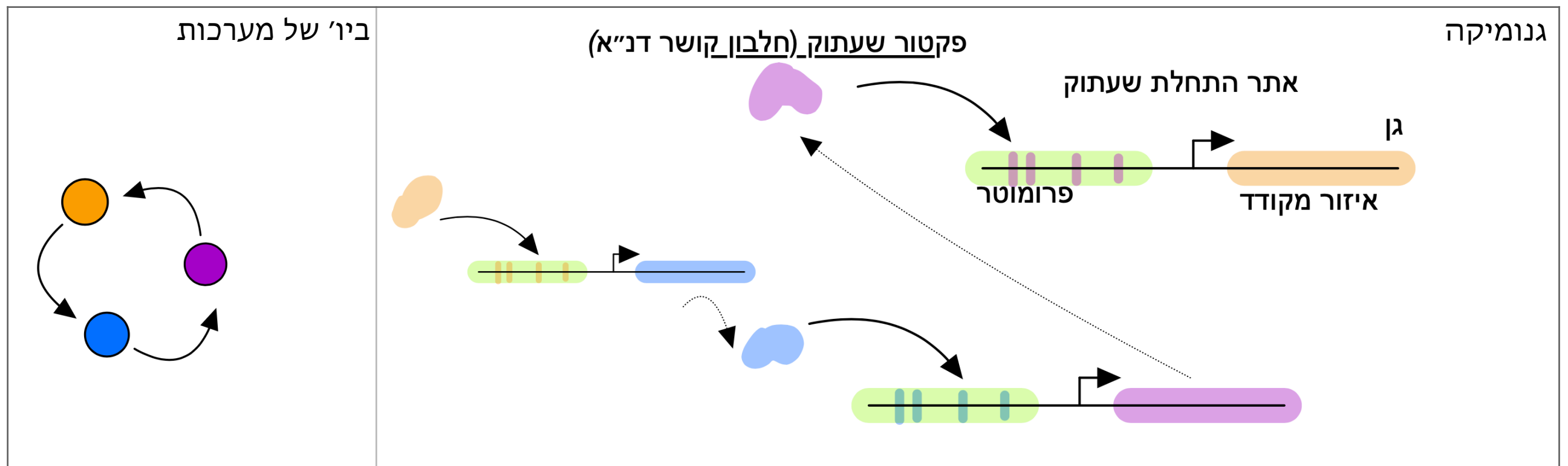
- ריצפנו חלבון חדש. ברצוננו לדעת מה הוא עושה.
- האם מוכרים חלבונים הדומים לו ברצף? מה תפקידם?

דוגמא 1

- ריצפנו חלבון חדש. ברצוננו לדעת מה הוא עושה.
- האם מוכרים חלבונים הדומים לו ברצף? מה תפקידם?
- מה זה "דומה"? מובהקות סטטיסטית? מודל רקע?

דוגמא 2

- גילינו שלושה פקטורי שעתוק*, כל אחד נקשר לפרומוטור† של קודמו ומבקר את פעילותו.
- האם מעניין? מפתיע? חשוב? פונקציונלי? נדיד?



*חלבונים הנקשרים לדנ"א ומעודדים שעתוקו לדנ"א
†איזור בקרה המצוי בתחילת כל גן

שבעת שלבי הקסם
לפתרון כל שאלת מחקר*

1. ניסוח השאלה המדעית

- מה אנחנו באמת רוצים לדעת?

- שימוש במושגים ביולוגיים

- דוגמא 1א: האם יש חלבונים משותפים בין חיות שונות?

- דוגמא 1ב: האם לחלבון א' וחלבון ב' יש אב קדמון משותף?

- דוגמא 2: האם קיימים מעגלי רגולציה ביצורים אוקריוטים?

1. ניסוח השאלה המדעית

- מה אנחנו באמת רוצים לדעת?

- שימוש במושגים ביולוגיים

- דוגמא 1א: האם יש חלבונים משותפים בין חיות שונות?

- דוגמא 1ב: האם לחלבון א' וחלבון ב' יש אב קדמון משותף?

- דוגמא 2: האם קיימים מעגלי רגולציה ביצורים אוקריוטים?

2. איסוף מידע רלוונטי

- תיאורטי ומעשי

- אבולוציה של רצפים, מודל מרחק בין רצפים, החלפה / הוספה / החסרה של אותיות, סיכויים להורשה של מוטציות

- בקרת שעתוק, אתרי קישור, רצפי בקרה [פרומוטורים],

גנומיקה

3. תרגום לשפה מתמטית (מודל)

- הגדרה פורמלית. משהו שתכנית מחשב יוכל לבדוק (דאטה אמיתי או סימולציה)
- מודל מתמטי / חישובי
- הגדרת ציון / סקור
- האם קיים במאגר רצף דומה מספיק (מרחק קטן מסף מסוים)

3. תרגום לשפה מתמטית (מודל)

- הגדרה פורמלית. משהו שתכנית מחשב יוכל לבדוק (דאטה

אמיתי או סימולציה)

- מודל מתמטי / חישובי

- הגדרת ציון / סקור

- האם קיים במאגר רצף דומה מספיק (מרחק קטן מסף מסוים)

4. אלגוריתם (אופטימיזציה)

- בהנתן שאלה/מודל/הסקור שהגדרנו, מצאו נקודת אופטימום

- מצא עימוד אופטימלי (=מרחק מינימלי) בין החלבון הנתון

וחלבון אחר

5. שערוד הפרמטרים במודל

- כמה "עולה" להחליף אות כלשהי ברצף ביולוגי? לאבד?
- שערוד הסתברות אירועים
- בהסתמך על איזה מאגרים?
- שואפים למודל "נכון" של אבולוציה של רצפים

5. שערוד הפרמטרים במודל

- כמה "עולה" להחליף אות כלשהי ברצף ביולוגי? לאבד?
- שערוד הסתברות אירועים
- בהסתמך על איזה מאגרים?
- שואפים למודל "נכון" של אבולוציה של רצפים

6. הערכת מובהקות סטטיסטית

- בהנתן מודל + פרמטרים ומאגר מידע, האם מפתיע אותנו לגלות מאורע כזה [או טוב ממנו] באקראי? [p-value]
- בהנתן רצף של 20 בסיסים מגנום האדם, האם מפתיע לגלותם [100% זהות] בגנום הבצל היבש?
- האם מפתיע לגלות רצף זהה של 20 בסיסים בגנום האדם והבצל היבש?

7. ויזואליזציה ופרשנות

- הצגת הדאטה באופן ברור וחסר הטיות
- עימוד דוגמאות ספציפיות (טובות, אקראיות)
- התפלגות ציוני עימוד
- מציאת בעיות בדאטה או הטיות במודל
- זיקוק תובנות ביולוגיות

7. ויזואליזציה ופרשנות

- הצגת הדאטה באופן ברור וחסר הטיות
- עימוד דוגמאות ספציפיות (טובות, אקראיות)
- התפלגות ציוני עימוד
- מציאת בעיות בדאטה או הטיות במודל
- זיקוק תובנות ביולוגיות

עצה לחיים: לפני הכל, הסתכלו על הדאטה בעיניים.

לשיעור של יום שלישי

- אנא שילחו שאלות ותהיות בנוגע לקורס, לנהלים, לדרישות, ובכלל. [tommy@cs.huji.ac.il]
- אנא ענו על הבוחן
- אנא השלימו את המטלה: יישמו המתודה המדעית [על שבעת שלביה, אם רלוונטיים] לבעיה לפי בחירתכם. [לאו דווקא קשור לביולוגיה או חומר הקורס. כל שאלה מדעית שתוכלו לנסח, לתאר דאטה רלוונטי, מודל, ופרמטרים זה סבבה. למשל "האם לכל המשפחות סיכוי 50/50 לבנים/בנות?".
- ניפגש ביום שלישי ברוטברג סי-220 בשעה 11.00.