

## מבוא (קצרצר) לדנ"א

המידע הגנומי שלנו מקודד במולקולות DNA המסודרות במבנה הקרוי כרומוזומים. במשך מרבית חיי התא, כרומוזום מורכב ממולקולות DNA יחידה וארוכה המסודרת במבנה כפול. בתא רגיל של האדם יש 46 כרומוזומים, מורכבים מ-22 זוגות של כרומוזומים כמעט זהים, וזוג אחד של כרומוזומים הזחים אצל נקבות (XX) אך שונים אצל זכרים XY. מולקולות DNA נראית כשתי שרשראות מקבילות, כאשר כל חרוז (נוקלאוטיד) מורכב מבסיס גרעין (nucleobases), פוספט, וסוכר. הפוספט של הבסיס הקודם מתקשר לסוכר של הבסיס הבא ליצירת המחרוזת. המידע הגנומי מצוי בבסיסי-הגרעין: ישנם ארבעה סוגים של בסיס גרעין, והם יכולים להקשר בכל רצף. הם מסומנים באותיות הלועזיות A,C,G,T

( A-adenine, C-cytosine, G-guanine, T-thymine ). השרשרת המקבילה מקודדת לכיוון ההפוך (ראו ציור), ומהווה תמונת ראי: מול כל בסיס A ניצב בסיס T, ומול כל בסיס C נמצא בסיס G. אם כן, אפשר לתאר את המידע הגנטי בכרומוזום כשרשרת ארוכה של אותיות ACCTTAGTTG... כמה ארוכה? הכרומוזום הארוך ביותר אצל אדם נקרא כרומוזום 1, והוא באורך של כ-250 מיליון בסיסים. הקצר ביותר (למעט כרומוזום X) באורך של כ-48 מיליון בסיסים. בכל תא בגוף נמצאים 23 זוגות כרומוזומים זהים, והם שמניעים את כל מערכת ייצור החלבונים בתא. המנגנון העיקרי הוא באמצעות גנים – גן הוא אזור בדנ"א שמכיל הוראות לייצור חלבון. ממעוף הציפור, חלבונים בתא נוצרים ע"פ עותקים של שרשרת הבסיסים המרכיבה את הגן: כל שלשה של בסיסים מקודדת לחומצה אמינית אחת, וביחד רצף של חומצות אמינו יוצרות חלבון.

המבנה הכפול של שרשרת ה-DNA הוא שמאפשר את ההעתקה שלה. כאשר תא מתחלק לשני תאים, גם השרשרת הכפולה בכל כרומוזום מתחלקת לשתי שרשראות בודדות (בהן יש אותו מידע), וכל תא חדש מקבל אחת השרשראות. בתא החדש, נוצר שוב השרשרת המתאימה בצד השני כאשר לכל בסיס פנוי מתחבר הבסיס הנגדי לו, ומתקבל כרומוזום מכל שרשרת בודדת.

## מספר העותקים של אזור גנומי

כאשר משווים כרומוזום בין אנשים או בין תאים של אותו אדם, מוצאים שני סוגי שינויים עיקריים. הראשון, והקל יותר לזיהוי, אלו מוטציות של בסיס בודד. סוג השינוי השני כולל תזוזה של רצפים; אנחנו נשים דגש על הוספה או החסרה של רצפים. הוספה או החסרה הן טעויות העתקה: חלק ממחרוזת הבסיסים הועתקה יותר מדי פעמים (copy number gain), ומופיעה במספר פעמים בחומר הגנטי של התא. גם התופעה ההפוכה יכולה לקרות, כאשר מחרוזת בסיסים לא הועתקה כלל וחסרה (החסרה – copy number loss). המחרוזות ששוכפלו יכולות להיות בקני מידה מגוונים, החל

מעשרות בסיסים ועד כרומוזומים שלמים. זהו חלק מהשוני שבין בני אדם בריאים; אולם בעת חלת הסרטן, כשמערכת הביקורת על העתקת התאים פגיעה, הטעויות האלו נפוצות הרבה יותר. לעיתים, אין להעתקות אלו השפעה כלל על תפקוד התא. אולם, פעמים אחרות מספר עותקים עודף או חסר יכול ליצור חוסר איזון בפעילות התא. לדוגמה, נמצא במחקר משנת 2005 בחנו 9 חולים גידולים סרטניים וברובם נמצא שהגן המייצר את הורמון הגדילה מועתק יותר פעמים; כנראה שהדבר גרם לגדילה המואצת של הרקמה הסרטנית. דוגמה נוספת היא תסמונת דאון, תסמונת התפתחותית הגורמת לעיכובים התפתחותיים, קשיים בדיבור והפרעות בריאותיות שונות. תסמונת דאון נגרמת מעותק שלישי (מלא או חלקי) של כרומוזום מספר 21. (כהערת אגב, אומרים שמספר העותקים הרגיל של מקטע הוא 2. קטע ששוכפל פעמיים באחד הכרומוזומים מופיע 3 פעמים, וכו').

#### גנום רפרנס - רצף השוואה של האדם (Human Reference Genome)

גנום רפרנס הוא רצף בסיסים (נוקלאוטידים) שמטרתו לייצג את הגנום האנושי. הרצף הורכב על ידי מדענים בצורה חישובית, מתוך רצפי ה-DNA של פרטים שונים, ולכן לא משקף גנום של פרט בודד אלא קונצזוס. הגרסה הראשונה של הגנום האנושי נוצרה במסגרת human genome project בשנת 2000. הגרסה נערכה בעזרת מימון ממשלתי אמריקאי, שהבטיח שהרצף האנושי יהיה פומבי. כפי שנראה בהמשך הקורס, גרסה כמעט מלאה של הרצף הוא תנאי שמאפשר שיטות רבות של מדידות גנומיות. אנחנו נשתמש בגרסאות יחסית עדכניות, hg18 ו hg19. ניתן להוריד גרסאות של הגנום ללא הרשמה באתר של אוניברסיטת UCSC, <http://hgdownload.cse.ucsc.edu/downloads.html>, או ישירות ל R בעזרת חבילות Bioconductor.