

المبادئ والتطبيقات في  
المعلوماتية الصحية والحيوية:  
اليوم الأول

# ورشة العمل

## اليوم الأول

### المحاور

التعريف بالمعلوماتية الحيوية وأهميتها الطبية.

أنواع البيانات في المعلوماتية الحيوية.

سير عمل المعلوماتية الحيوية في السياق الطبي.

التكامل بين البيانات الجينية والبيانات الإكلينيكية.

التحديات الأخلاقية والفنية لدمج البيانات الجينية.

تطبيق عملي: تجهيز وتنظيم البيانات الجينية والإكلينيكية باستخدام لغة البرمجة R

## اليوم الثاني

### المحاور

تطبيق عملي: التحليل الإحصائي للبيانات وربطها مع المتغيرات

تطبيق عملي: عرض النتائج عن طريق إنشاء الرسومات البيانية وتفسير المخرجات



## المحاور:



تحميل البيانات

1. التعرف بالمعلوماتية الحيوية وأهميتها الطبية.
2. أنواع البيانات في المعلوماتية الحيوية.
3. سير عمل المعلوماتية الحيوية في السياق الطبي.
4. التكامل بين البيانات الجينية والبيانات الإكلينيكية.
5. التحديات الأخلاقية والفنية لدمج البيانات الجينية.
6. مقدمة للغة R البرمجية، مع تنظيف البيانات وترتيبها

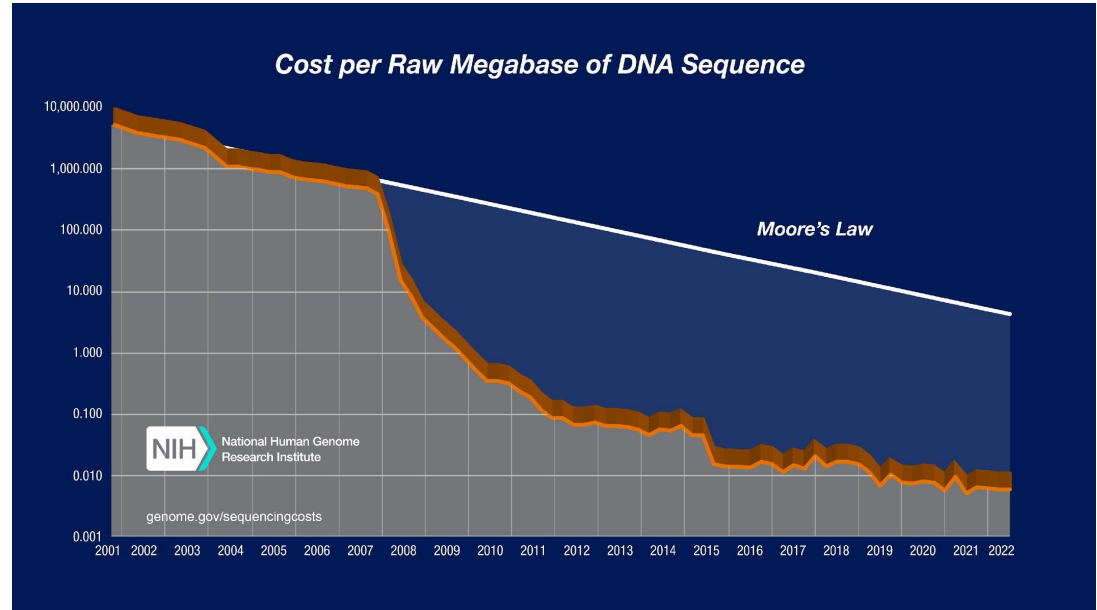
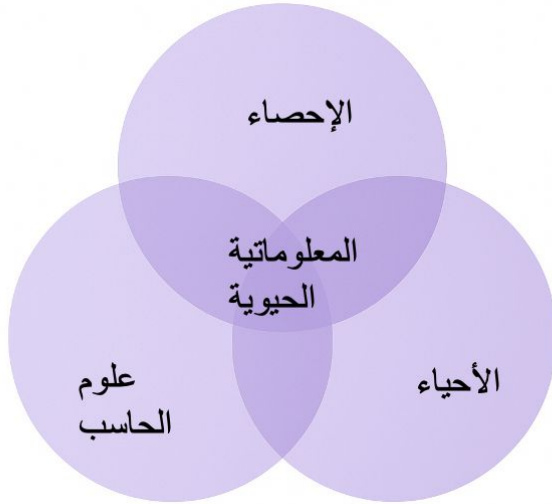


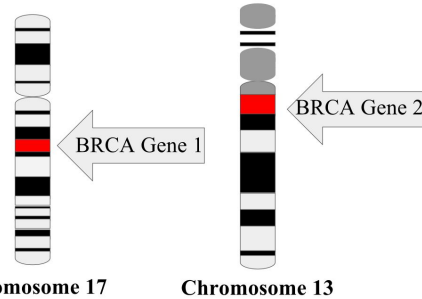
التعريف بالمعلوماتية الحيوية وأهميتها الطبية



# ماهي المعلوماتية الحيوية؟

المعلوماتية الحيوية: استخدام تقنيات الحوسبة والإحصاء والرياضيات لتحليل وتفسير البيانات البيولوجية، خاصة البيانات الجزيئية مثل تسلسل الجينوم.





# المعلوماتية الحيوية الإكلينيكية وأهميتها في الرعاية الطبية

تطوير أدوية وعلاجات دقيقة  
(Precision Medicine, Pharmacogenomics)



تحديد الطفرات المسببة للأمراض  
(مثل الطفرات الجينية في السرطانات والأمراض الوراثية)



دعم القرارات الإكلينيكية.



تحسين التشخيص المبكر والدقة  
(خصوصًا في السرطانات والأمراض النادرة)



# مقارنة بين المعلوماتية الصحية والمعلوماتية الحيوية

المجال	التركيز	أنواع البيانات
المعلوماتية الصحية	أنظمة الرعاية الصحية، إدارة بيانات المرضى، ودعم اتخاذ القرار السريري. تحسين جودة الرعاية الصحية وتعزيز كفاءة تقديم الخدمات	السجلات الطبية، التصوير الطبي، بيانات إدارة الرعاية الصحية
المعلوماتية الحيوية	تحليل وتفسير البيانات البيولوجية على المستويات الجزيئية والجينومية. المساهمة في التقدم في التكنولوجيا الحيوية والطب الدقيق.	تسلسلات بيولوجية ، بيانات وظيفية مثل DNA و RNA والبروتينات والمسارات الحيوية.

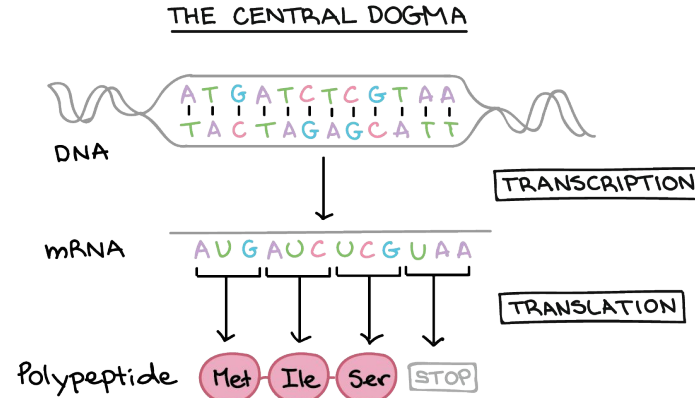




## المحور الثاني: أنواع البيانات في المعلوماتية الحيوية



# أنواع البيانات الحيوية

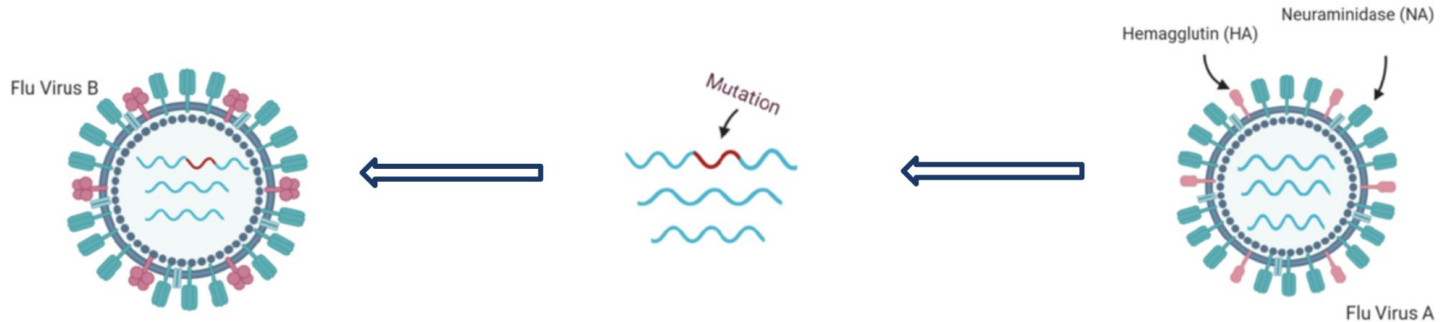
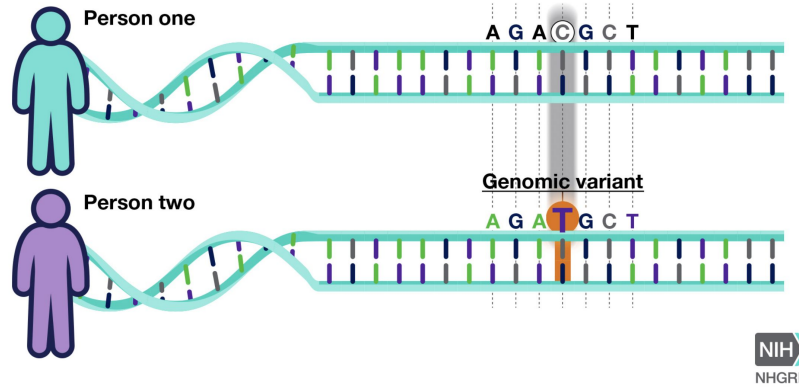


النوع	الوصف	مثال
Genomics	DNA تسلسل	بيانات WGS / WES
Transcriptomics	التعبير الجيني	RNA-seq
Proteomics	دراسة البروتينات	LC-MS
Metabolomics	دراسة الأيض	NMR/GC-MS

Source: <https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/translation/a/intro-to-gene-expression-central-dogma>



# ماهو التباين الجيني (Genomic variation) ؟

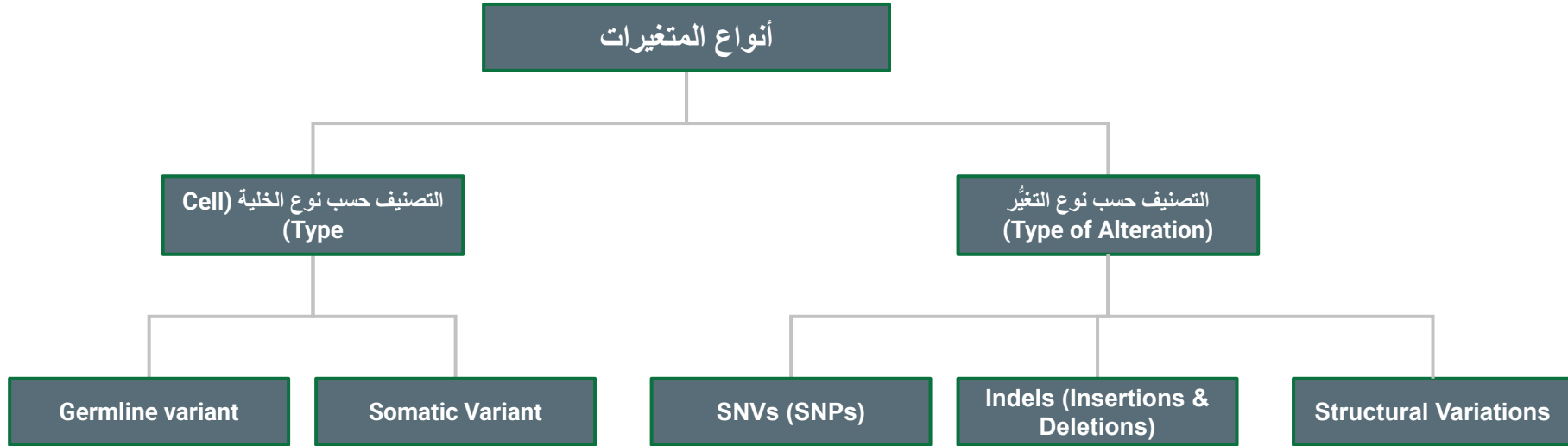




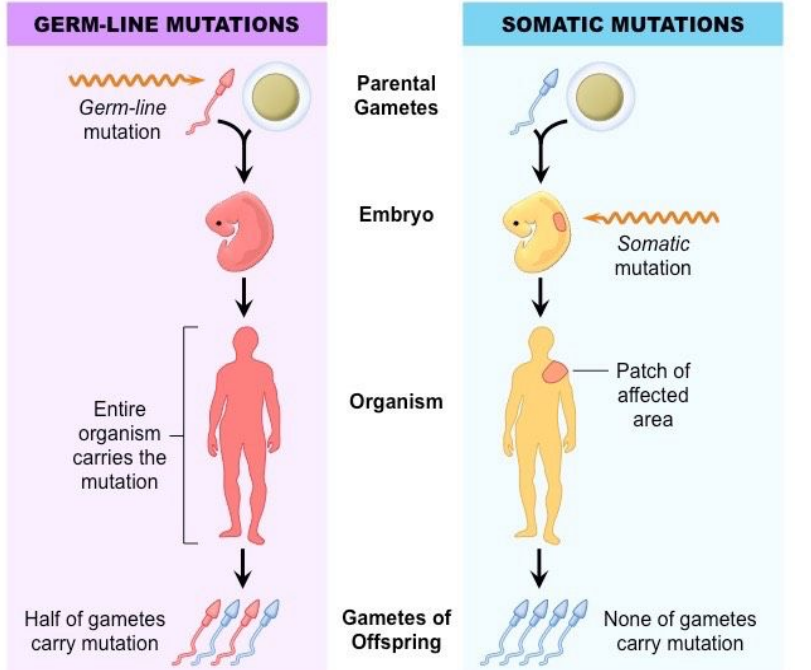
استخلاص المتغيرات

# ماهي المتغيرات ؟

اختلافات في تسلسل الحمض النووي مقارنةً بالجينوم المرجعي.



# التصنيف حسب نوع الخلية (Cell Type)



Source: <https://old-ib.bioninja.com.au/standard-level/topic-3-genetics/33-meiosis/somatic-vs-germline-mutatio.html>

## :Germline Variants

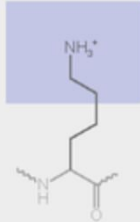
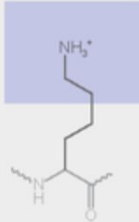
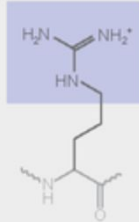
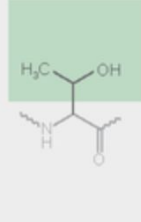
- تنشأ في الخلايا التناسلية (البويضات أو الحيوانات المنوية) وتُورث للأجيال القادمة.
- تكون موجودة في كل خلايا الجسم.

## :Somatic Variants

- حدث في الخلايا الجسدية خلال حياة الفرد بسبب عوامل مثل الطفرات البيئية أو الأخطاء أثناء انقسام الخلايا.
- لا تُورث للأجيال القادمة.
- توجد فقط في مجموعة فرعية من خلايا الجسم.



# تصنيف المتغيرات حسب تأثيرها على البروتين

	No Mutation لا توجد طفرة	Silent طفرة صامتة	Nonsense طفرة غير منطقية	Missense (طفرة مغلطة) مُغيرة للمعنى	
DNA level	TTC	TTT	ATC	TCC	TGC
mRNA level	AAG	AAA	UAG	AGG	ACG
protein level	Lys	Lys	STOP	Arg	Thr
					
				basic	polar

## Frameshift

تؤدي هذه الطفرة إلى تغيير كل الأحماض الأمينية التي سبقت بعد نقطة الطفرة، مما يؤدي غالبًا إلى إنتاج بروتين طويل جدًا أو قصير جدًا أو مشوه تمامًا



## المحور الثالث: سير عمل المعلوماتية الحيوية في السياق الطبي





معالجة ملفات الجينوم

المعالجة المسبقة  
للبيانات

استخلاص المتغيرات  
وتحليلها

## خطوات سير العمل

جمع البيانات الجينية (WGS, WES, RNA-seq)



ضبط الجودة ومعالجة البيانات (Quality Control)



اكتشاف المتغيرات الجينية (Variant Calling)



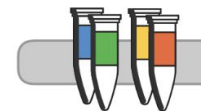
تحليل البيانات الإحصائي أو باستخدام الذكاء الاصطناعي



تفسير النتائج وربطها بالتشخيص أو العلاج



Sequencing



DNA Samples





# صيغة ملف MAF

- **Mutation Annotation Format (MAF):** ملف نصي مفصول بعلامة الجدولة (,) يحتوي على بيانات الطفرات الجينية.

- يتم توليده من ملفات **VCF** على مستوى المشروع عبر تجميع وتنقيح الطفرات (Somatic Aggregation Workflow).

Metadata

Header

Data

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=ID=20,length=62435964,assembly=B36,m55-f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NAO0001 NAO0002 NAO0003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0/0:48:1:51,51 1/0:48:8:51,51 1/1:43:5:...
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0/0:49:3:58,60 0/1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1/2:21:6:23,27 2/1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0/0:54:7:56,60 0/0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```



VCF File Format

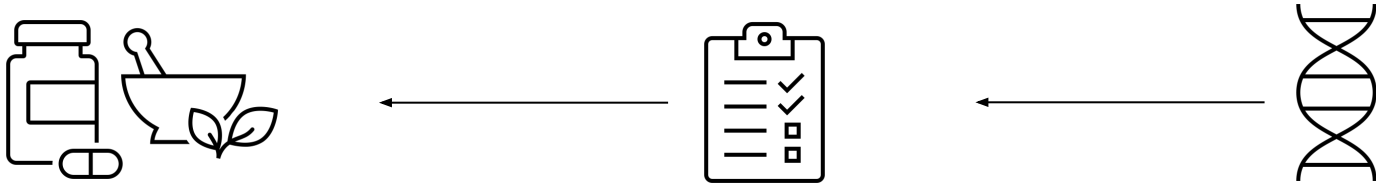


## المحور الرابع: التكامل بين البيانات الجينية والبيانات الإكلينيكية



# أهمية التكامل

- البيانات الجينية تفسر الأساس البيولوجي للأمراض.
  - البيانات الإكلينيكية توضح الأعراض والتاريخ الطبي للمريض.
  - الدمج بينهما = صورة شاملة تساعد في
1. التشخيص الدقيق.
  2. التنبؤ بمسار المرض وتقليل التكاليف.
  3. اختيار العلاج الأمثل.



# أمثلة:

1. الأورام: دمج الطفرات الجينية Somatic مع السجل الطبي —> اختيار علاج دقيق يناسب الورم.
2. أمراض القلب والسكري: التاريخ العائلي و مؤشرات إكلينيكية (ضغط دم، سكر) + SNPs —> التنبؤ بالمخاطر والمضاعفات بشكل أدق.
3. الأدوية (Pharmacogenomics): اختلاف جيني في إنزيم دوائي + بيانات وصف العلاج —> تعديل الجرعة أو اختيار دواء بديل.
4. الأمراض النادرة: تحليل تسلسل الجينوم + الأعراض الإكلينيكية —> تسريع التشخيص المرض النادر الذي قد يتأخر سنوات.



# المعايير العالمية لتكامل البيانات الجينية والإكلينيكية:

## Health Level Seven (HL7)

- معيار لتبادل السجلات الطبية الإلكترونية بين الأنظمة المختلفة.
- يضمن سهولة مشاركة بيانات المريض (تحاليل، فحوصات، زيارات).
- يُستخدم على نطاق واسع في المستشفيات والعيادات.



**Global Alliance**  
for Genomics & Health  
Collaborate. Innovate. Accelerate.

<https://www.ga4gh.org/>

## Global Alliance for Genomics and Health (GA4GH)

- تحالف عالمي لتوحيد تبادل البيانات الجينية.
- يضع معايير تسمح بمشاركة بيانات الجينوم بين المراكز البحثية والسريية.
- يسهل التعاون الدولي في الأبحاث والتشخيص.

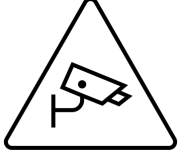


## المحور الخامس: التحديات الأخلاقية والفنية



# التحديات الأخلاقية

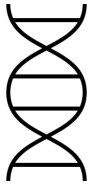
خصوصية البيانات الجينية والسجلات الصحية الإلكترونية: إمكانية التعرف على هوية المريض.



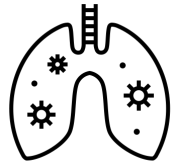
الموافقة المستنيرة **Informed Consent**: استخدام بيانات المرضى في البحث.



التمييز الجيني (**Genetic Discrimination**): خطر استغلال المعلومات الجينية من قبل جهات مثل شركات التأمين أو جهات العمل.

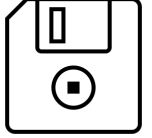


التنبؤ بالأمراض: ماذا لو أظهرت البيانات احتمالية إصابة شخص بمرض مستقبلي؟



# التحديات الفنية

**حجم البيانات الضخم** وصعوبة تخزينها ومعالجتها : جينوم واحد  $\approx$  مئات الجيجابايت، مشاريع كبيرة  $\approx$  تيرابايت/بيتابايت



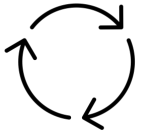
**جودة البيانات:** بيانات ناقصة، غير دقيقة، أو غير متناسقة.



**تنوع مصادر البيانات:** صور طبية، سجلات إلكترونية، بيانات جينية



نقص المعايير الموحدة للتكامل بين الأنواع المختلفة من البيانات دمج الجينوم مع Multi-Omics : البروتيوم و البيانات الإكلينيكية





إستراحة 10 دقائق قبل البدء بالجزء العملي

