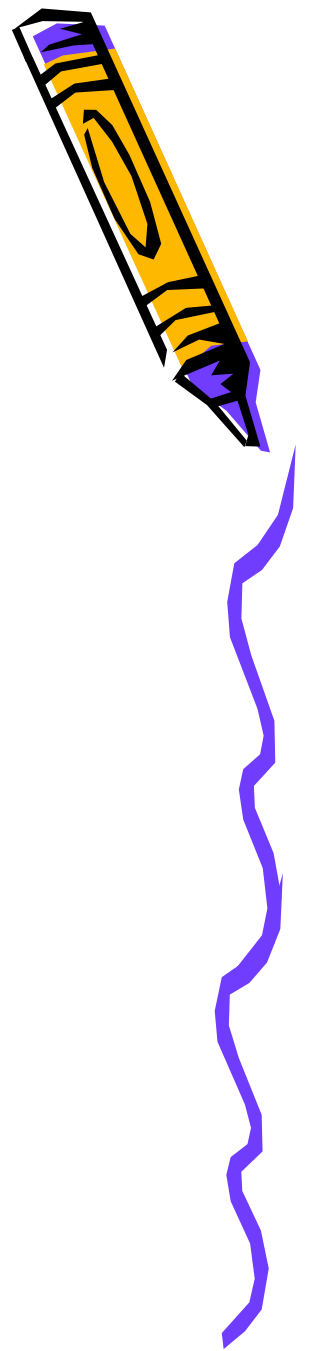


Computational Molecular Biology

- 計算分子生物學
- This is a course about the applications of Combinatorial Mathematics (組合數學) or Discrete Mathematics (離散數學) on Molecular Biology.
- **Not much Biology** will be involved.
- The main topic is "the sequencing" of DNA sequences.



DNA (去氧核糖核酸)



- DNA: Deoxyribonucleic Acid
- Nucleotide (核苷酸)

Adenine (A) : 腺嘌呤

Thymine (T) : 胸腺嘧啶

Guanine (G) : 鳥嘌呤

Cytosine (C) : 胞嘧啶



Coupled



- A on a strand is always coupled with a T on the other strand, and a C on a strand is always coupled with a G on the other strand.
- Those two strands are said to be **complimentary**.
- A DNA molecule is usually represented as a **single sequence**.



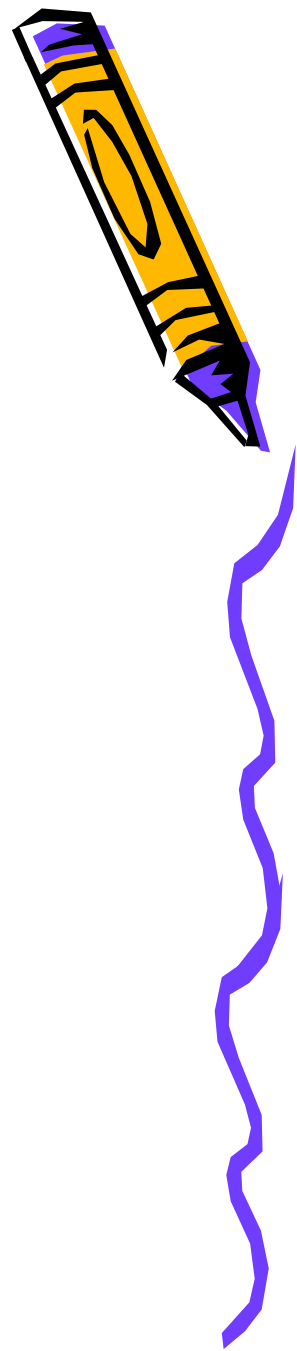
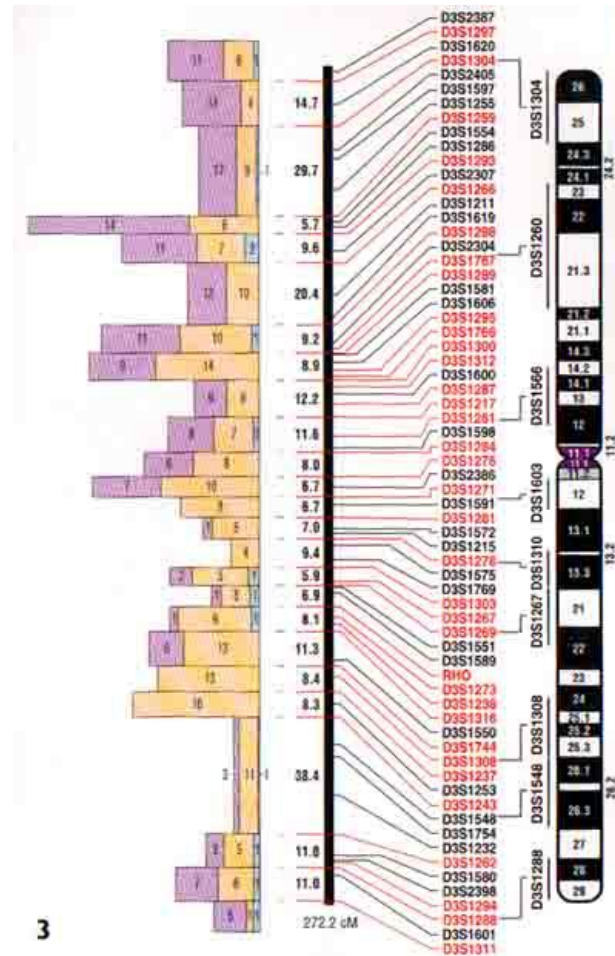
Chromosome (染色體)



- Chromosomes are made of DNA (deoxyribonucleic acid, 去氧核糖核酸), a double stranded molecule in which each strand is a long succession (sequence) of nucleotides.
- A - T, G - C



Human Chromosom 3



Genome



- The "molecules of heredity" (genetic information) are present in every cell of all living organisms.
- Each molecule is called a "chromosome".
- The genome is the set of all chromosomes.



Genes



- A "special" sequence of DNA is called a **gene**.
- Genes contain instructions of making **proteins**.
- Proteins act alone or in complexes to perform many cellular functions.



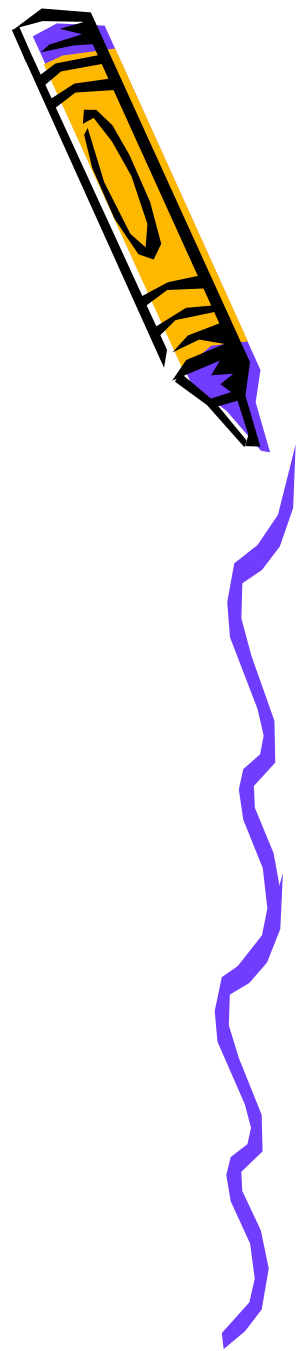
About Proteins



- Proteins are large biomolecules, or macromolecules, consisting of one or more long chains (sequences) of amino acid (氨基酸) residues.
- Proteins perform a vast array of functions within organisms: catalyzing metabolic reactions (新陳代謝), DNA replication and responding to stimuli.



Inside Proteins (蛋白質)

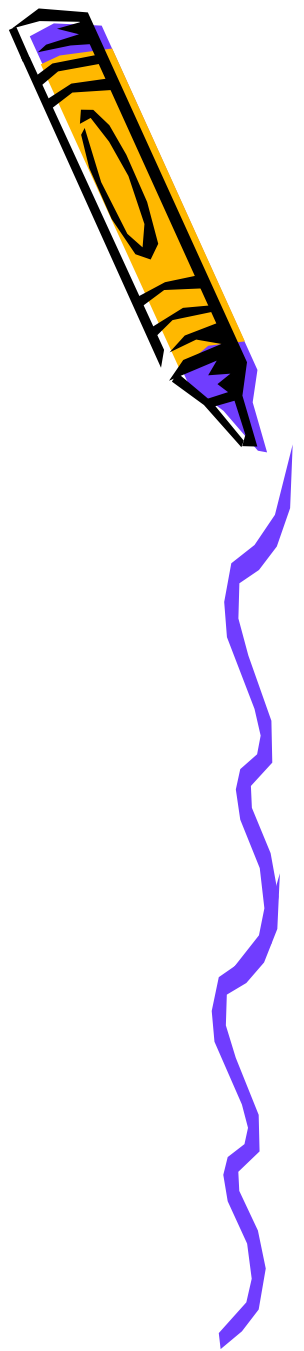


- The twenty *amino acids* commonly found in proteins:
 - **A** **Ala** Alanine
 - **C** **Cys** Cysteine
 - **D** **Asp** Aspartic Acid
 - **E** **Glu** Glutamic Acid



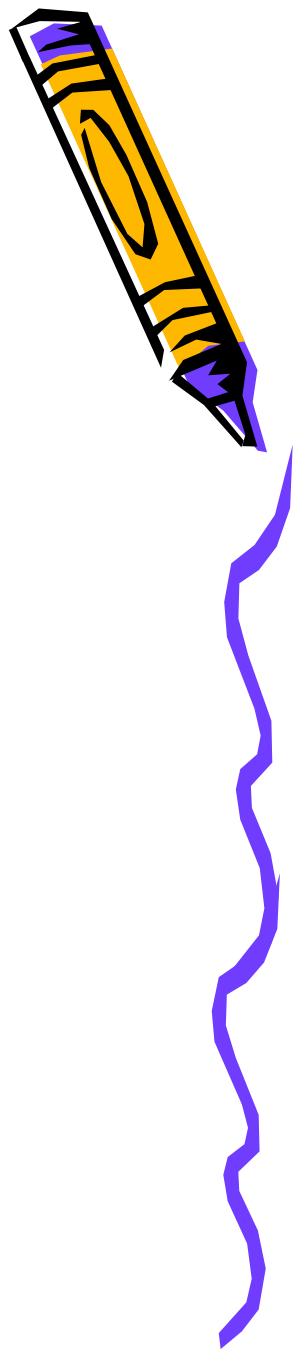
Continued ...

- F Phe Phenylalanine
- G Gly Glycine
- H His Histidine
- I Ile Isoleucine
- K Lys Lysine
- L Leu Leucine
- M Met Methionine
- N Asn Asparagine
- P Pro Proline

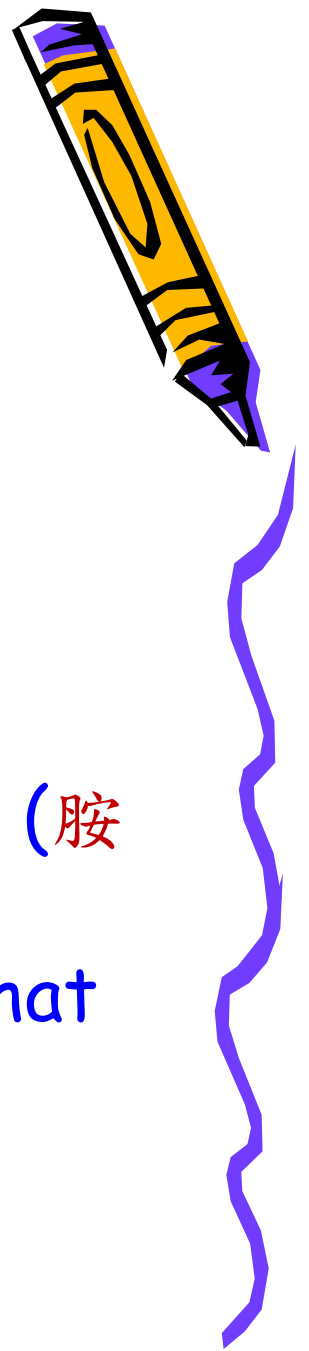


Continued ...

- Q Gln Glutamine
- R Arg Arginine
- S Ser Sreine
- T Thr Threonine
- V Val Valine
- W Trp Trptophan
- Y Tyr Tyrosine



More functions



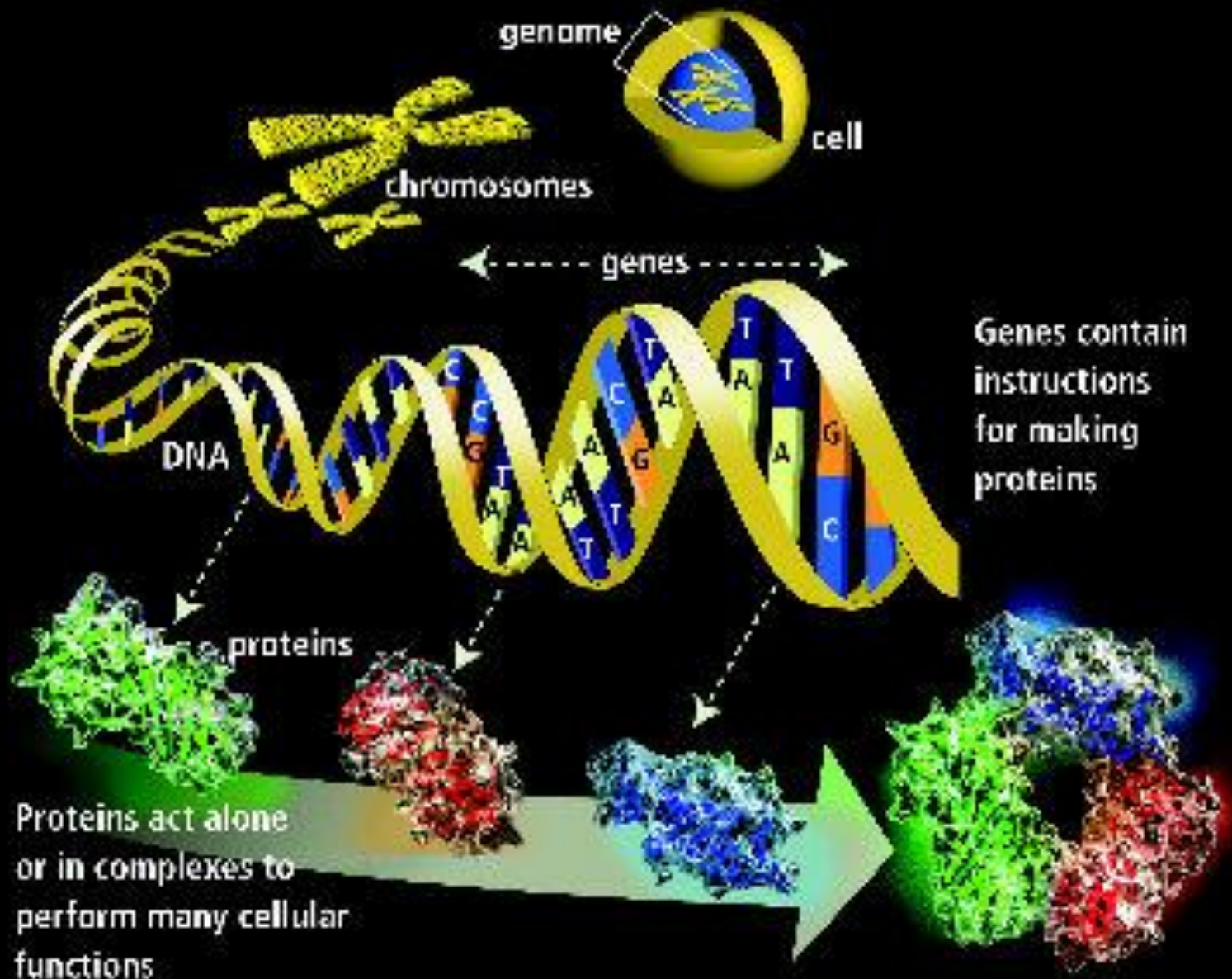
- Proteins provide structure to cells and organisms, and transporting molecules from one location to another.

(*) Proteins differ from one another primarily in their sequence of amino acids (氨基酸) of their genes results in **protein folding** into a specific 3-dim. Structure that determines its activity.



細胞與染色體

- 細胞：生物分子組合形成了能自我繁殖的單元。
- 細胞膜上有各式各樣的蛋白質，形成通道，讓細胞內外的分子能進進出出。
- 細胞裡面的DNA包裝染色體的構造。
- DNA 的資訊以3個字母唯一單元，稱為密碼子（codon），每個密碼子指定一種胺基酸。



U.S. DEPARTMENT OF ENERGY



Human Genome Project

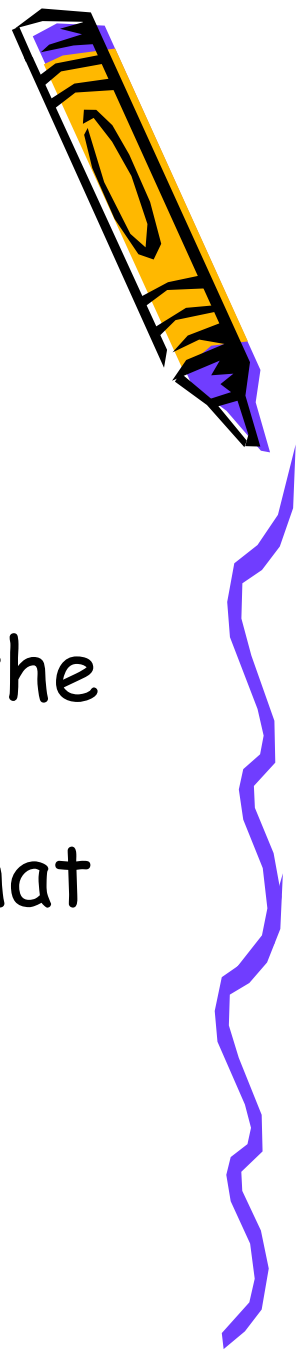
How does the human genome stack up?



Organism	Genome Size (Base pairs)	Estimated Genes
Human (人類)	3 billion	30,000
Laboratory mouse (白老鼠)	2.6 billion	30,000
Mustard weed (<i>A. thaliana</i>)	100 million	25,000
Roundworm (<i>C. elegans</i>)	97 million	19,000
Fruit fly (果蠅)	137 million	13,000
Yeast (酵母菌)	12.1 million	6,000
Bacterium (大腸桿菌)	4.6 million	3,200
Human immunodeficiency virus (HIV)	9700	9



Important Issues



- Develop new technologies in sequencing the DNA sequences.
- Transfer related technologies to the "private" sector, and address the ethical, legal, and issues (social) that may arise from **HGP**.



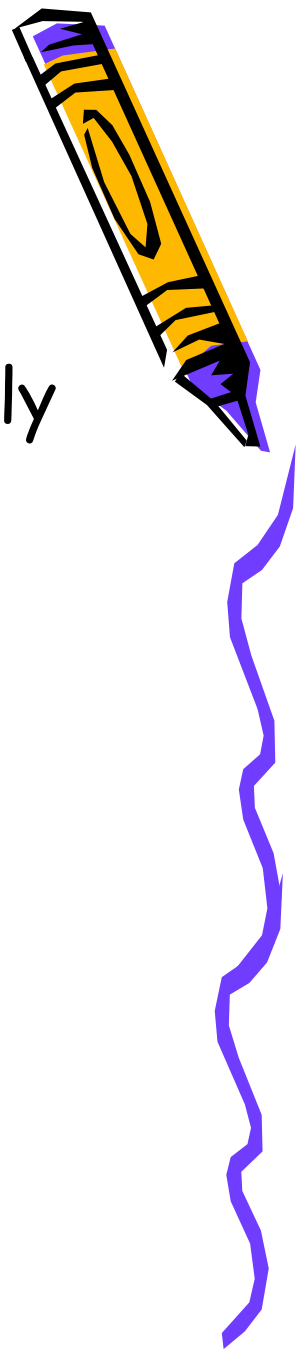
Goals



- Identify all the approximate 30,000 genes in human DNA.
- Determine the sequences of all 3 billion chemical base pairs that make up the human DNA.
- Store this sequence (databases).
- Improve tools for data analysis.



Milestones

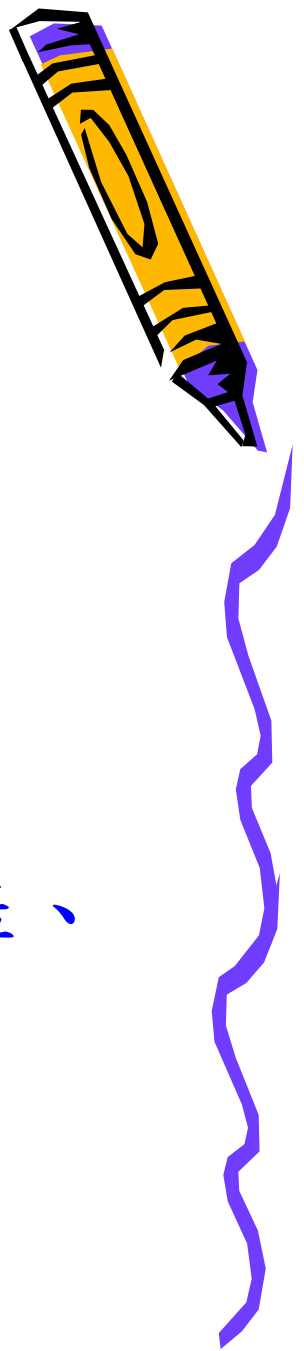


- In 1990, the project was initiated jointly by U. S. Department of Energy and the National Institute of Health.
- A working draft was completed of the entire human genome in 2000 and the analyses are published one year later.
- In April 2003, HGP sequencing is completed and the project is declared finished two years ahead of schedule.



主要課題

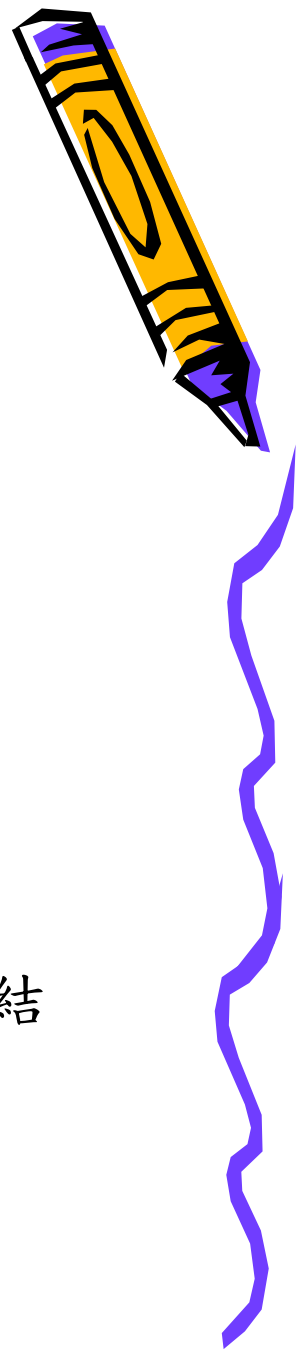
- 序列組合
- 序列分析
- 基因認定
- 生物資訊資料庫、種族樹建構、蛋白質三維結構推測…。



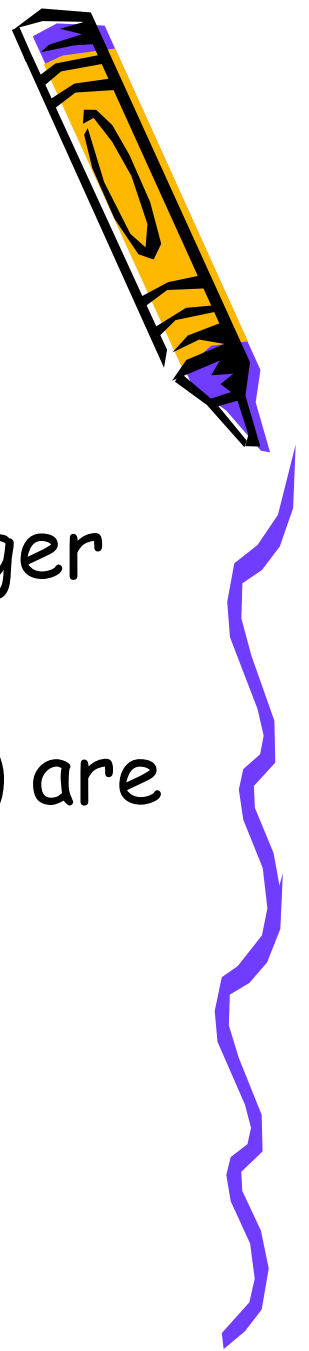
序列比對探究

• 說明

- 找序列中「相似」及「相異」的部份。
- 為何要比對序列？
- 為何要使用電腦比對？
- 如何使用電腦比對？
- 困難點：序列型態多樣、需建構不同的資料結構及演算法。如何讀取序列也是一項挑戰！



Challenges



- A sequence may evolve (演化) by modifying its organization at a larger scale.
- These large-scale mutations (突變) are called **structural variations** or **rearrangements** in the terms of modeling.



Types of rearrangements



1. **Deletions:** A segment of the genome is lost.
2. **Transpositions:** A segment of the genome moves to another location.
3. **Inversions or reversals:** A segment of the genome is reversed and the strands are exchanged.



Continued ...



4. Duplications

- (a) **Tandem duplications**: insert the copy next to the original.
- (b) **Retro-transpositions**: insert a copy of a gene at an arbitrary location in the genome.
- (c) **Whole genome duplications**: Copy either the whole genome or some of its chromosomes.



More



5. Reciprocal translocation: A segment of a chromosome that contains a **telomere** is exchanged with a segment of another chromosome that also contains a **telomere**.

Telomere: 端粒是真核生物染色體末端的DNA重複序列，作用是保持染色體的完整性和控制細胞分裂週期。由於DNA複製的機制，每次染色體複製後，延遲股上的染色體末端必無法被複製。因此，真核生物在染色體末端演化出端粒以作為可被重複遺棄的部分。



More

6. Fusion: Two chromosomes are joined into one.

7. Fission: One chromosome splits into two (inverse of fusion).

8. Horizontal transfer: A segment of the genome is copied from one genome to another. (This is common mainly in unicellular organisms.)



Notices



- All the operations act on a segment **at the level of DNA segments** rather on nucleotides.
- This is why a genome is often represented by a sequence of segments **that are found in an almost identical state in several species, not cut by rearrangements.**



Homologous (同族)



- Two segments are said to be homologous if they derive from a common ancestor and are distinguished by a replication event (they end up in two different genomes) or a duplication event (they both belong to the same genome).
- Genes are often taken as those homologous segments due to their functional utility (rarely cut by rearrangement).





Genomics and Its Impact on Science and Society: *The Human Genome Project and Beyond*

U.S. Department of Energy Genome Programs
<http://doegenomes.org>





Human Genome Project

Goals:

- identify all the approximate 30,000 genes in human DNA,
- determine the sequences of the 3 billion chemical base pairs that make up human DNA,
- store this information in databases,
- improve tools for data analysis,
- transfer related technologies to the private sector, and
- address the ethical, legal, and social issues (ELSI) that may arise from the project.

Milestones:

- 1990: Project initiated as joint effort of U.S. Department of Energy and the National Institutes of Health
- June 2000: Completion of a working draft of the entire human genome
- February 2001: Analyses of the working draft are published
- April 2003: HGP sequencing is completed and Project is declared finished two years ahead of schedule



What does the draft human genome sequence tell us?

By the Numbers

- The human genome contains 3 billion chemical nucleotide bases (A, C, T, and G).
- The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases.
 - The total number of genes is estimated at around 30,000--much lower than previous estimates of 80,000 to 140,000.
- Almost all (99.9%) nucleotide bases are exactly the same in all people.
- The functions are unknown for over 50% of discovered genes.



What does the draft human genome sequence tell us?

How It's Arranged

- The human genome's gene-dense "urban centers" are predominantly composed of the DNA building blocks G and C.
- In contrast, the gene-poor "deserts" are rich in the DNA building blocks A and T. GC- and AT-rich regions usually can be seen through a microscope as light and dark bands on chromosomes.
- Genes appear to be concentrated in random areas along the genome, with vast expanses of noncoding DNA between.
- Stretches of up to 30,000 C and G bases repeating over and over often occur adjacent to gene-rich areas, forming a barrier between the genes and the "junk DNA." These CpG islands are believed to help regulate gene activity.
- Chromosome 1 has the most genes (2968), and the Y chromosome has the fewest (231).



What does the draft human genome sequence tell us?

The Wheat from the Chaff

- Less than 2% of the genome codes for proteins.
- Repeated sequences that do not code for proteins ("junk DNA") make up at least 50% of the human genome.
- Repetitive sequences are thought to have no direct functions, but they shed light on chromosome structure and dynamics. Over time, these repeats reshape the genome by rearranging it, creating entirely new genes, and modifying and reshuffling existing genes.
- The human genome has a much greater portion (50%) of repeat sequences than the mustard weed (11%), the worm (7%), and the fly (3%).



What does the draft human genome sequence tell us?

How the Human Compares with Other Organisms

- Unlike the human's seemingly random distribution of gene-rich areas, many other organisms' genomes are more uniform, with genes evenly spaced throughout.
- Humans have on average three times as many kinds of proteins as the fly or worm because of mRNA transcript "alternative splicing" and chemical modifications to the proteins. This process can yield different protein products from the same gene.
- Humans share most of the same protein families with worms, flies, and plants; but the number of gene family members has expanded in humans, especially in proteins involved in development and immunity.
- Although humans appear to have stopped accumulating repeated DNA over 50 million years ago, there seems to be no such decline in rodents. This may account for some of the fundamental differences between hominids and rodents, although gene estimates are similar in these species. Scientists have proposed many theories to explain evolutionary contrasts between humans and other organisms, including those of life span, litter sizes, inbreeding, and genetic drift.



What does the draft human genome sequence tell us?

Variations and Mutations

- Scientists have identified about 3 million locations where single-base DNA differences (SNPs) occur in humans. This information promises to revolutionize the processes of finding chromosomal locations for disease-associated sequences and tracing human history.
- The ratio of germline (sperm or egg cell) mutations is 2:1 in males vs females. Researchers point to several reasons for the higher mutation rate in the male germline, including the greater number of cell divisions required for sperm formation than for eggs.



Future Challenges: What We Still Don't Know

- Gene number, exact locations, and functions
- Gene regulation
- DNA sequence organization
- Chromosomal structure and organization
- Noncoding DNA types, amount, distribution, information content, and functions
- Coordination of gene expression, protein synthesis, and post-translational events
- Interaction of proteins in complex molecular machines
- Predicted vs experimentally determined gene function
- Evolutionary conservation among organisms
- Protein conservation (structure and function)
- Proteomes (total protein content and function) in organisms
- Correlation of SNPs (single-base DNA variations among individuals) with health and disease
- Disease-susceptibility prediction based on gene sequence variation
- Genes involved in complex traits and multigene diseases
- Complex systems biology including microbial consortia useful for environmental restoration
- Developmental genetics, genomics



Anticipated Benefits of Genome Research

Molecular Medicine

- improve diagnosis of disease
- detect genetic predispositions to disease
- create drugs based on molecular information
- use gene therapy and control systems as drugs
- design “custom drugs” (pharmacogenomics) based on individual genetic profiles

Microbial Genomics

- rapidly detect and treat pathogens (disease-causing microbes) in clinical practice
- develop new energy sources (biofuels)
- monitor environments to detect pollutants
- protect citizenry from biological and chemical warfare
- clean up toxic waste safely and efficiently



Anticipated Benefits of Genome Research-cont.

Risk Assessment

- evaluate the health risks faced by individuals who may be exposed to radiation (including low levels in industrial areas) and to cancer-causing chemicals and toxins

Bioarchaeology, Anthropology, Evolution, and Human Migration

- study evolution through germline mutations in lineages
- study migration of different population groups based on maternal inheritance
- study mutations on the Y chromosome to trace lineage and migration of males
- compare breakpoints in the evolution of mutations with ages of populations and historical events



Anticipated Benefits of Genome Research-cont.

DNA Identification (Forensics)

- identify potential suspects whose DNA may match evidence left at crime scenes
- exonerate persons wrongly accused of crimes
- identify crime and catastrophe victims
- establish paternity and other family relationships
- identify endangered and protected species as an aid to wildlife officials (could be used for prosecuting poachers)
- detect bacteria and other organisms that may pollute air, water, soil, and food
- match organ donors with recipients in transplant programs
- determine pedigree for seed or livestock breeds
- authenticate consumables such as caviar and wine



Anticipated Benefits of Genome Research-cont.

Agriculture, Livestock Breeding, and Bioprocessing

- grow disease-, insect-, and drought-resistant crops
- breed healthier, more productive, disease-resistant farm animals
- grow more nutritious produce
- develop biopesticides
- incorporate edible vaccines incorporated into food products
- develop new environmental cleanup uses for plants like tobacco



Medicine and the New Genetics

Gene Testing

Pharmacogenomics

Gene Therapy

Anticipated Benefits:

- improved diagnosis of disease
- earlier detection of genetic predispositions to disease
- rational drug design
- gene therapy and control systems for drugs
- personalized, custom drugs



YGGG-0007

Gene Therapy

基因治療的發展歷程

基因治療是指利用分子生物學方法將特定的遺傳物質導入病患體內，使之達成標的(Target)基因產物，從而使疾病得到治療，為現代醫學和分子生物學相結合而誕生的新技術。

Continued ...

基因療法發展過程中，出現了臨床試驗上的死亡案例，隨著負面效應逐漸擴展，基因療法甚至受到相關領域專家學者的輕視。儘管歐洲監管機構於2012年批准部分這類型的療法，仍因患者在治療後引發嚴重的胰臟炎病症，使得基因療法的效果與安全性備受質疑，相關研究及試驗的腳步也暫時緩下來。

近年來，基因治療在一系列疾病的臨床試驗中表現不俗，包含血友病、鐮刀型紅血球疾病以及多種癌症等，逐漸展現出其應用價值及發展的可能性，使得投資者因此回流，而美國食品藥品管理局(FDA)近年批准的基因療法，無疑是對這領域的鼓勵與肯定。

Methodology

- 1960年Rogers以及Pfunderer證明病毒可以調控基因移轉，進而有了基因治療的概念，1960年代後期，便逐漸發展出一些特定功能的細胞株，其目的為測試外來DNA是否能永久且穩定的表現並遺傳給子代。歷經30年的研究後，人類史上第一次的基因療法於1990年9月首度問世，試驗者為患有遺傳性疾病腺苷脫氨酶(Adenosine Deaminase, ADA)缺乏症的四歲女童。由於患者對細菌、病毒以及真菌皆缺乏免疫力，因此嚴重時可能危害到生命。醫院從患者體內取出白血球，並將能夠製造ADA的基因轉殖入白血球中，再將這些被基因校正過的白血球注射回患者體內，為基因療法歷史上的重要里程碑。

快速發展

2012年，歐盟首度批准第一個基因治療藥物Glybera後，全球每年批准的基因治療臨床試驗已超越過往20年間的最大值，FDA也於2019年1月宣布，預計在2025年前每年批准10到20項細胞及基因治療產品，意味著基因治療展現出其無限的發展性與未來性，導致市場需求量增加。

除市場需求之外，在定價方面，基因治療藥物也要價不斐，以Glybera為例，在2014年正式上市時還是當時最貴的藥物(111萬歐元)，如此高額的定價，是因為需要從有限的患者收回藥物的研發成本，這種高收益的特殊性質，使得投資者抑或是生醫藥廠願意挹注資金及資源進入基因治療領域。

積極發展

目前基因治療使用的載體包含腺相關病毒（Adeno-Associated Virus, AAV）、慢病毒(Lenti Virus, LV)、反轉錄病毒(Retro Virus, RV)等。2018年全球基因治療市場規模約為 5.4 億美元，由於嵌合抗原受體 T細胞療法(Chimeric Antigen Receptor T-cell Therapy, CAR-T)的臨床試驗結果近年屢傳佳報，激勵許多生物製藥開發商及學術研究團體搶進。

根據Roots Analysis 2019年8月發表的研究報告，2030年CAR-T 療法的市場規模預計將有望達到110億美元，2019-2030年的年複合成長率上看 31%。基因、細胞治療市場正以倍數蓬勃發展，市場規模預計於2027 年成長至356.7億美元，是2019年的10倍以上。意味著細胞及基因治療市場正處於急速發展階段。

Two types of therapy (Treatment)

(1) 基因治療分成兩種方式：體外基因治療法(Ex vivo)以及體內基因治療法(In vivo)。體外基因治療(亦可視為細胞治療)是取出患者特定的細胞(如：嵌合抗原受體 T細胞，造血幹細胞等)至體外進行基因工程改良基因缺陷後，再選出機能表現正常且符合需求的細胞，移植回患者體內。

(2) 體內基因治療則是利用事先已經過基因工程改造後的基因，直接注射至患者體內進行治療。體內基因治療又可分為非基因體編輯和基因體編輯兩大類，非基因體編輯利用合成的DNA或RNA(如：RNAi，microRNA等)調節基因功能；基因體編輯則是利用基因工程的方式，插入、移除或取代有缺陷的基因。而在體內基因治療的基因遞送方面可分類成病毒載體(RV、AAV、LV等)及非病毒載體(Liposome、Nanoparticles等)兩種方式。

Among the bests

嵌合抗原受體 T細胞療法(Chimeric Antigen Receptor T-cell Therapy, CAR-T)

嵌合抗原受體 T細胞療法是先從患者的血液中分離出T細胞，接著利用基因工程技術製造出CAR混合基因(目前做法是以LV或 RV送入)，再將CAR基因轉殖到T細胞內，形成嵌合抗原受體 T細胞(CAR-T cell)。將CAR-T細胞在體外增生至足夠的細胞數之後，注射到病人身上，這些CAR-T細胞便可以與癌細胞結合，並消滅癌細胞。

How to implement?

FDA (美國)六項指引及一項草案

為了促進基因治療產品發展的同時，一併確保其安全性與有效性，FDA於2020年1月28日公布六項最新的細胞與基因治療指引以及一項「根據美國孤兒藥法規解釋基因治療產品之相同性」(Interpreting Sameness of Gene Therapy Products under the Orphan Drug Regulations) 草案。該指引及草案提供研發人員與製造業者重要資訊包含了臨床新藥需如何符合相關規定，並公開揭露產品之安全、特性、品質、純度與強度。然而，在六項指引中有三項是為特定疾病基因治療產品之生產指引，例如成人與兒童之血友病、視網膜疾病，以及其他罕見疾病之療法重點，包含廠商須如實提供產品之生產、臨床前期、臨床試驗、臨床試驗設計等重要研究數據與資料。

Details

GMP等級病毒載體生產

病毒載體無論在細胞治療或是基因治療上都至關重要，因此如何大量生產臨床試驗用量甚至是商業GMP等級規模的病毒載體，無疑是門高深的學問。不論最終是否會成功商業化，臨床試驗之生物製劑(或藥品)其生產、製造與管制，phase I、phase II都需符合GMP之精神，phase III應符合「西藥藥品優良製造規範」(PIC/S：Guide to Good Manufacturing Practice for Medicinal Products)之要求。

病毒載體製備流程為一開始將複數個質體(plasmid)共轉染(co-transfection)到宿主細胞(host cell)，依據病毒載體的差異，所需的質體數量也不盡相同，收集中間產物再經由純化手續，去除不純物，進而得到高純度之病毒載體。病毒載體的製造需仰賴複數質體共轉染至宿主細胞，在宿主細胞部分可區分為貼附性以及懸浮性兩大系統，在病毒生產上游階段依據小量生產與大量生產的需求，而有各自匹配的儀器耗材。

Continued ...

由於貼附性與懸浮性系統差異過大，大型工業化生產需要重新摸索所有的條件，如：細胞培養、轉染等。經過轉染步驟，其中間產物含有大量病毒載體外的不純物質，為增強後續的轉導效率，如何移除目標以外的不純物質，也是業界不斷精進的目標。在下游純化的策略以AAV為例可分成兩種方式：一、經由梯度超高速離心步驟，移除不純物，進而獲得較高純度的AAV。二、藉由兩道管柱純化步驟，同樣可以得到較高純度的AAV (Full AAV)。

病毒載體的製造過程中(質體、宿主細胞、中間產物至最終產物)，會進行適當的品質管制，如細胞庫及質體檢驗、產品安全性檢驗等，以確保整個製造過程及最終產物的品質合乎GMP規範。檢驗項目亦隨著檢體來源而有所變異，如宿主細胞的檢驗項目包含無菌測試等；批次放行檢驗則涵蓋較多檢驗項目如純度、品質、安定性等。

Continued ...

在病毒載體的品質方面，因偵測方法、偵測目標序列、宿主細胞株或製程添加物之差異，而造成病毒力價(Viral Vector Titer)或(Transduction Unit)數據略有高低，因此，於製程中，除需確保製造過程的一致性外，也需統一檢驗的偵測方法。針對單一檢驗項目的檢測方式有許多種，如在AAV的純度方面，穿透式電子顯微鏡(Transmission Electron Microscopy, TEM)等，為目前常用的檢測手法，可從外觀上區分出Empty AAV以及Full AAV。

綜合前述，病毒載體之製備從開發、臨床試驗用量乃至商業規模，依序進程為：測試階段→小量研發→大量研發→GMP先導測試→GMP製程。在各階段相對應的檢驗方法，應同時開發並進行確效，確保病毒載體可應用於臨床試驗甚至商品化。

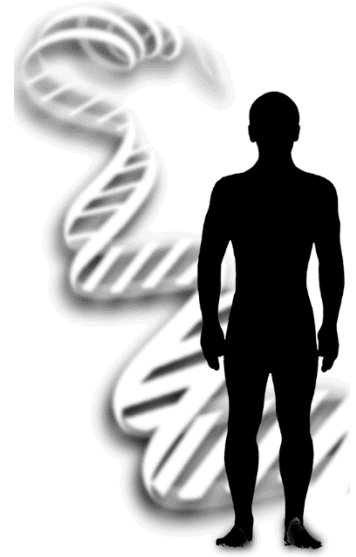
Not an easy job!

由於基因及細胞治療產品的開發步驟，非常繁瑣費時，為了進一步提高研發效率，降低成本以及研發風險，許多製藥公司所採取的策略為：公司進行前期研發，委外CDMO (Contract Development & Manufacturing Organization)公司生產臨床試驗用量或是商業規模的病毒載體。且臨床試驗及商用規模之產品皆需符合GMP精神與規範，CDMO公司可以協助製藥公司節省漫長的廠房建設時間，使產品以最快的速度推向市場，搶占市場先機。優質的CDMO對於整個基因及細胞治療產品，從前端研發到臨床前期產品生產均可介入，包括：生產細胞建立、細胞庫及質體檢驗、病毒載體檢驗方法開發、病毒載體生產、產品安全性檢驗、批次放行檢驗、臨床(前)毒理試驗、臨床檢體檢驗等，提供一條龍式的服務，減少樣品運送及委外所耗費的時間成本，藉以因應急速成長的基因及細胞治療市場。

Continued ...

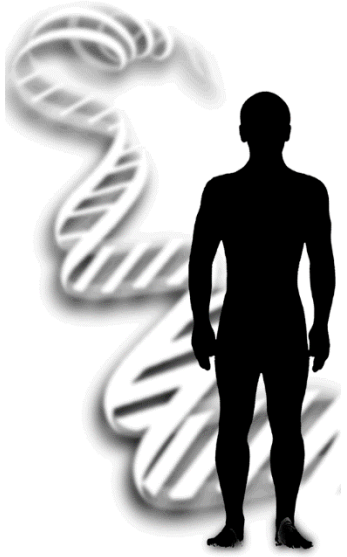
在病毒載體的品質方面，因偵測方法、偵測目標序列、宿主細胞株或製程添加物之差異，而造成病毒力價(Viral Vector Titer)或(Transduction Unit)數據略有高低，因此，於製程中，除需確保製造過程的一致性外，也需統一檢驗的偵測方法。針對單一檢驗項目的檢測方式有許多種，如在AAV的純度方面，穿透式電子顯微鏡(Transmission Electron Microscopy, TEM)等，為目前常用的檢測手法，可從外觀上區分出Empty AAV以及Full AAV。

綜合前述，病毒載體之製備從開發、臨床試驗用量乃至商業規模，依序進程為：測試階段→小量研發→大量研發→GMP先導測試→GMP製程。在各階段相對應的檢驗方法，應同時開發並進行確效，確保病毒載體可應用於臨床試驗甚至商品化。



ELSI: Ethical, Legal, and Social Issues

- **Privacy and confidentiality of genetic information.**
- **Fairness in the use of genetic information** by insurers, employers, courts, schools, adoption agencies, and the military, among others.
- **Psychological impact, stigmatization, and discrimination** due to an individual's genetic differences.
- **Reproductive issues** including adequate and informed consent and use of genetic information in reproductive decision making.
- **Clinical issues** including the education of doctors and other health-service providers, people identified with genetic conditions, and the general public about capabilities, limitations, and social risks; and implementation of standards and quality-control measures.



ELSI Issues (cont.)

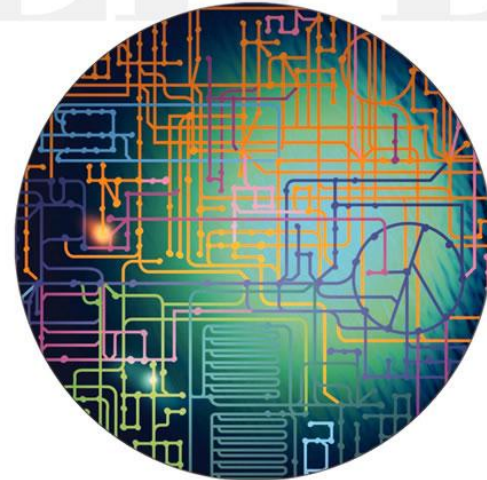
- **Uncertainties associated with gene tests for susceptibilities and complex conditions** (e.g., heart disease, diabetes, and Alzheimer's disease).
- **Fairness in access to advanced genomic technologies.**
- **Conceptual and philosophical implications** regarding human responsibility, free will vs genetic determinism, and concepts of health and disease.
- **Health and environmental issues** concerning genetically modified (GM) foods and microbes.
- **Commercialization of products** including property rights (patents, copyrights, and trade secrets) and accessibility of data and materials.

Beyond the HGP: What's Next?



HapMap

*Chart genetic variation
within the human genome*



U.S. DEPARTMENT OF ENERGY

Systems Biology

*Exploring Microbial Genomes for
Energy and the Environment*



U.S. DEPARTMENT OF ENERGY

Genomes to Life: A DOE Systems Biology Program

*Exploring Microbial Genomes for Energy and the
Environment*

Goals

- identify the protein machines that carry out critical life functions
- characterize the gene regulatory networks that control these machines
- characterize the functional repertoire of complex microbial communities in their natural environments
- develop the computational capabilities to integrate and understand these data and begin to model complex biological systems

GTL Applications in Energy Security and Global Climate Change

Within a Decade

Develop knowledge base for cost-effective cleanup strategies



Long Term

2020

Save billions of dollars in toxic waste cleanup and disposal

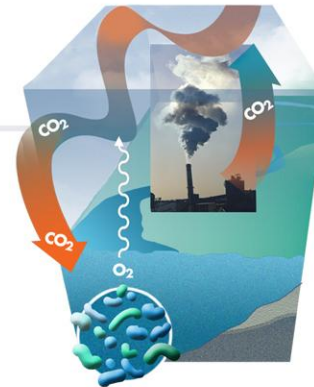


2040

Understand earth's natural carbon cycle and design strategies for enhanced carbon capture



Help stabilize atmospheric carbon dioxide to counter global warming

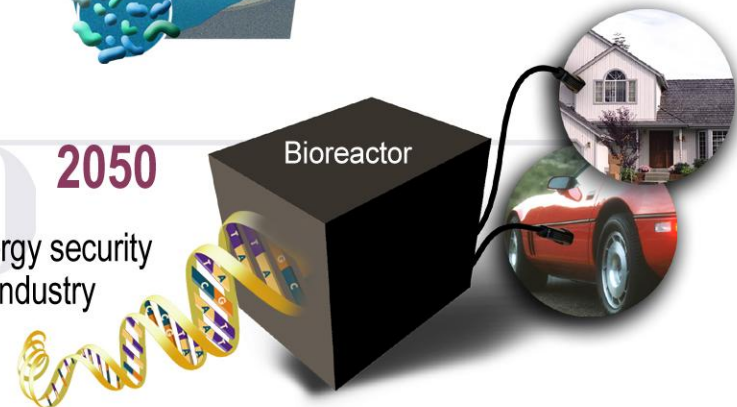


2050

Increase biological sources of fuels and electricity



Contribute to U.S. energy security
• Biohydrogen-based industry in place



HapMap

An NIH program to chart genetic variation within the human genome



- Begun in 2002, the project is a 3-year effort to construct a map of the patterns of SNPs (single nucleotide polymorphisms) that occur across populations in Africa, Asia, and the United States.
- Consortium of researchers from six countries
- Researchers hope that dramatically decreasing the number of individual SNPs to be scanned will provide a shortcut for identifying the DNA regions associated with common complex diseases
- Map may also be useful in understanding how genetic variation contributes to responses in environmental factors