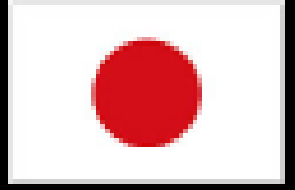




Korea-China-Japan Bioinformatics Training Course

- JST: Japan Science and Technology -

"Introduction to evolutionary and comparative genomics"

March 11, 2011
Ocean Suites Hotel, Jeju, Korea

Takashi Gojobori

Center for Information Biology and
DDBJ (DNA Data Bank of Japan)
National Institute of Genetics,
Mishima, Japan

Mission: Message

- Basis of Population genetics/genomics
- Relationship of intra-pupation genetic diversity with molecular/evolution
- Evolutionary implication of duplication
- Evolution

  -Conservation and Diversification-
- Conservation

  - Sequence level

  - Gene set level
- Evolution of the Neural system

Natural Selection
（Darwinism）

Law of Inheritance
＋
Mutation

Neo-Darwinism
（The Synthesis Theory of Evolution）

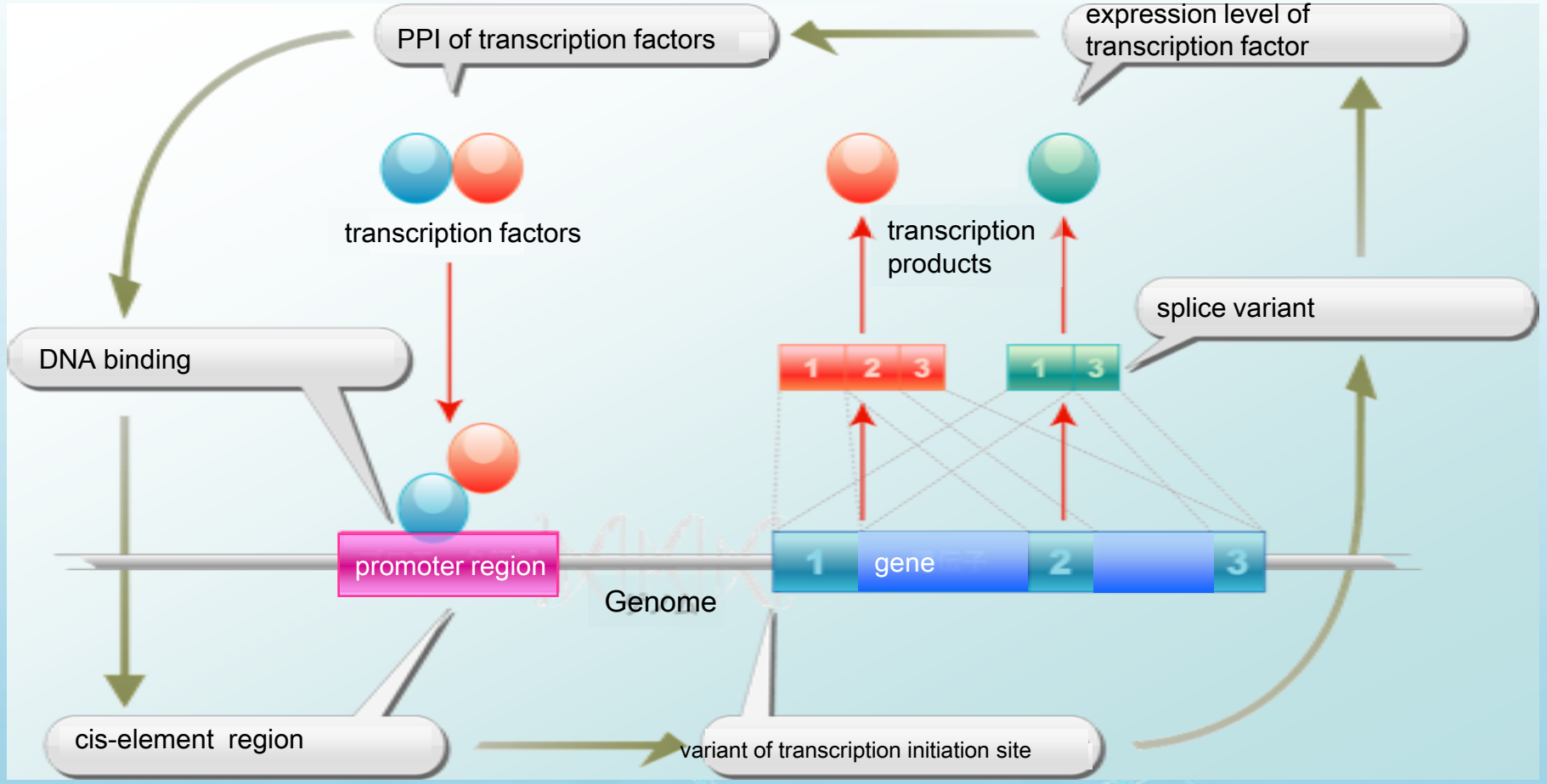
(How about the molecular level?=> Genetic polymorphism?

Genome・DNA・Genes
(All cells)

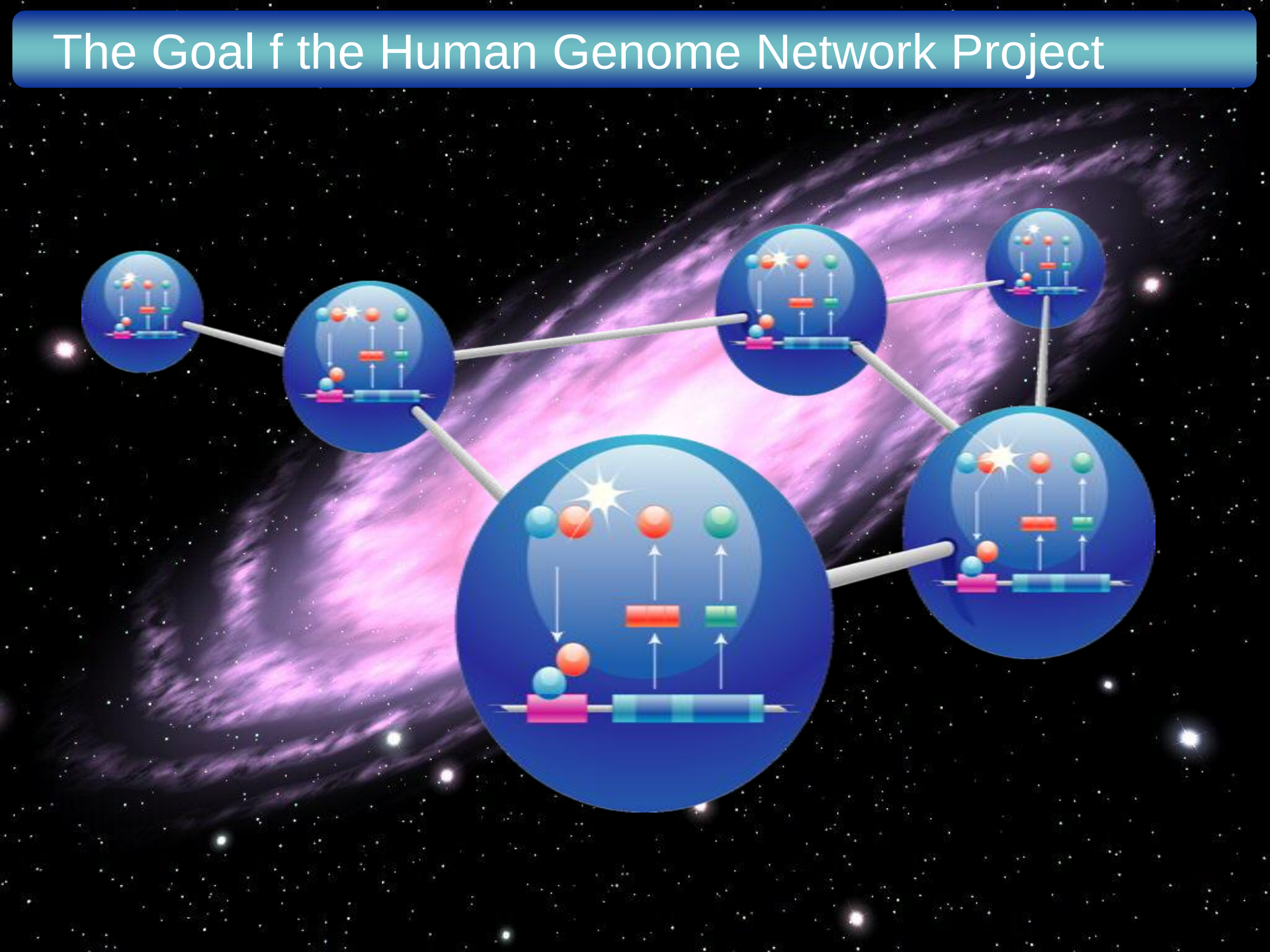
mRNA (cDNA)
（Some Cells/Tissues/Organs）

Proteins
（Some Cells/Tissues/Organs）

The Goal of Human Genome Network Project – Elucidation of an Elementary Process



The Goal f the Human Genome Network Project



Genome =

Gene + Chromosome

- Coined by Hans Winkler (1920)

- Modified in functional context by Hitoshi Kihara (1930)

Biological hierarchy

Evolution

Ecosystem (Environments)

Population (Human population)

Individual (Human)

Organ (Lung, Stomach)

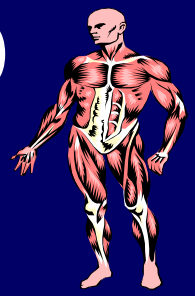
Tissue (Epidermal tissue)

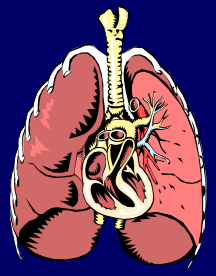
Cell (Red blood cell)

Organella (Mitochondria)

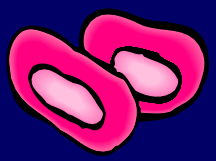
Bio-molecules (DNA, RNA, Proteins)

Molecules (H_2O, O_2)

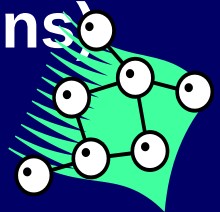




Integration



Red Blood Cell



The Underlying Ideas of Population Genetics

Key issues of relevant subjects

- Population genetics = To study the time change of genetic variability within a population or between populations in order to understand the evolutionary mechanisms.
- Polymorphism = A particular locus is called to be polymorphic when the allele frequency of the most common allele on the locus is less than 1% (or 5%).
- Linkage disequilibrium : $D = X_{12} - x_1 x_2$, where X_{12} is a haplotype frequency, and x_1 and x_2 are allele frequencies.

Major factors for changing allele frequencies

- **Mutation**
- **Random genetic drift -- random mating**
  => size effect
  => (bottleneck effect / founders' effect / Wright effect)
- **Natural selection**
  - - negative selection (purifying selection)
  - neutral
  - positive selection

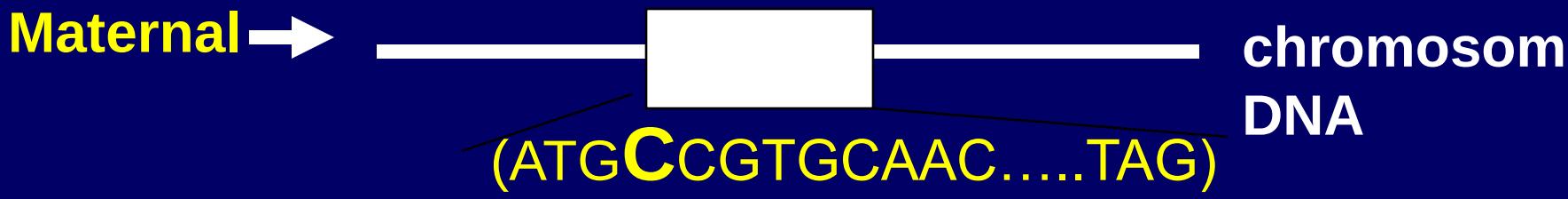
What is a gene ?

Chromosome No. 1 : The container of DNAs

For human, 23 chromosomes x 2 sets = 46 chromo

Locus, Loci : Location of a gene

allele : "A"



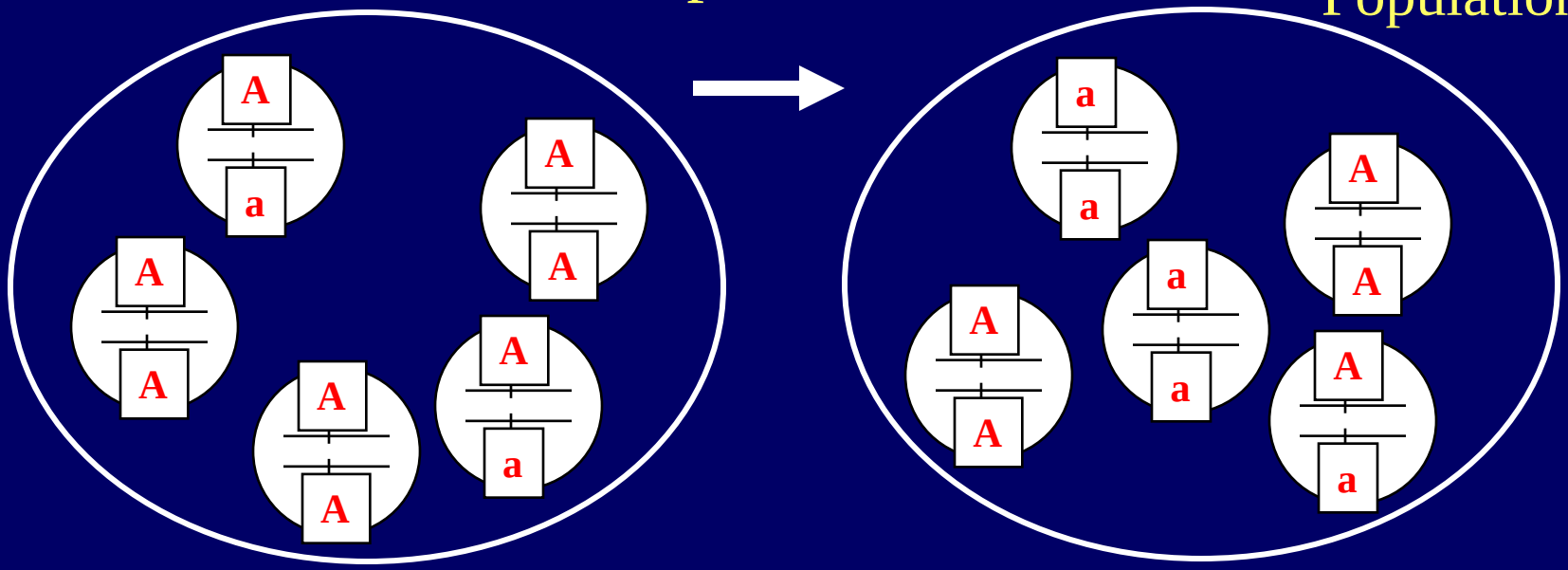
allele : "a"

Maternal and Paternal distinction (Genetic imprinting)
=> Usually impossible.

populationX

Gene pool

Population X



Frequency of allele A $x_A = \frac{8}{10} = 80\%$

Frequency a $x_a = \frac{2}{10} = 20\%$

Frequency A $x_A = \frac{5}{10} = 50\%$

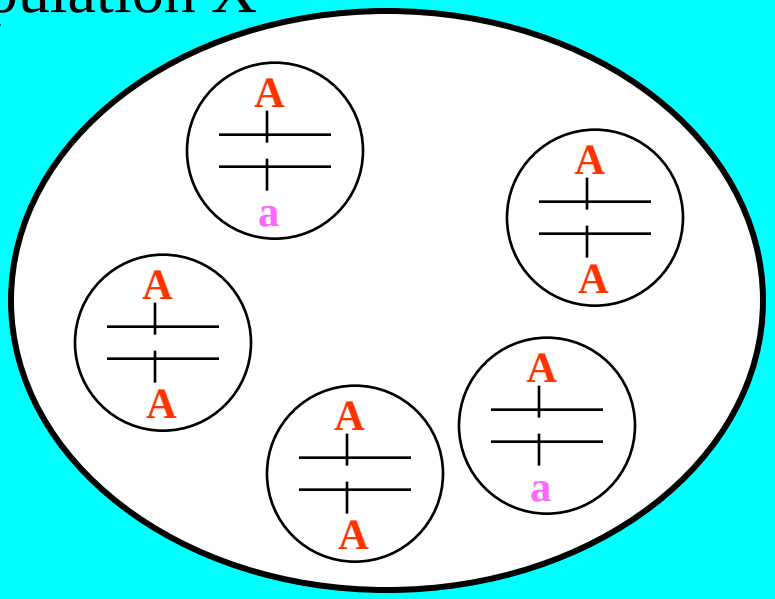
Frequency a $x_a = \frac{5}{10} = 50\%$

Genetic composition > allele frequency

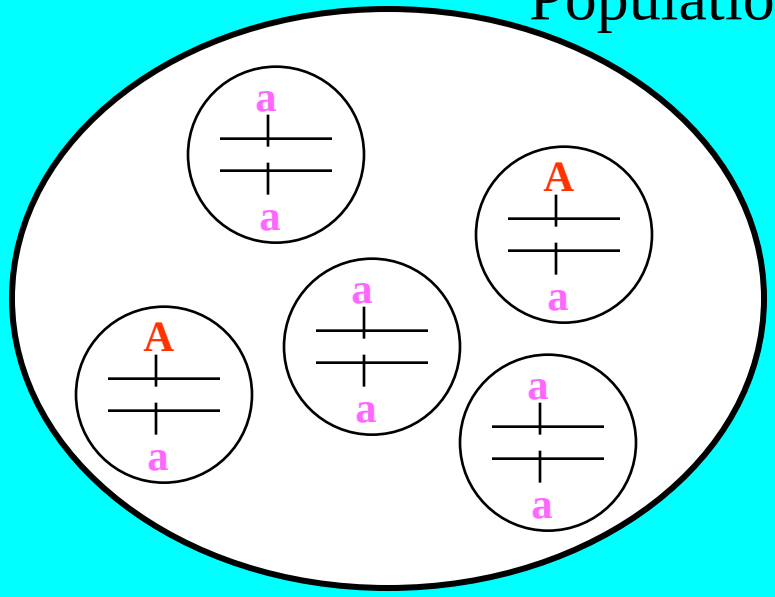
Population genetics => Time change of allele frequencies

Nei's Genetic Distance

Population X



Population Y



Frequency A $x_A = \frac{8}{10} = 80\%$

Frequency a $x_a = \frac{2}{10} = 20\%$

Frequency A $y_A = \frac{2}{10} = 20\%$

Frequency a $y_a = \frac{8}{10} = 80\%$

Genetic similarity between populations X and Y

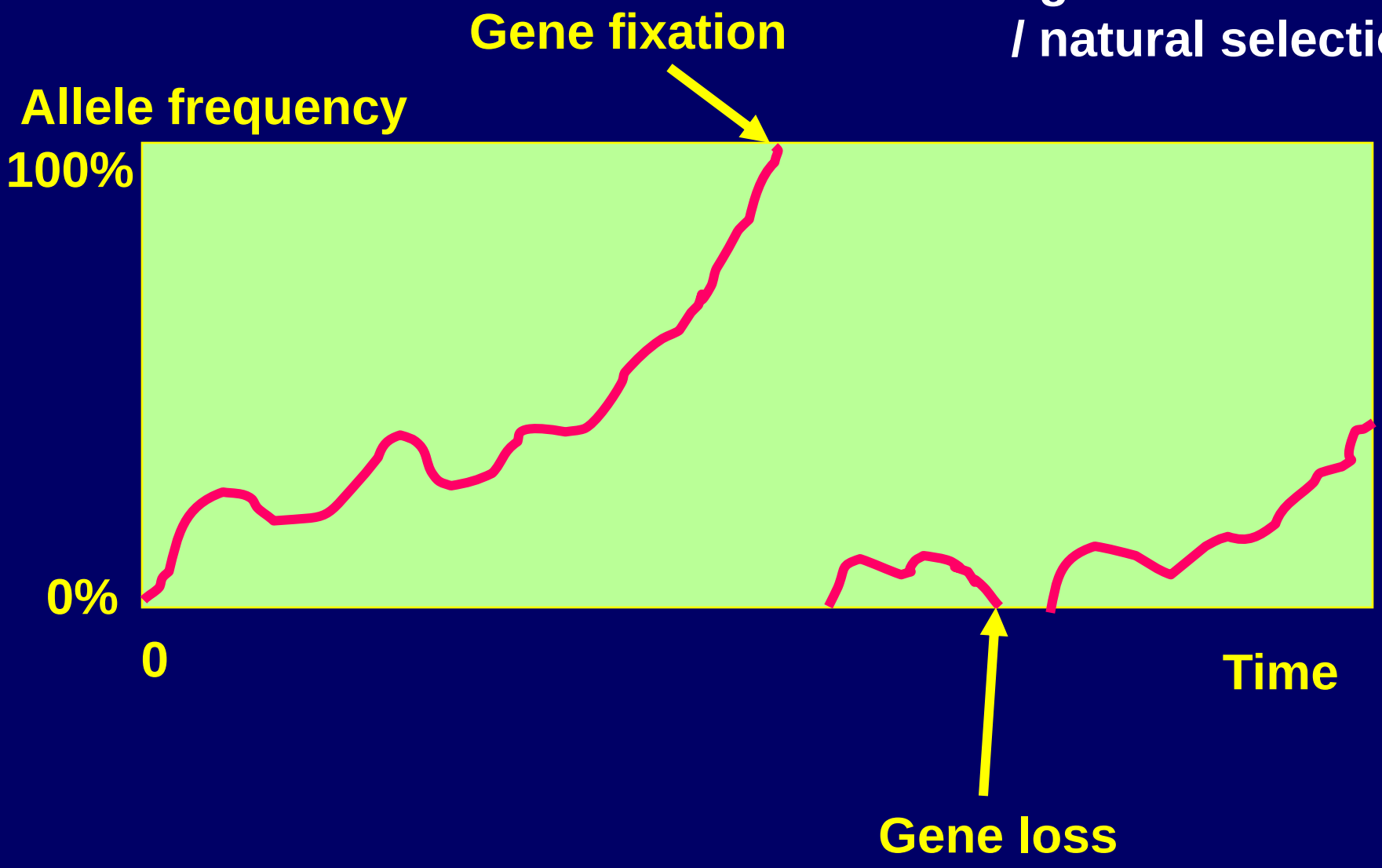
$$I = \frac{x_A y_A + x_a y_a}{\sqrt{x_A^2 + x_a^2}\sqrt{y_A^2 + y_a^2}}$$

$$\text{generally, } I = \frac{\sum x_i y_i}{\left(\sqrt{\sum x_i^2}\right)\left(\sqrt{\sum y_i^2}\right)}$$

Nei's genetic distance, $D = -\log_e I$

Time change of allele frequency

/ mutation
/ genetic drift
/ natural selection



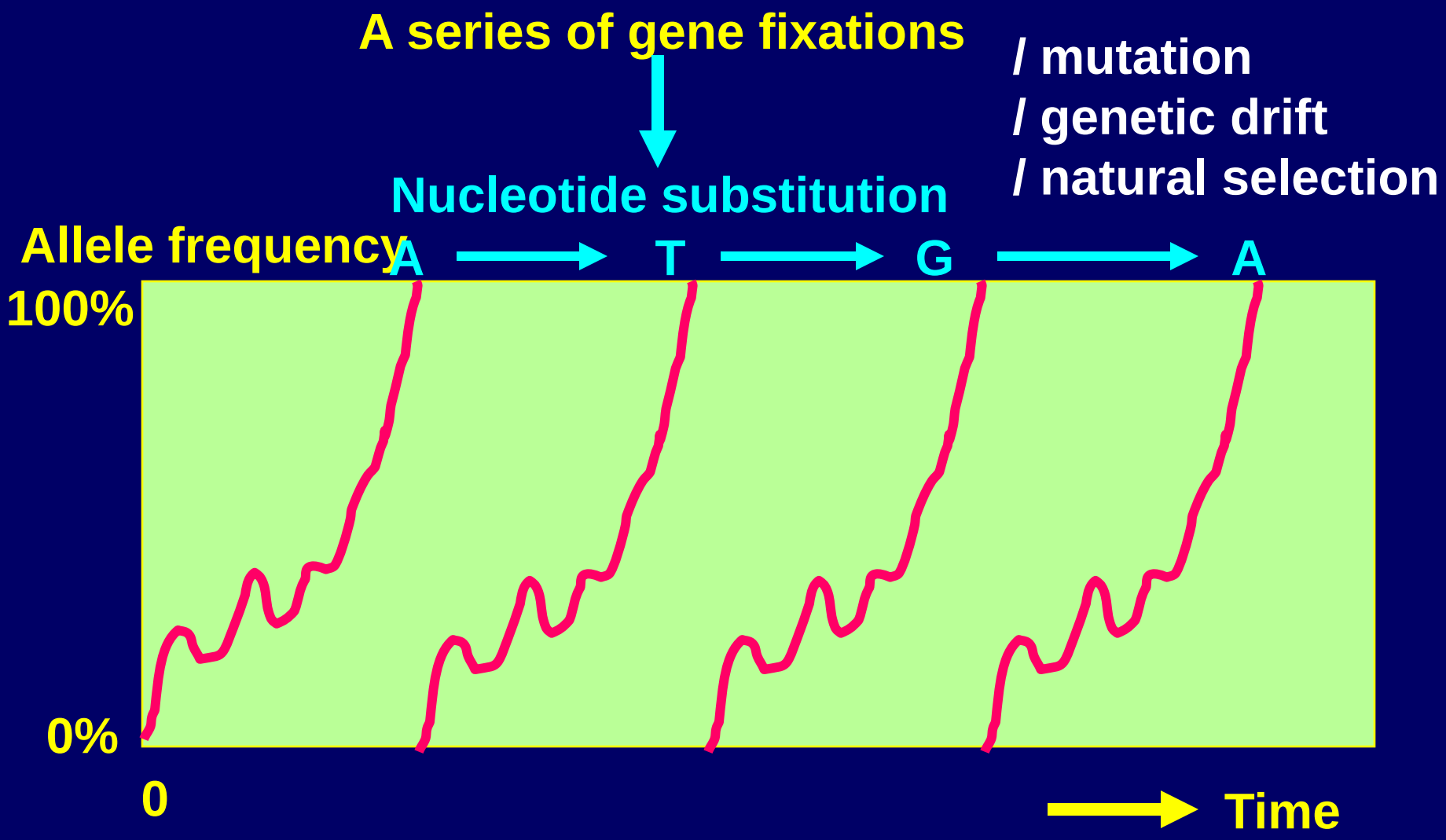
Mechanisms of frequency changes

- **Mutation**
- **Selection (positive or negative selection)**
- **Random genetic drift (random mating)**

=> Size effect

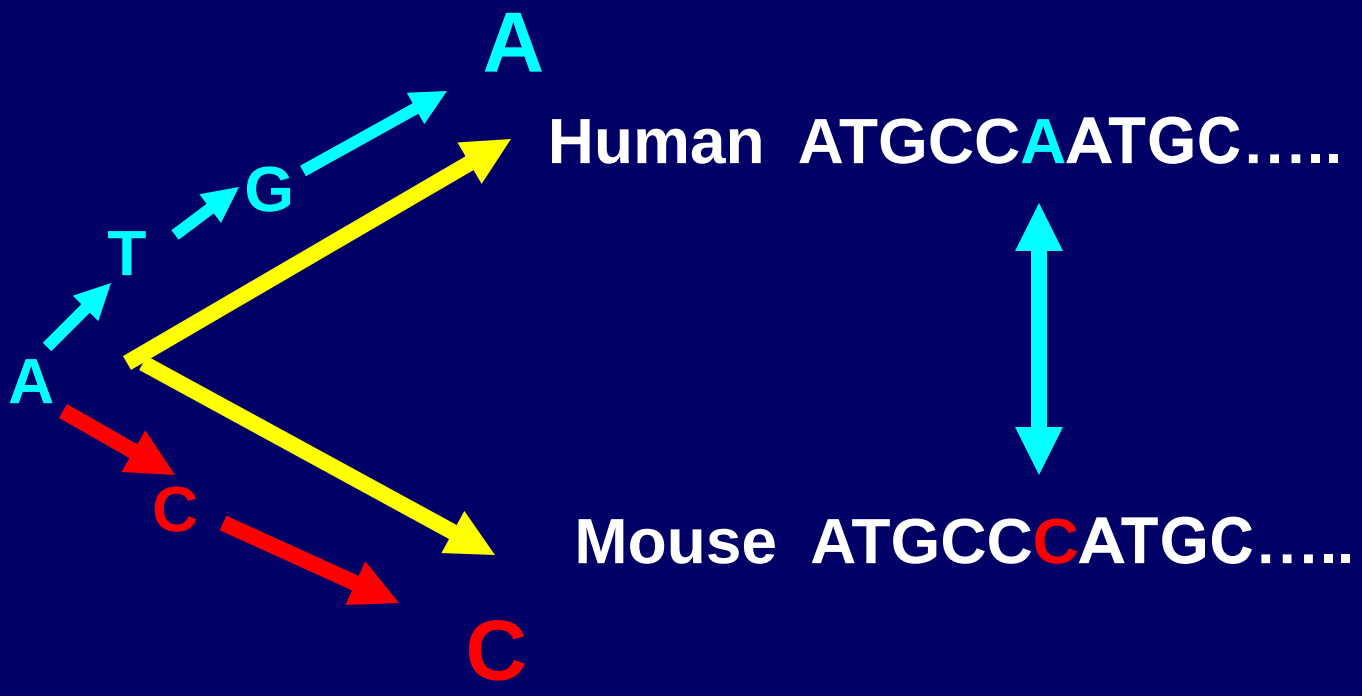
- Migration and others

Relationship with Molecular Evolution

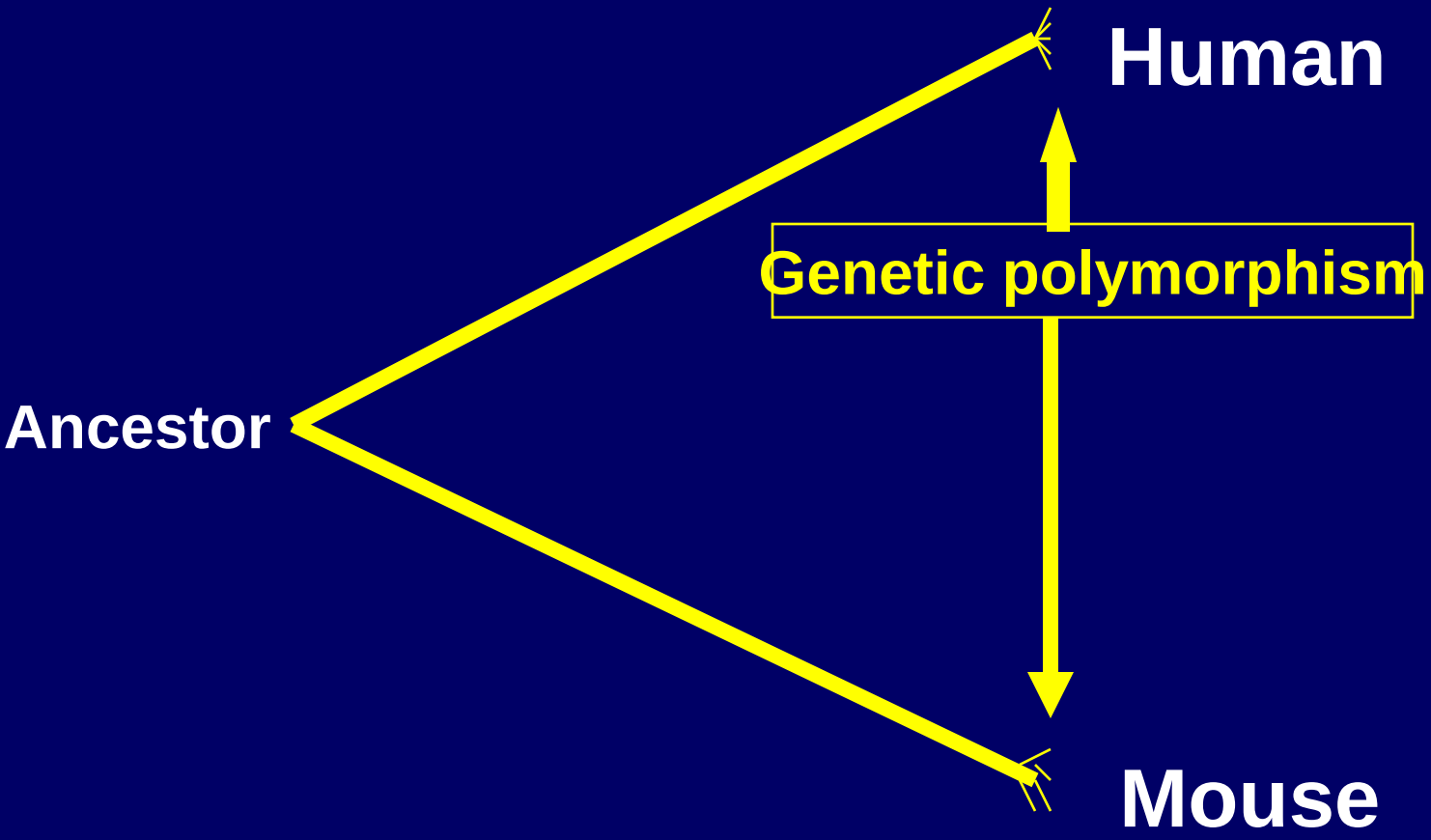


Relationship between population genetics and molecular evolution

Molecular Evolution : Nucleotide substitutions



Phylogenetic tree and genetic polymorphism



Evolutionary rate and Mutation rate

- Evolutionary rate (u)
  (a rate of nucleotide substitutions or
  a rate of amino acid substitutions or
  a rate of Indel, and so on)
- Mutation rate (v)

- u = v (neutral) (= $2Nv \cdot \{1/(2N)\}$)
  > v (positive selection, Darwinian)
  < v (negative selection, purifying selection)

Remarks

- **Nucleotide substitution as mutation**
  (Nucleotide substitution mutation)
  - a new-arisen mutation
  - a polymorphic variant (ex. SNP)
- **Nucleotide substitution through fixation process**
  (Nucleotide substitution)

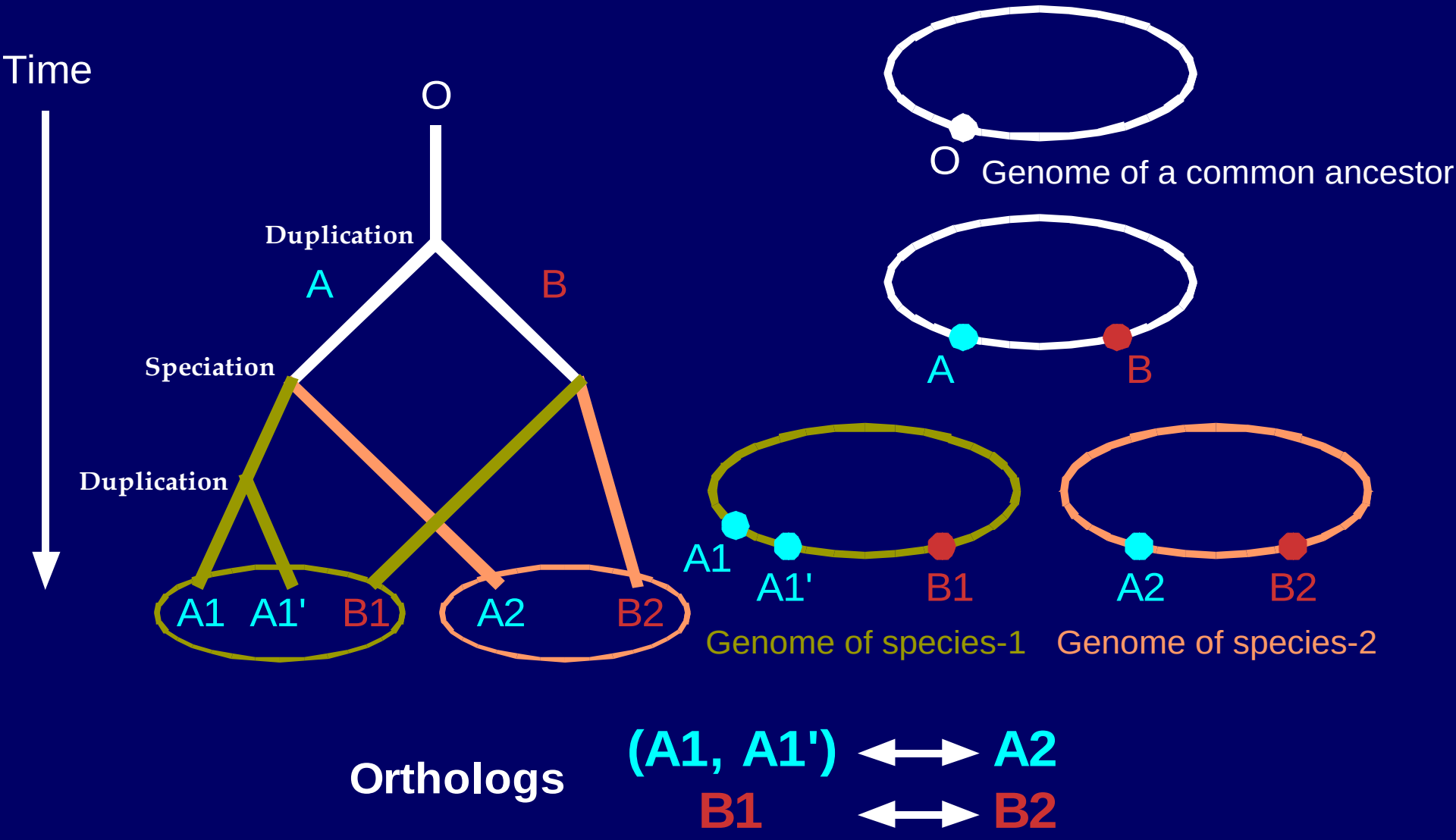
- **Synonymous substitution vs. Nonsynonymous substitution**
- **Mutation rate (pseudogene, non-codingr egion)**
- **Degree of selective constraints**
  (selective coefficient)

Classification of gene pairs based on their evolutionary history

- Orthologues
  — **Gene pairs that have diverged along with speciation.**

- Paralogues
  — **Gene pairs that emerged from a gene duplication event**
- Xenologue
  — **A gene that emerged from horizontal gene transfer beyond species barrier**

Detection of orthologous pairs

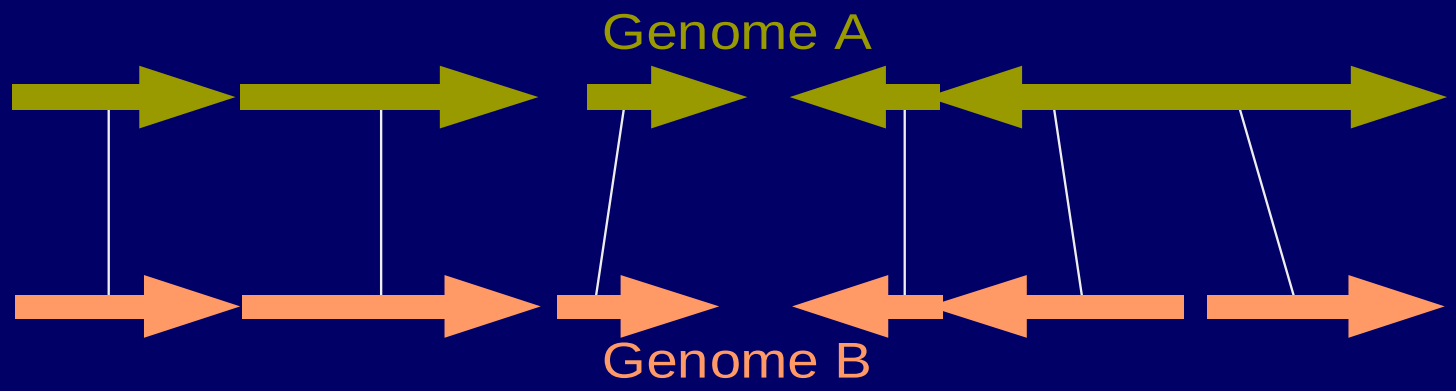


Orthologs **(A1, A1') ⟷ A2**

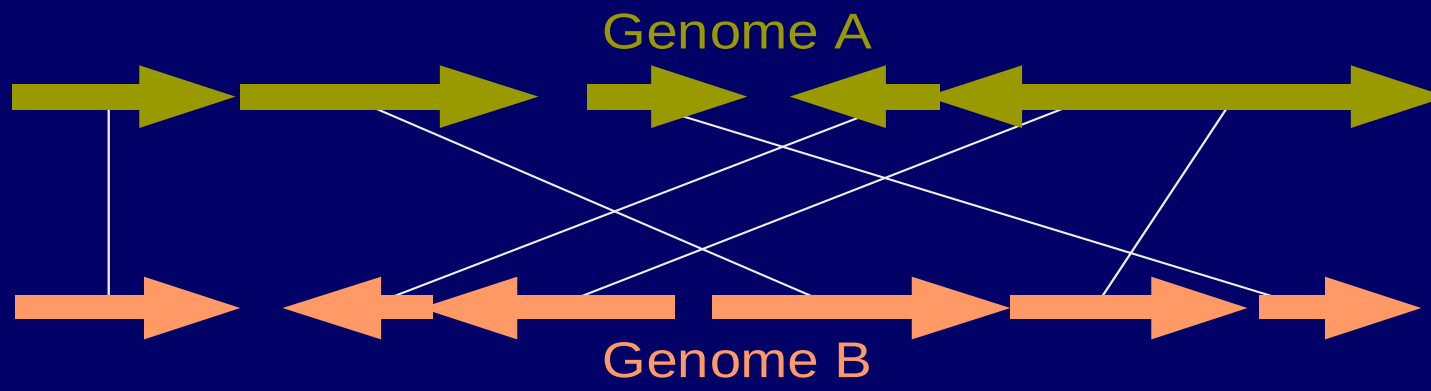
B1 ⟷ B2

Graphical representation of structural changes between two genomes

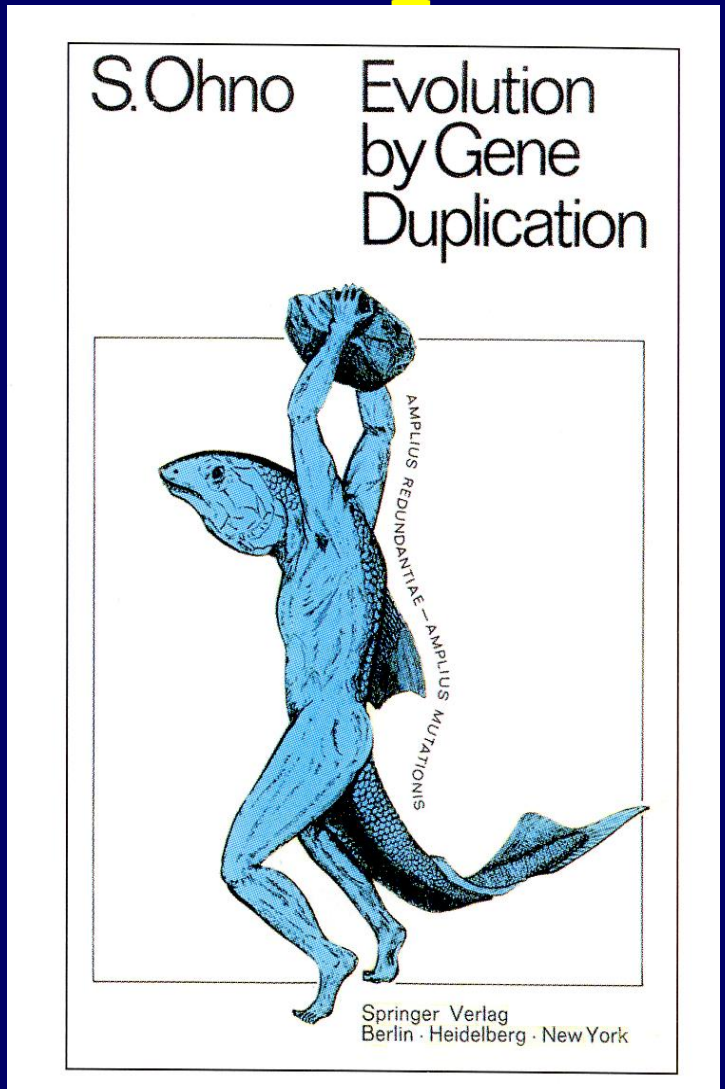
(a) Conserved genome structure



(b) Shuffled genome structure



Evolution by Gene Duplication

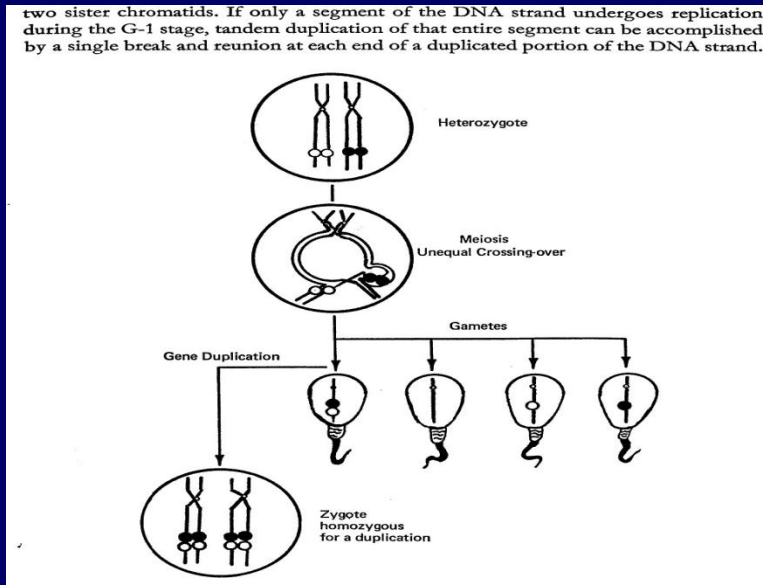


(1970)

Gene and Genome Duplication

- Major role of gene duplication in evolution
- Duplicates preserved by ***neo-functionalization***
- 2R hypothesis: vertebrates went through at least one (probably 2-round) whole genome duplications

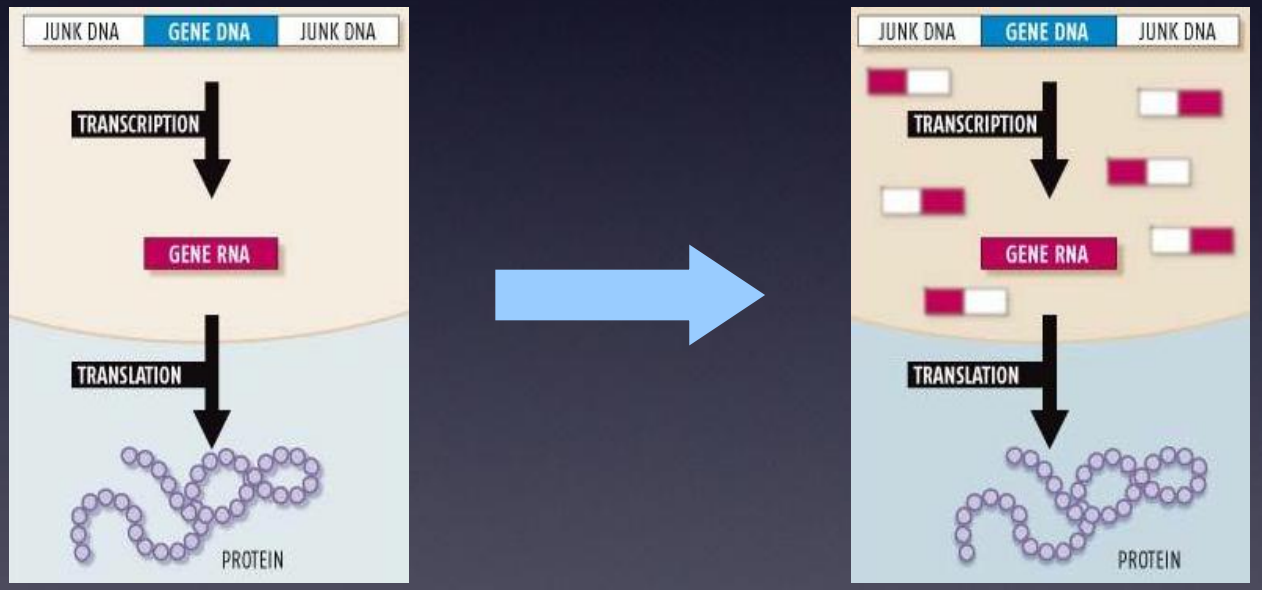
- ***Neo-functionalization***
- ***sub-functionalization***
- ***Non-functionalization***



Junk DNA

- **Term coined in the article "So much 'junk DNA' in our genome"**

(Brookhaven Symposium on Biology, 1972)



Population genetics is the basis of molecular or genome evolution

- Disease gene hunting : Polymorphic markers of a human population.
- Utilization of polymorphisms of other organisms => mouse SNP, rat SNP, macaque SNP, and so on.
- Utilization of comparative genomics => to undersatnd the process of functional diversification of genes and gene network.
- From forward and reverse genetics to "inverse genetics" (transduction of genetic elements of other species. (S. Brenner)

Evolution

- Explanation of the seemingly contradictory feature “Conservation and Diversification”
- Conservation =>
  - Sequence level
    (non-coding, gene, genome)
  - Molecular structure level
  - Gene set level
  - Network level (interaction)
  - Mechanism level

=> Functional Implications

Nature Genetics (2004) 36(7) 760-766

Horizontal gene transfer disclosed by Bayesian inference in prokaryotic genomes

Yoji Nakamura[1], Takeshi Itoh[2], Hideo Matsuda[3] and Takashi Gojobori[1,2]

[1] *Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan*

[2] *Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, TIME24 Bldg. 10F, 2-45 Aomi, Koto-ku, Tokyo 135-0064, Japan*

[3] *Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan*

Highly Conserved Upstream Sequences for Transcription Factor Genes and its Evolutionary Implication to Regulatory Network

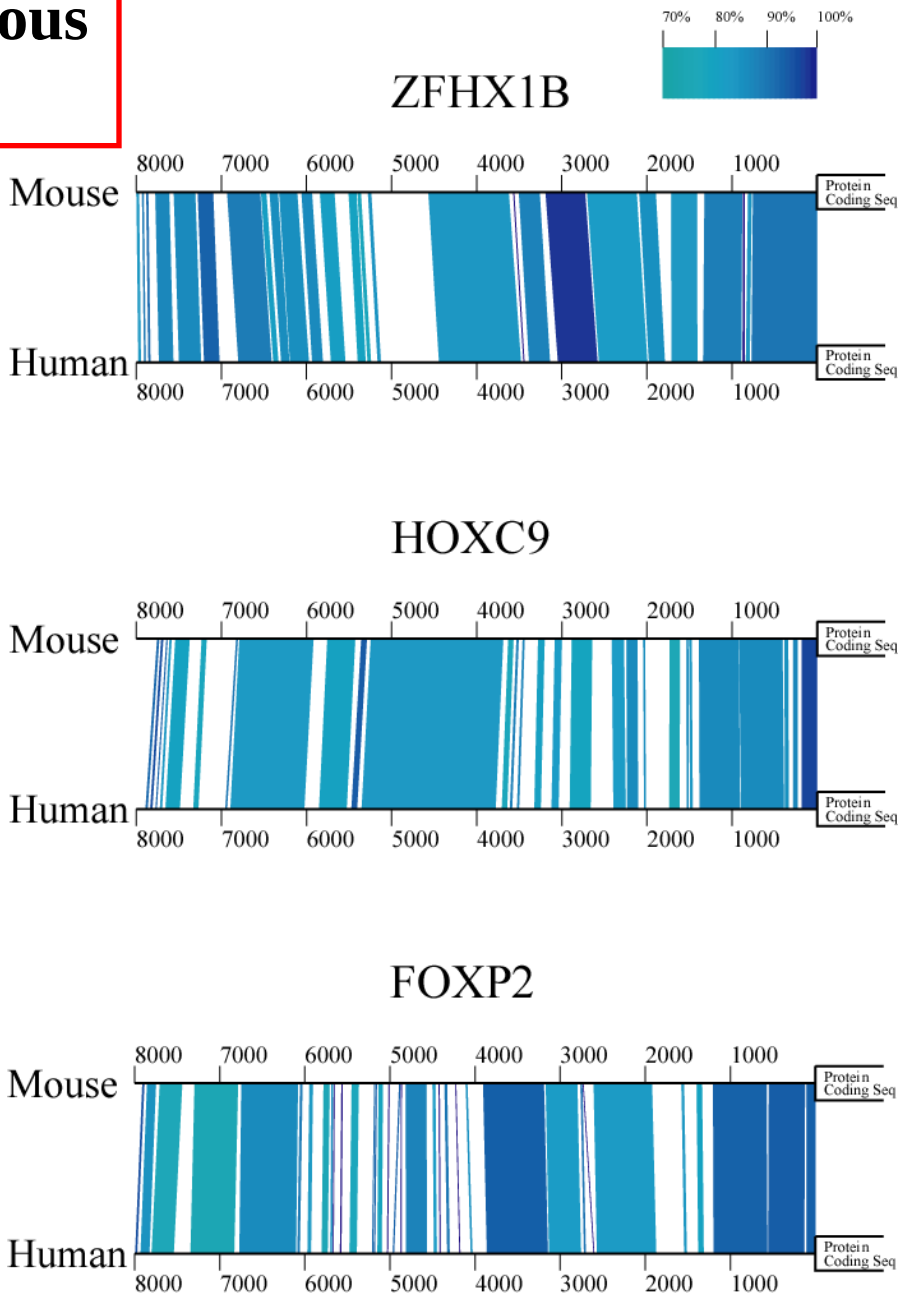
H. Iwama and T. Gojobori

***Proc. Natl. Acad. Sci.* (2004) 101:17156-61**

Top-3 Upstream-Conserved Human-Mouse Orthologous Genes

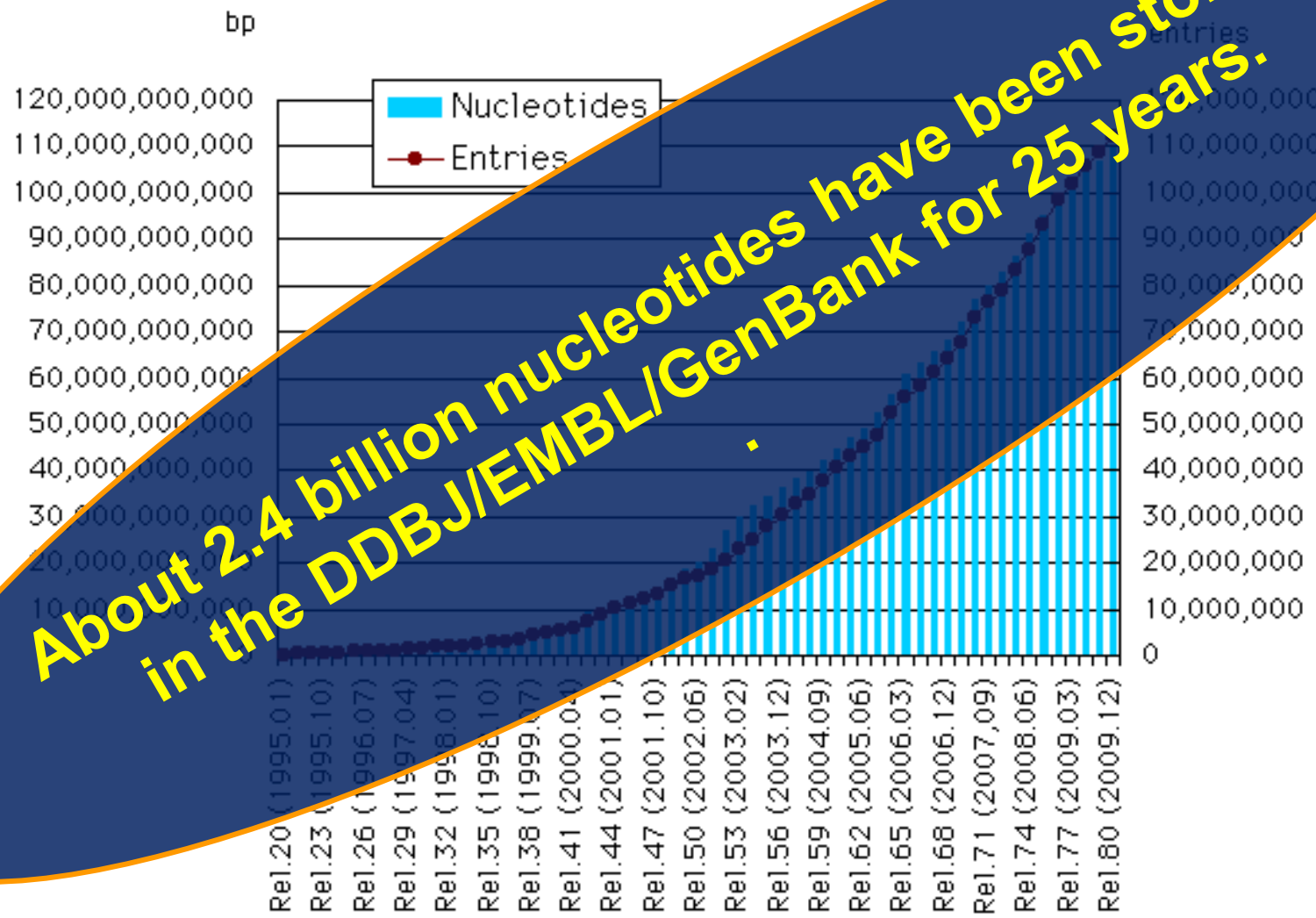
Within the top-10 upstream-conserved genes, 9 genes were ***transcription factor*** genes. ($p < 2*10^{-8}$)

62 genes of the top-200 upstream-conserved genes were also ***transcription factor*** genes. ($p < 5*10^{-15}$)



Contents

DDBJ/EMBL/GenBank database growth



About 2.4 billion nucleotides have been stored in the DDBJ/EMBL/GenBank for 25 years.

* Note : CON and TPA divisions are not counted in the Release

Sciencexpress Report

Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays

Radoje Drmanac,1* Andrew B. Sparks,1† Matthew J. Callow,1† Aaron L. Halpern,1† Norma L. Burns,1†,Bahram G. Kermani,1† Paolo Carnevali,1† Igor Nazarenko,1† Geoffrey B. Nilsen,1† George Yeung,1† Fredrik Dahl,1† Andres Fernandez,1† Bryan Staker,1† Krishna P. Pant,1† Jonathan Baccash,1 Adam P. Borcherding,1 Anushka Brownley,1 Ryan Cedeno,1 Linsu Chen,1 Dan Chernikoff,1 Alex Cheung,1 Razvan Chirita,1 Benjamin Curson,1 Jessica C. Ebert,1 Coleen R. Hacker,1 Robert Hartlage,1 Brian Hauser,1 Steve Huang,1
Yuan Jiang,1 Vitali Karpinchyk,1 Mark Koenig,1 Calvin Kong,1 Tom Lanman,1 Catherine Le,1 Jia Liu,1 Celeste E. McBride,1 Matt Morenzoni,1 Robert E. Morey,1 § Karl Mutch,1 Helena Perazich,1 Kerry Perry,1 Brock A. Peters,1 Joe Peterson,1 Charit L. Pethiyagoda,1 Kaliprasad Pothuraju,1 Claudia Richter,1 Abraham M. Rosenbaum,2 Shaunak Roy,1 Jay Shafto,1 Uladzislau Sharanhovich,1 Karen W. Shannon,1|| Conrad G. Sheppy,1 Michel Sun,1 Joseph V. Thakuria,2 Anne Tran,1 Dylan Vu,1 Alexander Wait Zaranek,2 Xiaodi Wu,3 Snezana Drmanac,1 Arnold R. Oliphant,1William C. Banyai,1 Bruce Martin,1 Dennis G. Ballinger,1* George M. Church,2 Clifford A. Reid1
1Complete Genomics, Inc., 2071 Stierlin Court, Mountain View, CA 94043, USA. 2Department of Genetics, Harvard Medical School, Cambridge, MA, USA. 3School of Medicine, Washington University, St. Louis, St. Louis, MO, USA.

Genome sequencing of large numbers of individuals promises to advance the understanding, treatment, and prevention of human diseases, among other applications. We describe a genome sequencing platform that achieves efficient imaging and low reagent consumption with combinatorial probe anchor ligation (cPAL) chemistry to independently assay each base from patterned nanoarrays of self-assembling DNA nanoballs (DNBs). We sequenced three human genomes with this platform, generating an average of 45- to 87-fold coverage per genome and identifying 3.2 to 4.5 million sequence variants per genome. Validation of one genome data set demonstrates a sequence accuracy of about 1 false variant per 100 kilobases. The high accuracy, affordable cost of $4,400 for sequencing consumables and scalability of this platform enable complete human genome sequencing for the detection of rare variants in large-scale genetic studies.

A human genome (45X) costs you only $4,400 !!

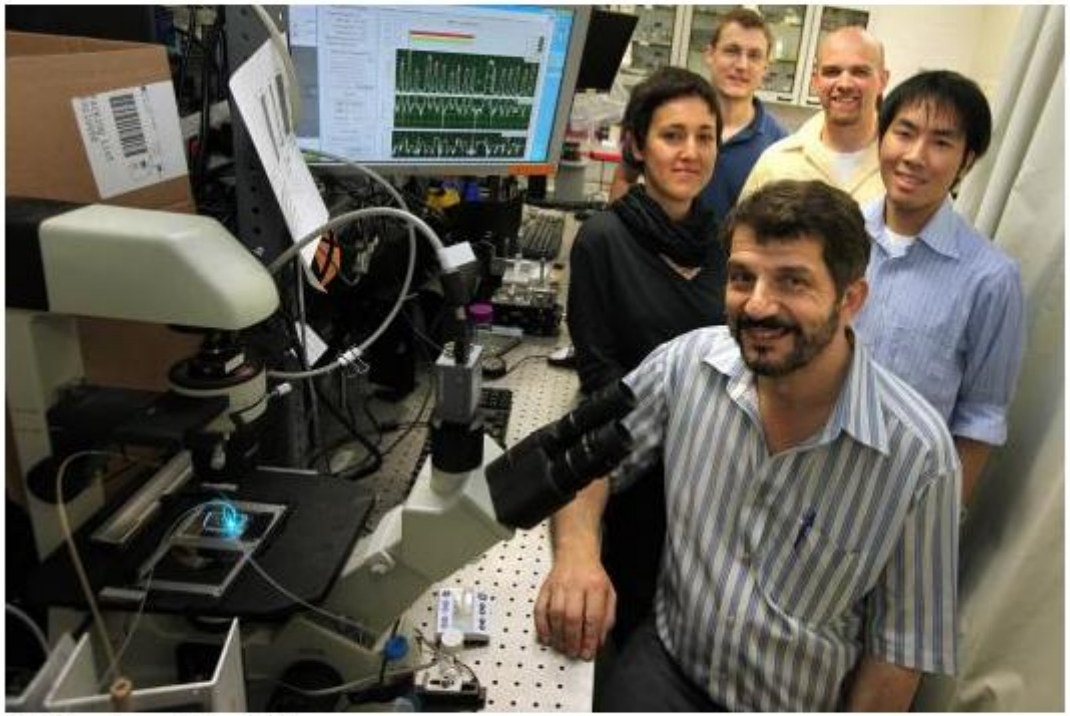
HiSeq **2000** system

www.illuminakk.co.jp/

HiSeq 2000 is the first commercially available sequencer to enable to obtain ~30x coverage of two human genomes in a single run for under $10,000 (USD)* per sample. HiSeq 2000 makes it possible f Individual labs to take on the largest and most complex sequencing t the lowest cost. The ability to process larger numbers of samples and to decode larger and more complex genomes means that virtually any sequencing project is now within reach.

The Boston Globe

Start-up aims to sequence human genomes for $30, in just a few hours



(Pat Greenhouse/Globe Staff)

By Carolyn Y. Johnson
Globe Staff / June 7, 2010

CAMBRIDGE — The race to sequence genomes faster and cheaper has a new entrant — a start-up spun out of a Harvard University laboratory that focuses on emulsions, or mixtures of liquids like those found in mayonnaise and salad dressings. Deciphering the first human genome, a massive technical feat, took more than a decade and cost about $3 billion, but the price and time have been dropping rapidly in the 10 years since — down to about $20,000, powered by new technologies that take days or weeks.

The new company, GnuBio, is in the very early stages of its development, but it said last week that its technology could sequence a human genome in hours and for **just $30.**

http://www.boston.com/business/healthcare/articles/2010/06/07/start_up_aims_to_sequence_human_genomes_for_30_in_just_a_few_hours/

Science (2009) 326:1541-1545 (11 December)

Science 11 December 2009:
Vol. 326. no. 5959, pp. 1541 - 1545
DOI: 10.1126/science.1177074

REPORTS

Mapping Human Genetic Diversity in Asia

The HUGO Pan-Asian SNP Consortium*,†

Asia harbors substantial cultural and linguistic diversity, but the geographic structure of genetic variation across the continent remains enigmatic. Here we report a large-scale survey of autosomal variation from a broad geographic sample of Asian human populations. Our results show that genetic ancestry is strongly correlated with linguistic affiliations as well as geography. Most populations show relatedness within ethnic/linguistic groups, despite prevalent gene flow among populations. More than 90% of East Asian (EA) haplotypes could be found in either Southeast Asian (SEA) or Central-South Asian (CSA) populations and show clinal structure with haplotype diversity decreasing from south to north. Furthermore, 50% of EA haplotypes were found in SEA only and 5% were found in CSA only, indicating that SEA was a major geographic source of EA populations.

The HUGO Pan-Asian SNP Consortium

Mahmood Ameen Abdulla,[1] Ikhlak Ahmed,[2] Anunchai Assawamakin,[3,4] Jong Bhak,[5] Samir K. Brahmachari,[2] Gayvelline C. Calacal,[6] Amit Chaurasia,[2] Chien-Hsiun Chen,[7] Jieming Chen,[8] Yuan-Tsong Chen,[7] Jiayou Chu,[9] Eva Maria C. Cutiongco-de la Paz,[10] Maria Corazon A. De Ungria,[6] Frederick C. Delfin,[6] Juli Edo,[1] Suthat Fuchareon,[3] Ho Ghang,[5] Takashi Gojobori,[11,12] Junsong Han,[13] Sheng-Feng Ho,[7] Boon Peng Hoh,[14] Wei Huang,[15] Hidetoshi Inoko,[16] Pankaj Jha,[2] Timothy A. Jinam,[1] Li Jin,[17,38]† Jongsun Jung,[18] Daoroong Kangwanpong,[19] Jatupol Kampuansai,[19] Giulia C. Kennedy,[20,21] Preeti Khurana,[22] Hyung-Lae Kim,[18] Kwangjoong Kim,[18] Sangsoo Kim,[23] Woo-Yeon Kim,[5] Kuchan Kimm,[24] Ryosuke Kimura,[25] Tomohiro Koike,[11] Supasak Kulawonganunchai,[4] Vikrant Kumar,[8] Poh San Lai,[26,27] Jong-Young Lee,[18] Sunghoon Lee,[5] Edison T. Liu,[8]† Partha P. Majumder,[28] Kiran Kumar Mandapati,[22] Sangkot Marzuki,[29] Wayne Mitchell,[30,31] Mitali Mukerji,[2] Kenji Naritomi,[32] Chumpol Ngamphiw,[4] Norio Niikawa,[40] Nao Nishida,[25] Bermseok Oh,[18] Sangho Oh,[5] Jun Ohashi,[25] Akira Oka,[16] Rick Ong,[8] Carmencita D. Padilla,[10] Prasit Palittapongarnpim,[33] Henry B. Perdigon,[6] Maude Elvira Phipps,[1,34] Eileen Png,[8] Yoshiyuki Sakaki,[35] Jazelyn M. Salvador,[6] Yuliana Sandraling,[29] Vinod Scaria,[2] Mark Seielstad,[8]† Mohd Ros Sidek,[14] Amit Sinha,[2] Metawee Srikummool,[19] Herawati Sudoyo,[29] Sumio Sugano,[37] Helena Suryadi,[29] Yoshiyuki Suzuki,[11] Kristina A. Tabbada,[6] Adrian Tan,[8] Katsushi Tokunaga,[25] Sissades Tongsima,[4] Lilian P. Villamor,[6] Eric Wang,[20,21] Ying Wang,[15] Haifeng Wang,[15] Jer-Yuarn Wu,[7] Huasheng Xiao,[13] Shuhua Xu,[38]† Jin Ok Yang,[5] Yin Yao Shugart,[39] Hyang-Sook Yoo,[5] Wentao Yuan,[15] Guoping Zhao,[15] Bin Alwi Zilfalil,[14] Indian Genome Variation Consortium[2]

NATURE ARTICLES

Complex landscapes of somatic rearrangement in human breast cancer genomes

Philip J. Stephens1, David J. McBride1, Meng-Lay Lin1, Ignacio Varela1, Erin D. Pleasance1, Jared T. Simpson1, Lucy A. Stebbings1, Catherine Leroy1, Sarah Edkins1, Laura J. Mudie1, Chris D. Greenman1, Mingming Jia1, Calli Latimer1, Jon W. Teague1, King Wai Lau1, John Burton1, Michael A. Quail1, Harold Swerdlow1, Carol Churcher1, Rachael Natrajan2, Anieta M. Sieuwerts3, John W. M. Martens3, Daniel P. Silver4, Anita Langerød5, Hege E. G. Russnes5, John A. Foekens3, Jorge S. Reis-Filho2, Laura van 't Veer6, Andrea L. Richardson4,7, Anne-Lise Børresen-Dale5,8, Peter J. Campbell1, P. Andrew Futreal1 & Michael R. Stratton1,9

Multiple somatic rearrangements are often found in cancer genomes; however, the underlying processes of rearrangement and their contribution to cancer development are poorly characterized. Here we use a paired-end sequencing strategy to identify somatic rearrangements in breast cancer genomes. There are more rearrangements in some breast cancers than previously appreciated. Rearrangements are more frequent over gene footprints and most are intrachromosomal. Multiple rearrangement architectures are present, but tandem duplications are particularly common in some cancers, perhaps reflecting a specific defect in DNA maintenance. Short overlapping sequences at most rearrangement junctions indicate that these have been mediated by non-homologous end-joining DNA repair, although varying sequence patterns indicate that multiple processes of this type are operative. Several expressed in-frame fusion genes were identified but none was recurrent. The study provides a new perspective on cancer genomes, highlighting the diversity of somatic rearrangements and their potential contribution to cancer development.

Nature (2009) 462: 1005-1012 31 December, 2009

Vol 463|18 February 2010|doi:10.1038/nature08768 nature

ARTICLES

Signatures of mutation and selection in the cancer genome

Graham R. Bignell[1]*, Chris D. Greenman[1]*, Helen Davies[1], Adam P. Butler[1], Sarah Edkins[1], Jenny M. Andrews[1], Gemma Buck[1], Lina Chen[1], David Beare[1], Calli Latimer[1], Sara Widaa[1], Jonathon Hinton[1], Ciara Fahey[1], Beiyuan Fu[1], Sajani Swamy[1], Gillian L. Dalgliesh[1], Bin T. Teh[2], Panos Deloukas[1], Fengtang Yang[1], Peter J. Campbell[1], P. Andrew Futreal[1] & Michael R. Stratton[1,3]

The cancer genome is moulded by the dual processes of somatic mutation and selection. Homozygous deletions in cancer genomes occur over recessive cancer genes, where they can confer selective growth advantage, and over fragile sites, where they are thought to reflect an increased local rate of DNA breakage. However, most homozygous deletions in cancer genomes are unexplained. Here we identified 2,428 somatic homozygous deletions in 746 cancer cell lines. These overlie 11% of protein-coding genes that, therefore, are not mandatory for survival of human cells. We derived structural signatures that distinguish between homozygous deletions over recessive cancer genes and fragile sites. Application to clusters of unexplained homozygous deletions suggests that many are in regions of inherent fragility, whereas a small subset overlies recessive cancer genes. The results illustrate how structural signatures can be used to distinguish between the influences of mutation and selection in cancer genomes. The extensive copy number, genotyping, sequence and expression data available for this large series of publicly available cancer cell lines renders them informative reagents for future studies of cancer biology and drug discovery.

(18 February, 2010)

Vol 463|18 February 2010|doi:10.1038/nature08822 nature

ARTICLES

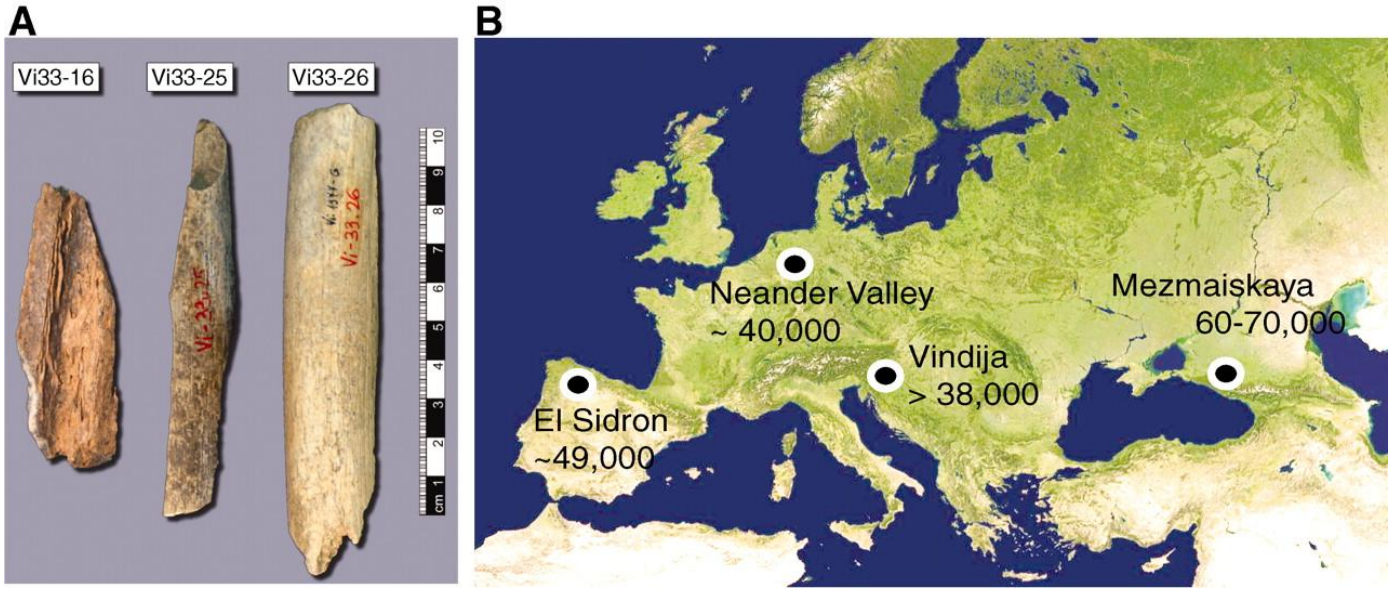
The landscape of somatic copy-number alteration across human cancers

Rameen Beroukhim[1,3,4,5]*, Craig H. Mermel[1,3]*, Dale Porter[8], Guo Wei[1], Soumya Raychaudhuri[1,4], Jerry Donovan[8], Jordi Barretina[1,3], Jesse S. Boehm[1], Jennifer Dobson[1,3], Mitsuyoshi Urashima[9], Kevin T. Mc Henry[8], Reid M. Pinchback[1], Azra H. Ligon[4], Yoon-Jae Cho[6], Leila Haery[1,3], Heidi Greulich[1,3,4,5], Michael Reich[1], Wendy Winckler[1], Michael S. Lawrence[1], Barbara A. Weir[1,3], Kumiko E. Tanaka[1,3], Derek Y. Chiang[1,3,13], Adam J. Bass[1,3,4], Alice Loo[8], Carter Hoffman[1,3], John Prensner[1,3], Ted Liefeld[1], Qing Gao[1], Derek Yecies[3], Sabina Signoretti[3,4], Elizabeth Maher[10], Frederic J. Kaye[11], Hidefumi Sasaki[12], Joel E. Tepper[13], Jonathan A. Fletcher[4], Josep Tabernero[14], José Baselga[14], Ming-Sound Tsao[15], Francesca Demichelis[16], Mark A. Rubin[16], Pasi A. Janne[3,4], Mark J. Daly[1,17], Carmelo Nucera[7], Ross L. Levine[18], Benjamin L. Ebert[1,4,5], Stacey Gabriel[1], Anil K. Rustgi[19], Cristina R. Antonescu[18], Marc Ladanyi[18], Anthony Letai[3], Levi A. Garraway[1,3], Massimo Loda[3,4], David G. Beer[20], Lawrence D. True[21], Aikou Okamoto[22], Scott L. Pomeroy[6], Samuel Singer[18], Todd R. Golub[1,3,23], Eric S. Lander[1,2,5], Gad Getz[1], William R. Sellers[8] & Matthew Meyerson[1,3,5]

A powerful way to discover key genes with causal roles in oncogenesis is to identify genomic regions that undergo frequent alteration in human cancers. Here we present high-resolution analyses of somatic copy-number alterations (SCNAs) from 3,131 cancer specimens, belonging largely to 26 histological types. We identify 158 regions of focal SCNA that are altered at significant frequency across several cancer types, of which 122 cannot be explained by the presence of a known cancer target gene located within these regions. Several gene families are enriched among these regions of focal SCNA, including the *BCL2* family of apoptosis regulators and the NF-κB pathway. We show that cancer cells containing amplifications surrounding the *MCL1* and *BCL2L1* anti-apoptotic genes depend on the expression of these genes for survival. Finally, we demonstrate that a large majority of SCNAs identified in individual cancer types are present in several cancer types.

(18 February, 2010)

Fig. 1 Samples and sites from which DNA was retrieved



R. E. Green et al., Science 328, 710-722 (2010)

Published by AAAS

Our recent activities of the related subjects

- Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, Ikuta K, Jern P, **Gojobori T**, Coffin JM, Tomonaga K. Endogenous non-retroviral RNA virus elements in mammalian genomes Identification of endogenous non-retroviral RNA virus elements in mammalian genomes ***Nature (2010)*** 463(7277): 84-87.
- Hwang, JS., Takaku, Y., Momose, T., Adamczyk, P., Özbek, S., Ikeo, K., Khalturin, K., Hemmrich, G., Bosch, T., Holstein, T., David, C., and **Gojobori, T**. (2010). Nematogalectin, a nematocyst protein with GlyXY and Galectin domains, demonstrates nematocyte-specific alternative splicing in Hydra. ***Proc. Natl. Acad. Sci. USA (20101)*** *Epub.* in October.

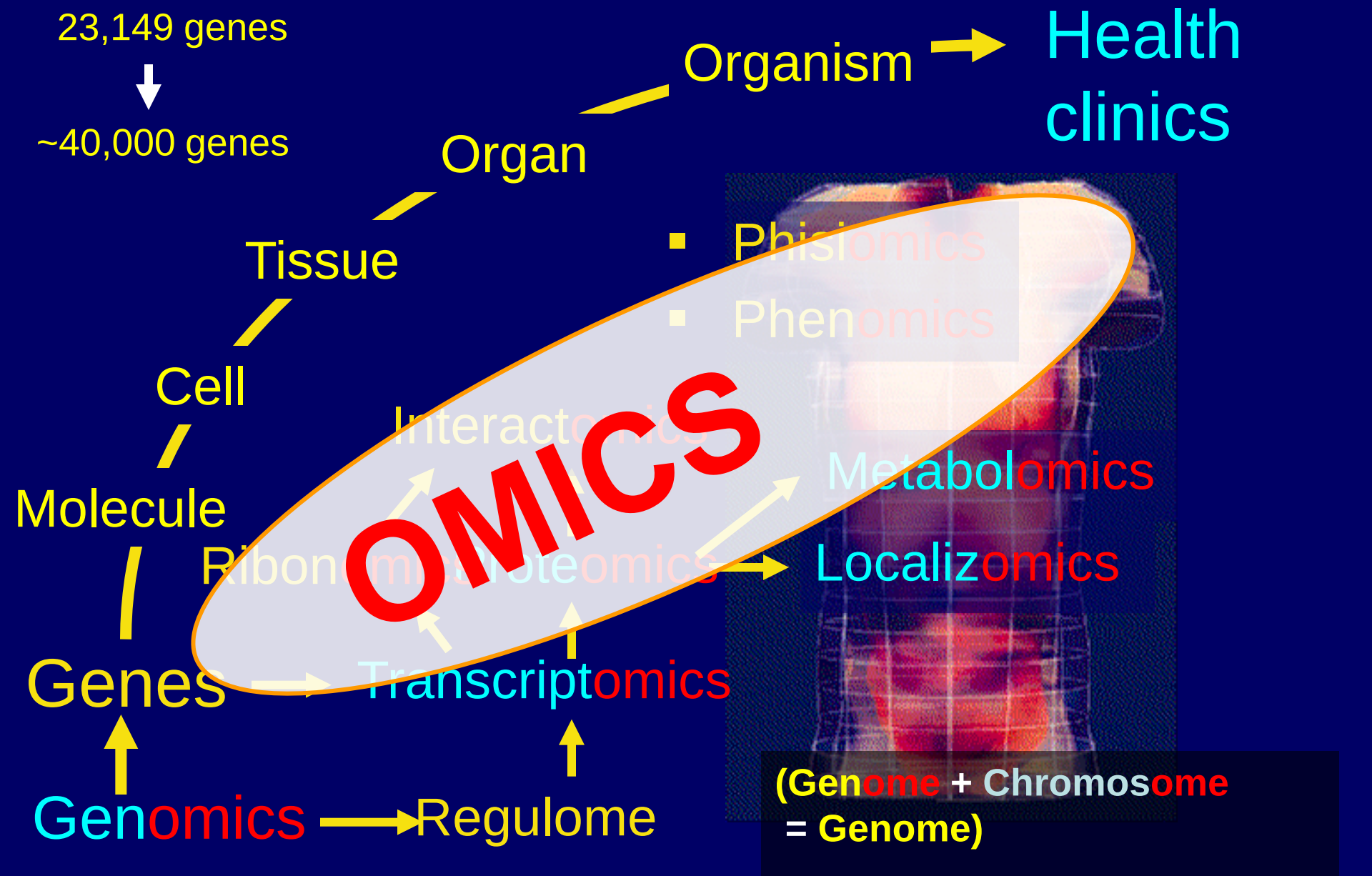
- Chapman, JA., Hayakawa, S., Hirose, M., Hwang, JS., Ikeo, K., Nishimiya-Fujisawa, C., Ogura, A., **Gojobori, T**, Fujisawa, T., Steele, RE. et al. (**2010**). The Dynamic Genome of Hydra ***Nature*** 464(7288): 592-596.

- FANTOM Consortium, Suzuki, H., Furuno, M, **Gojobori T**, Ikeo, K. 97 authors, and Hume DA; Riken Omics Science Center, Arakawa T, Hayashizaki Y (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. ***Nature Genetics*** 41(5): 553-62.

From the Genome Revolution to *Sequencing Revolution*

Sequencing =>
/ Genome sequencing
/ Meta-genomics
/ Gene Expression
Transcription- EST, SAGE, and CAGE)
/ miRNA, functional non-coding RNAs, siRNAs
(Translational regulation)
/ CHIP-Seq (CHIP-Chip, CHIP-Pet)
/ PPI (Two hybrid System)
/ Epi-genomics (Methylated sites)

Post-genome epoc: Future perspectives



- Problem -

- Sydney Brenner says,

  “Low input,

  High throughput, and

  No output” !

D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Hannick, W. Hide, D. Hill, R. Kania, M. Schaeffer, S. St Pierre, S. Tweigger, and S. Rhee

Nature (2008) 455: 47-50

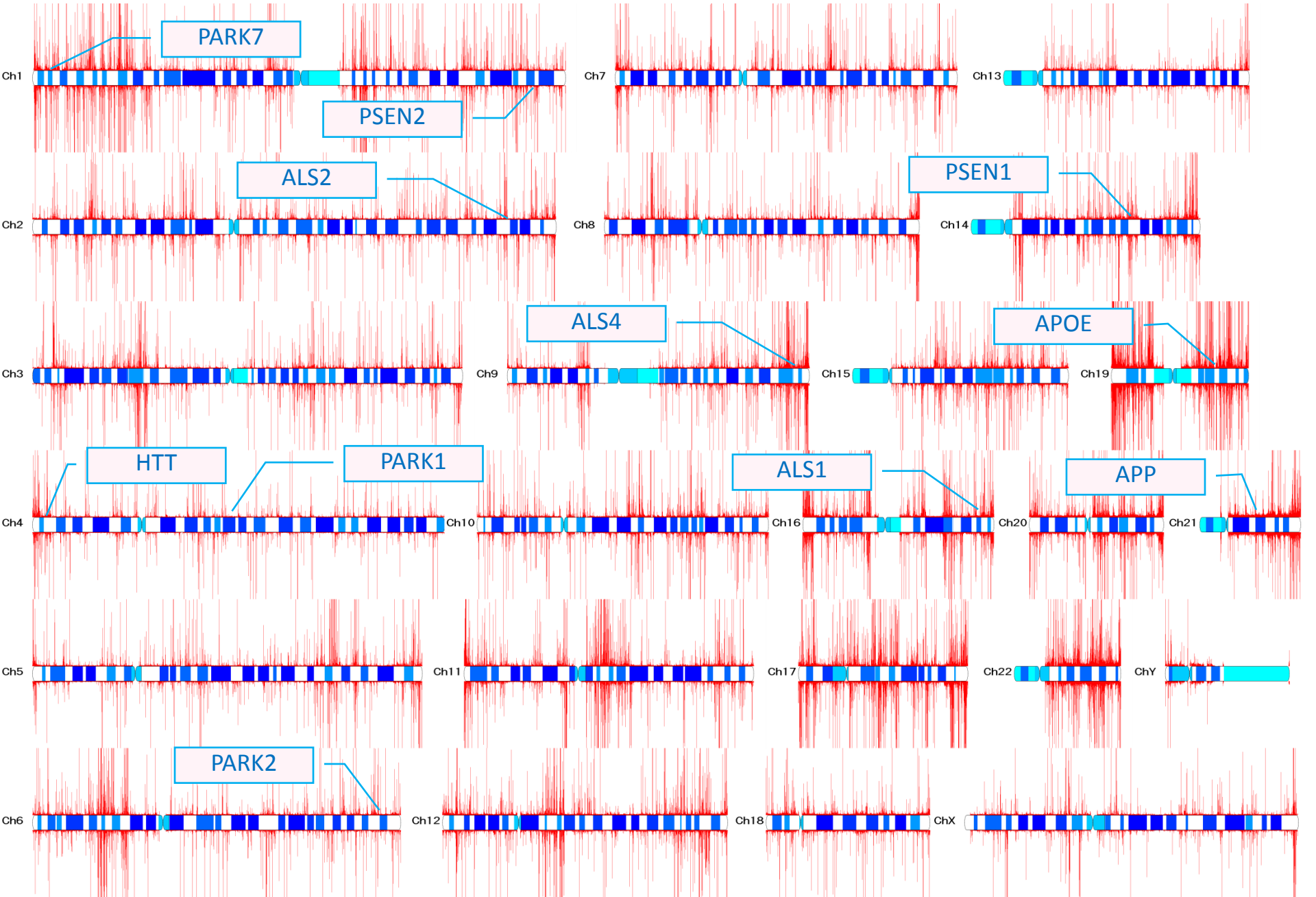
Genome Network Project (2005-2010)

http://genomenetwork.nig.ac.jp/

FANTOM Consortium, Suzuki, H., Furuno, M, **Gojobori T**, Ikeo, K. 97 authors, and Hume DA; Riken Omics Science Center, Arakawa T, Hayashizaki Y.
***Nature Genetics* (2009)41(5): 553-62**.

(Of course, English version is available)

Human CAGE Tag (2009/01) 46,205,347Tags



Contents

Evolution of the Central Nervous System (CNS) and brain.

Contents

Where can we obtain genes expressed specifically in a human brain?

“Human Full-Length cDNA Annotation Invitational” (H-Invitational)

August 25 - September 5, 2002

- Systematic Identification of Human Genes and its Biological Significance -

Co-organized byJBIRC and DDBJ/NIG

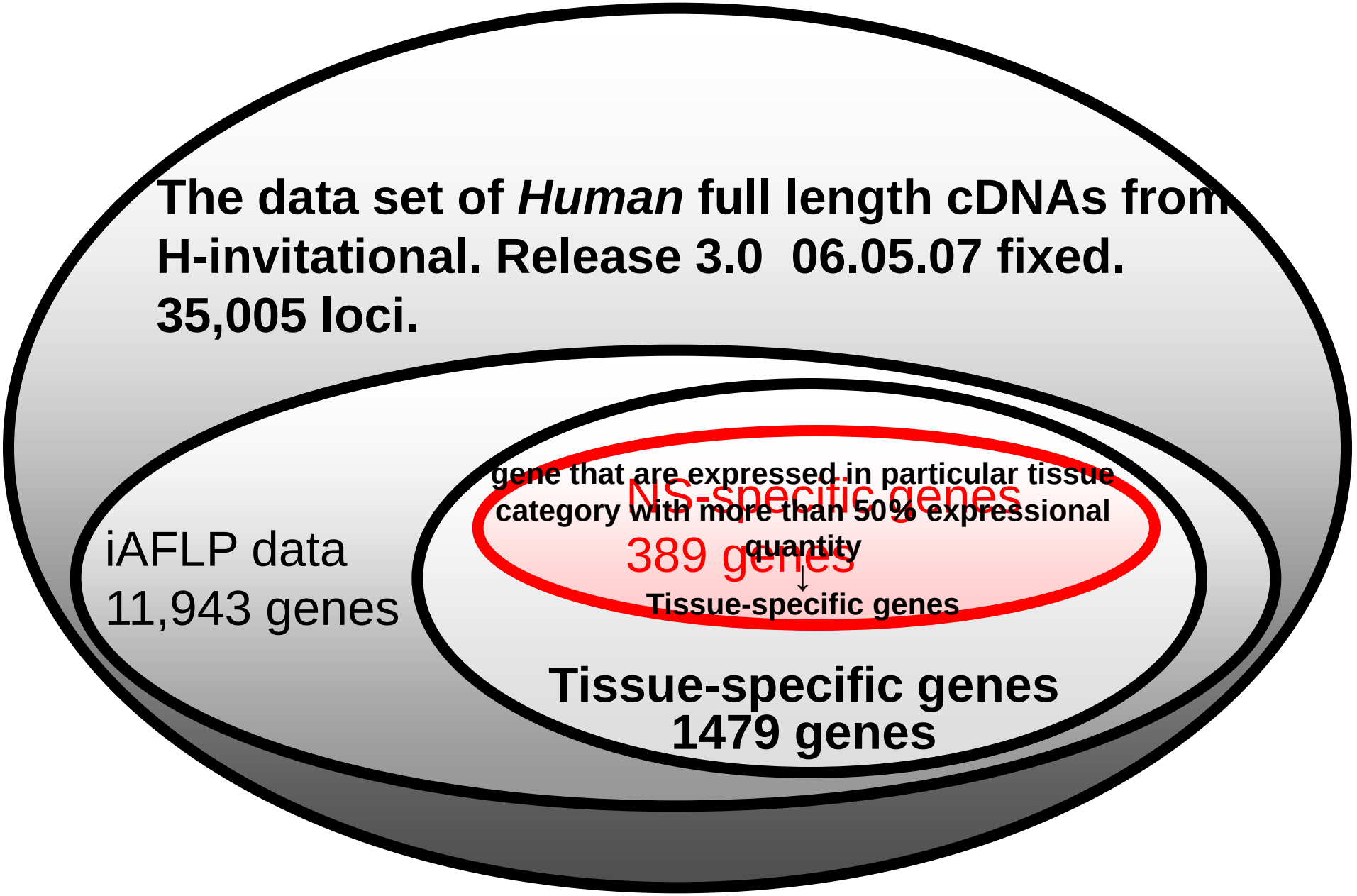
Attended by more than 118 people from 40 organizations such as

JBIRC, DDBJ, NCBI, EBI, Sanger Centre,NCI-MGC, DOE, NIH, DKFZ, CNHGC(Shanghai), RIKEN, Tokyo U, MIPS, CNRS, MCW, TIGR, CBRC, Murdoch U, U Iowa, Karolinska Int., WashU, U Cincinnati, Tokyo MD U, KRIBB, South African Bioinfor Inst, U College London, Reverse Proteomics Res. Inst., Kazusa DNA Inst, Weizmann Inst, Royal Inst. Tech. Sweden, Penn State U, Osaka U, Keio U, Kyushu U, TIT, **Ludwig Inst. Brazil**, Kyoto U, German Can.Inst., and NIG

Supported by

JBIC, METI, MEXT, NIH, and DOE

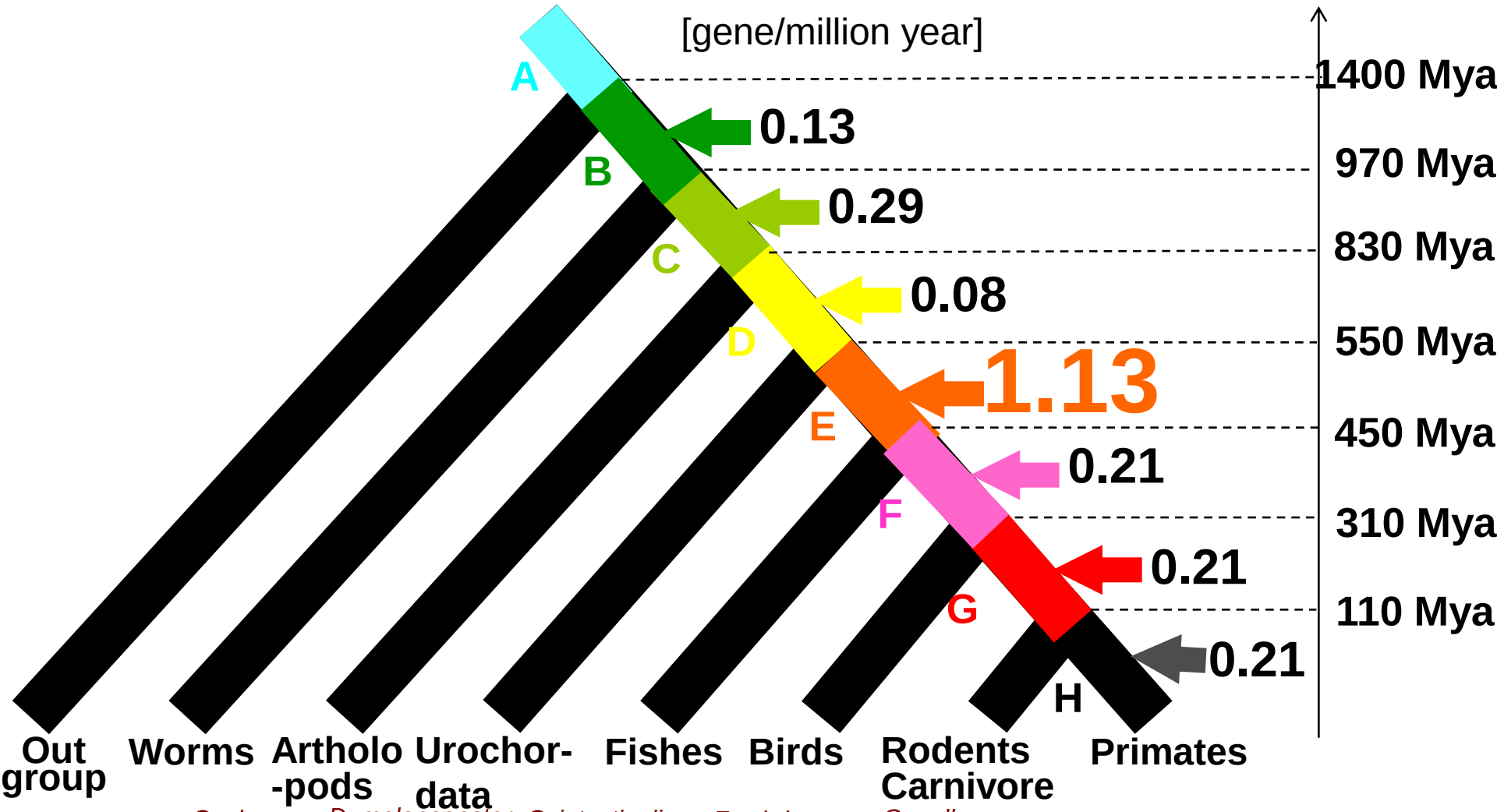
3 Materials: The data set used for this study



5-1 Results: The emergence time of the NS-specific genes

29% of the 389 NS-specific genes emerged during period E.

5-3 Results: The emergence rate of the NS-specific genes



Many genes had emerged during the evolutionary process during period E.

Dehal *et al.* 2002. Science vol298 p2157-2166

<Answer>

A) When did the NS genes emerge?

- The common ancestor of eukaryotes with no NS had already possessed 9% of the 389 human NS-specific genes.
- The emergence of the NS-specific genes have concentrated during the period E (from ascidians to vertebrate).

➔ An emergence pattern of the NS-specific genes is very dynamic.

B) What kinds of NS genes emerged?

- The genes whose functions are related to NS emerged during all periods.
- The genes of receptor activities emerged most during period B (from organisms having no NS to worms) and period E (from ascidians to vertebrate).

➔ An increase of receptor-related genes had facilitated the evolution of the NS during periods B and E (neurons and their network formation)

The examined genes specifically expressed in a human brain emerged even before the worms, and 30% of them emerged just before bony fish!!

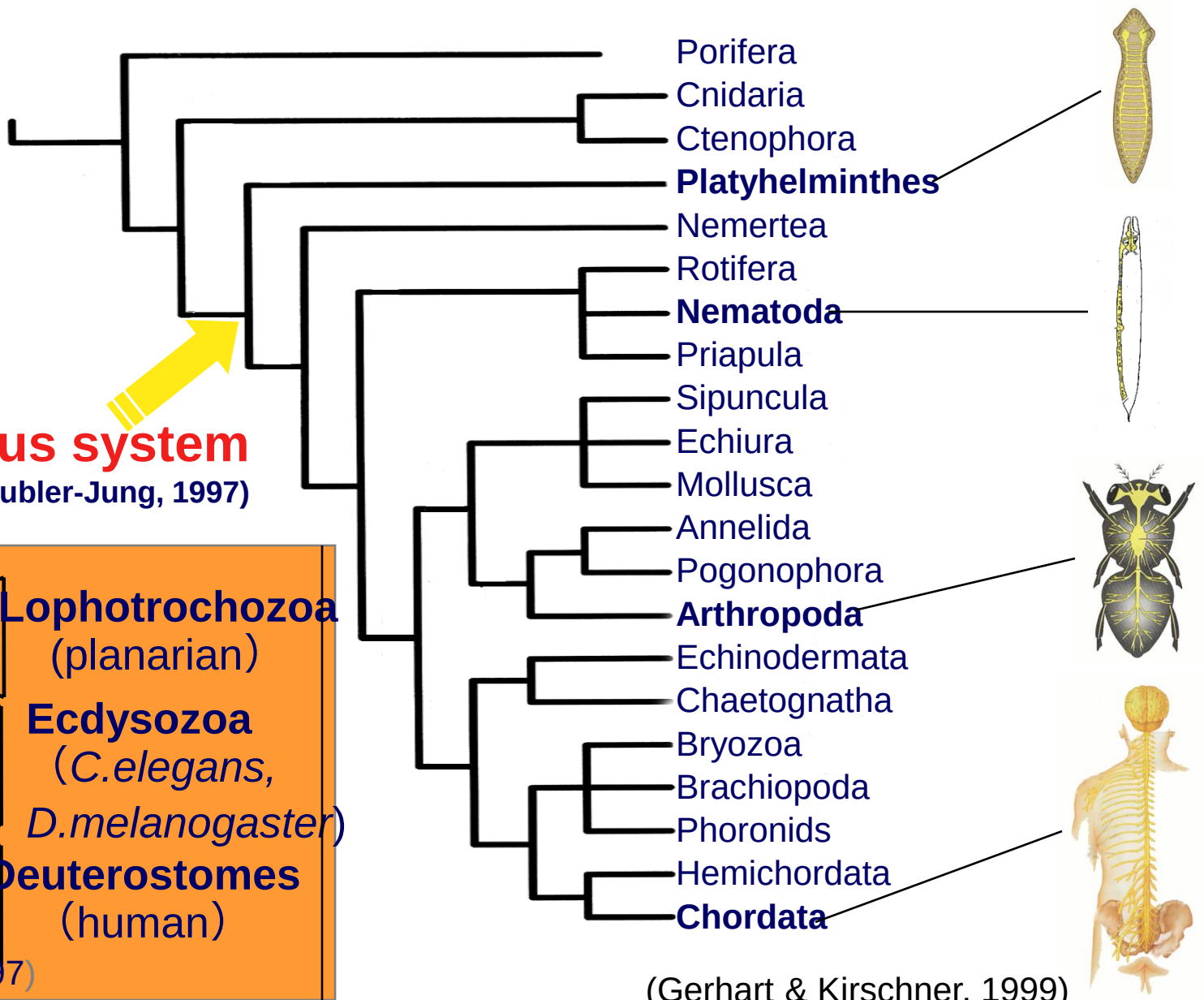
Contents

2) What kind of genes are expressed in a planarian that is an organism having the most primitive brain?

3) What kind of genes are expressed in a hydra that does not have CNS, but have only neural cells and nematocytes (motion-controlling cells)?

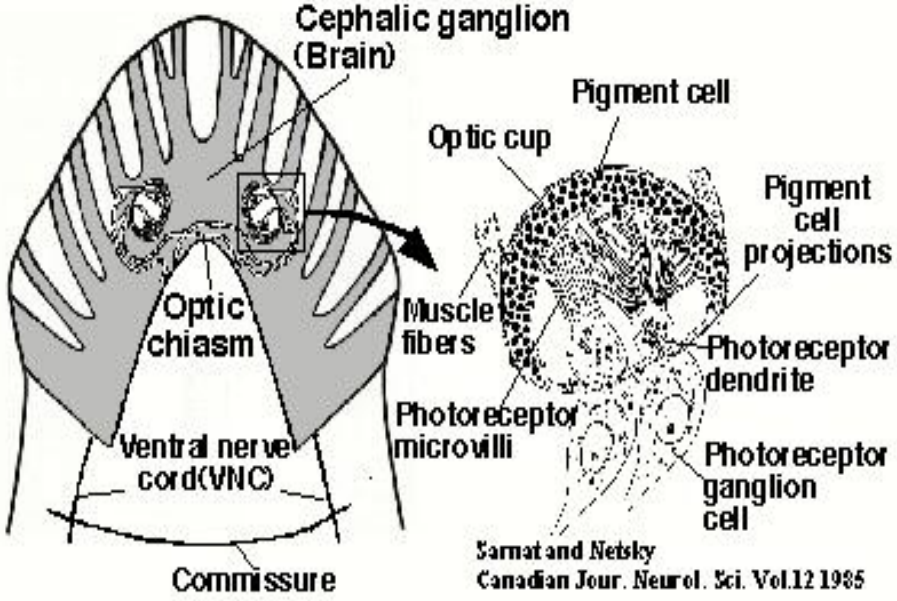
4) Summary

Phylogenetic relationship of metazoan and CNS

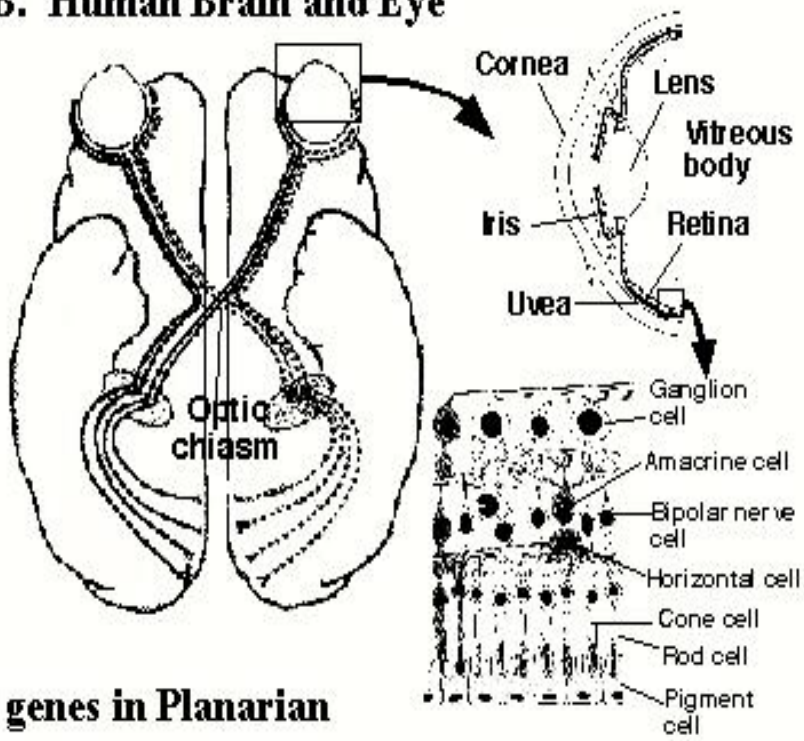


What is the same, what is the differences between Planarian and Human Brain?

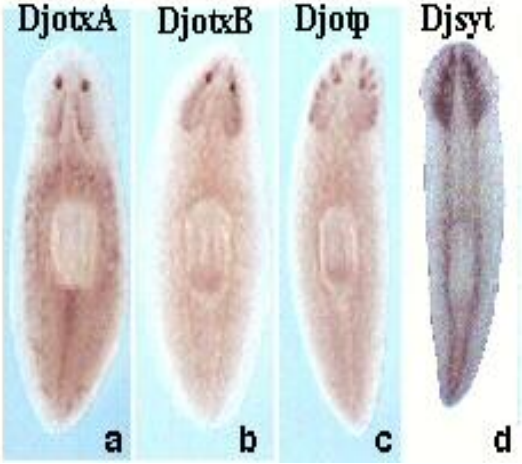
A. Planarian Brain and Eye



B. Human Brain and Eye



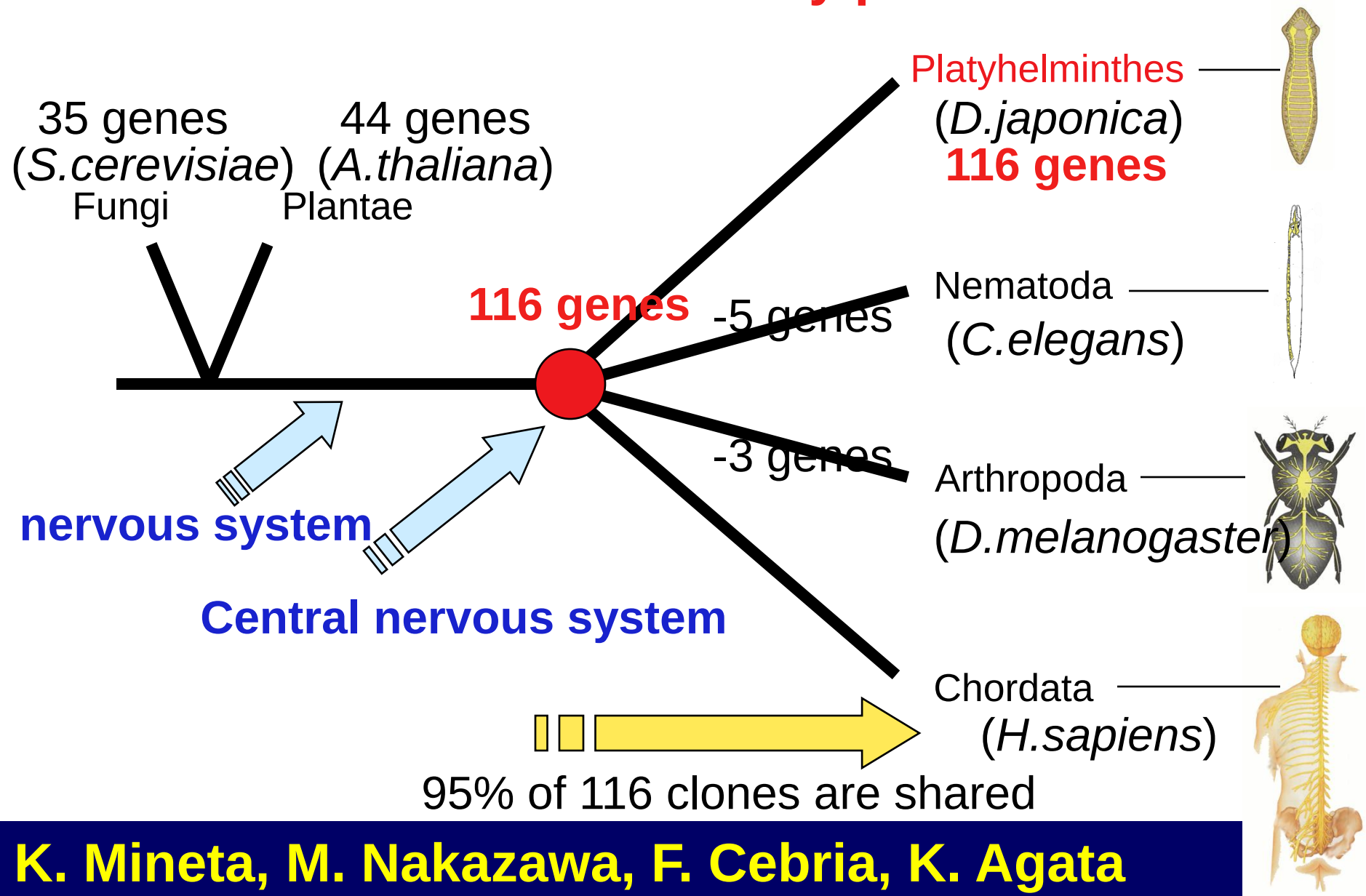
C. Whole mount *in situ* hybridization of brain-related genes in Planarian



These are the planarian cephalic ganglion (brain)-related genes already isolated in previous works. DjotxA and otxB, Djotp and Djsyt are orthodenticle (Otx), orthopedia (Otp) and synaptotagmin homologues, respectively.

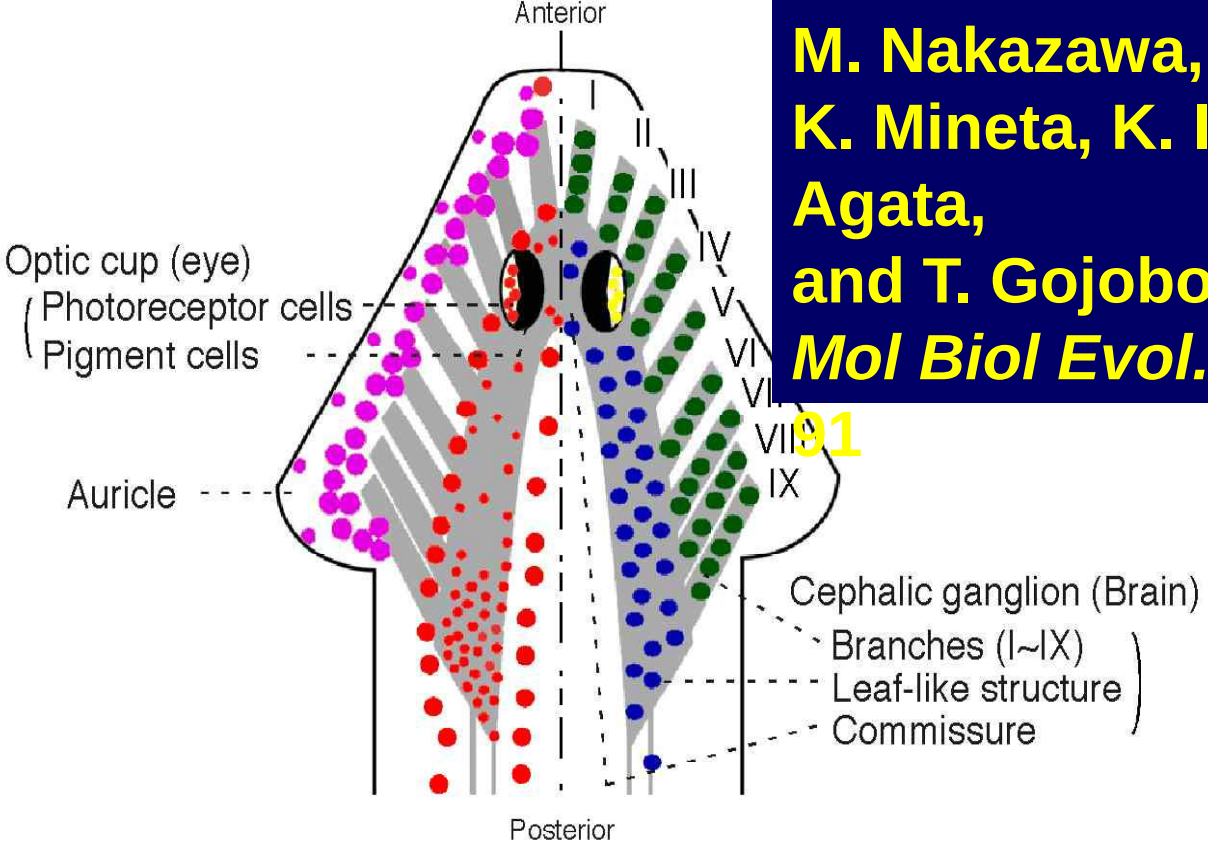
a, b, c: Umesono et al. Dev Genes Evol. 209:31-39 1999
d: Tazaki et al. BBRC 260:426-432 1999

SUMMARY of the evolutionary process of the CNS



K. Mineta, M. Nakazawa, F. Cebria, K. Agata
K. Ikeo, and T. Gojobori (2003) *PNAS* 100(13):7666-

M. Nakazawa, F. Cebria, K. Mineta, K. Ikeo, K. Agata,
and T. Gojobori (2003)
Mol Biol Evol. 20(5): 784-791



Type	EST clone	Expression area & patterns	Inferred function
A	0944_HH	Entire region in the CNS	Fundamental for nervous system
B	0005_E	Leaf-like structures	Signal processing & transporting to VNCs
C	1008_HH	Brain branches	Signal transduction
D	0251_HH	Eye	Visual system
E	0821_HN	Periphery of the head region	Sensory cells
F	0721_HH	Whole region of head	Head formation & maintenance
G	0053_E	Gradient expression in the brain	Signal processing

Figure 5 Cytoarchitecture map of the planarian CNS

The represented patterns of each type (A to G) are schematically drawn using different colors at a half side.

Double whole-mount immunostaining with

VC-1 (arrestin) and **anti-DjSYT** antibodies (synaptotagmin)

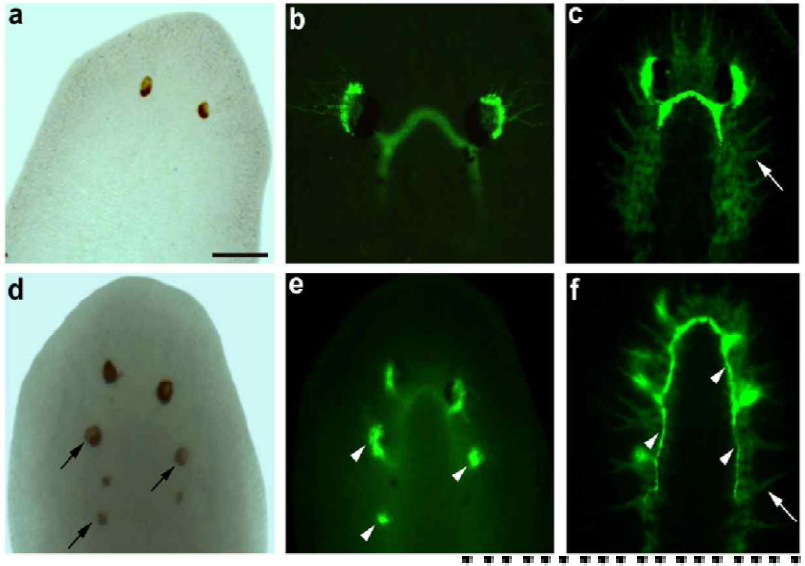


Figure 7 Ectopic eyes appear after 721_HH dsRNA inje
Bright field view of control (**a**) and 721_HH injected (**d**) an

In control animals a brain with its normal pattern of centra (**c**, pale green; arrow points to lateral branches). In 721_H appear in more posterior regions (arrows in **f**).

FGFR-related gene *nou-darake* restricts brain tissues to the head region of planarians

Francesc Cebrià*†, Chiyoko Kobayashi*†, Yoshihiko Umesono*, Masumi Nakazawa‡, Katsuhiko Mineta‡§, Kazuho Ikeo‡§, Takashi Gojobori‡§, Mari Itoh||, Masanori Taira||, Alejandro Sánchez Alvarado¶ & Kiyokazu Agata*#

Nature (2002) 419: 620-

<Answer>

1) A half of planarian genes specifically expressed in the head were shared with other higher organisms including human.
2) Those genes are expressed in a specific manner in a planarian head. It appears that there are 7 different regions. It implies that a planarian brain has already functional regionalization.
3) The emergence of the brain seems to be very old and complex already.

The emergence of the brain seems to be very old and complex already!!

Contents

Phylogenetic relationship of metazoan and CNS

Porifera
Cnidaria
Ctenophora
Platyhelminthes
Nemertea
Rotifera
Nematoda
Priapula
Sipuncula
Echiura
Mollusca
Annelida
Pogonophora
Arthropoda
Echinodermata
Chaetognatha
Bryozoa
Brachiopoda
Phoronids
Hemichordata
Chordata

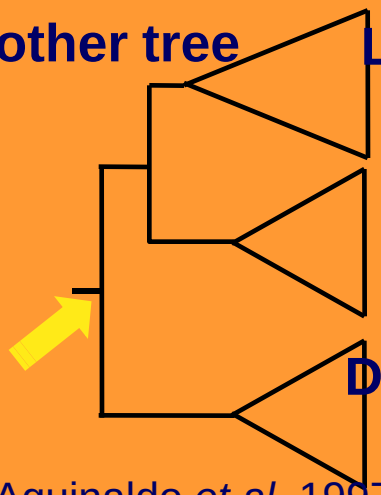
Hydra

(Gerhart & Kirschner, 1999)

Central nervous system

(Arendt & Nubler-Jung, 1997)

other tree



Lophotrochozoa
(planarian)

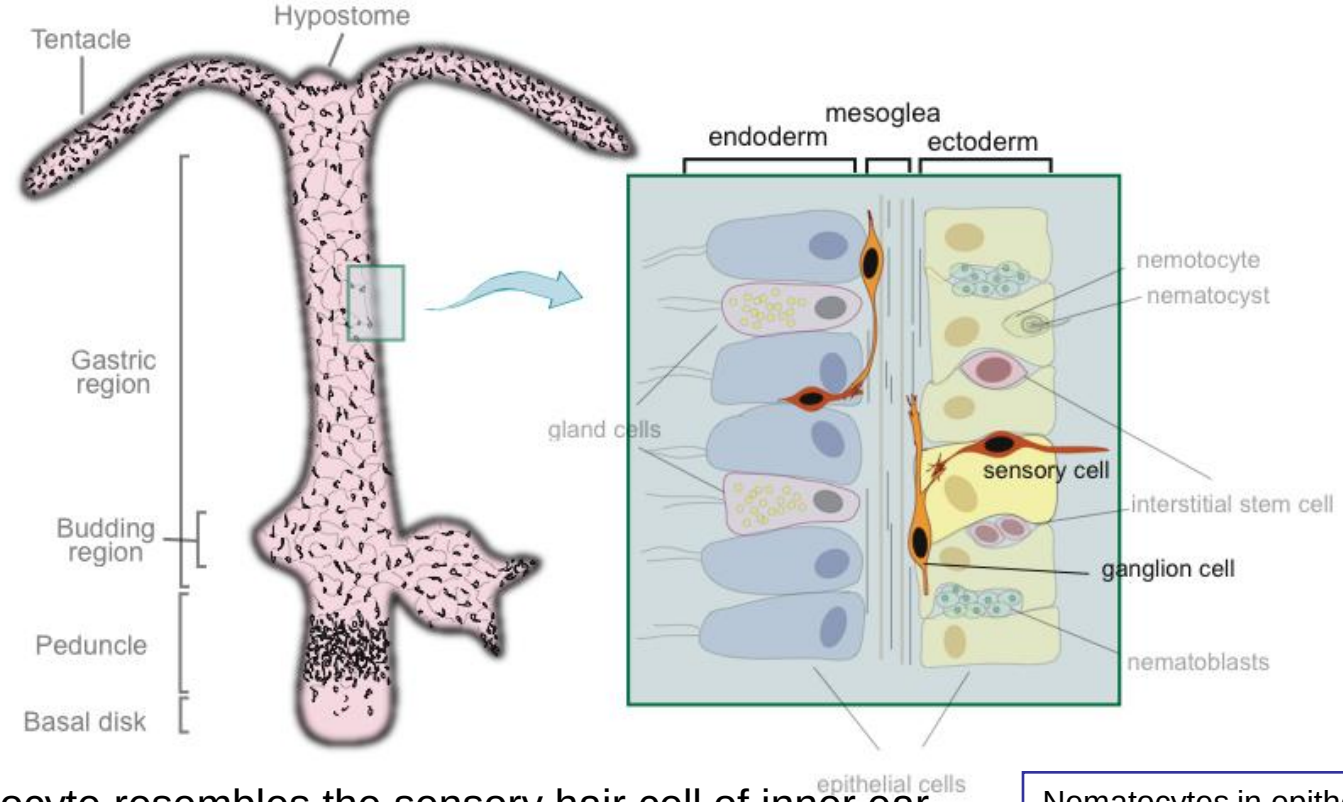
Ecdysozoa
(*C.elegans*, *D.melanogaster*)

Deuterostomes
(human)

(Aguinaldo *et al*, 1997)

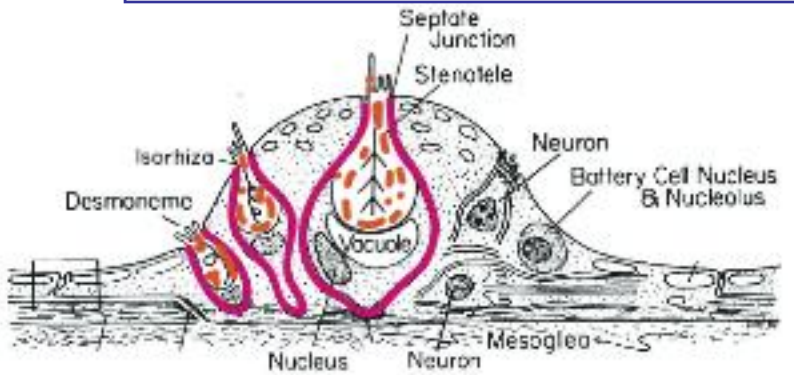
The origin of nervous system: Cnidaria--Hydra

(1) Diffuse Nerve Net



(2) Nematocyte resembles the sensory hair cell of inner ear.

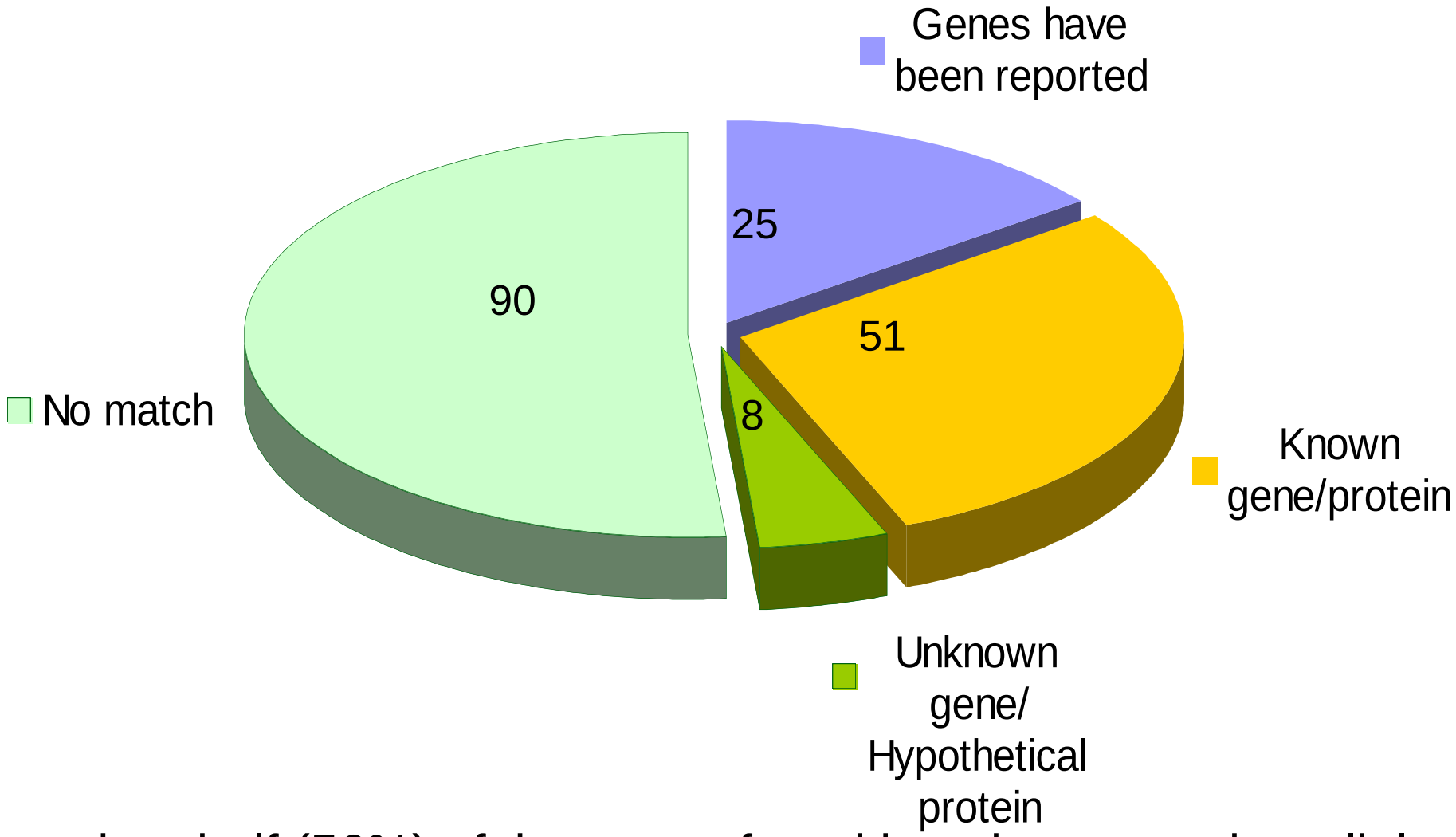
Nematocytes in epithelial battery cell of hydra



Organ of Corti

The figure is modified from Fig. 2 in Kass-Simon G. and Scappaticci, J.J. (2001) The behavioral and developmental physiology of nematocysts. Can. J. Zool. 80:177

174 genes (normal >> epithelial)



More than half (56%) of the genes found by microarray share little or no information to our current knowledge.

<Answer>

1) A half of hydra genes specifically expressed in the neural cells and nematocytes were shared with other higher organisms including human.
2) Those genes are expressed in a specific manner in those cells, though they do not have the central nervous system. It appears that there are unknown network that is different from the CNS and brain.
3) The emergence of genes specifically expressed in the neural cells seems to be very old.

The genes specifically expressed in hydra neural cells seems to have specific patterns of expression!

Chapman, JA., Hayakawa, S., Hirose, M., Hwang, JS., Ikeo, K., Nishimiya-Fujisawa, C., Ogura, A., **Gojobori, T**, Fujisawa, T., Steele, RE. et al.

The Dynamic Genome of Hydra.

Nature (2010) 464(7288): 592-6.

Summary

- Earlier existence of gene sets that were potentially capable of forming the nervous system
- Diversification of gene sets forming gene network and probably regulatory systems
- Lineage-specific evolution of neural-related genes.
- Explosive emergence of neural human genes just before the evolutionary appearance of vertebrates.
- 3D visualized database is useful for comparative gene expressionics
- Integration – logical connection of biological information scattering at different levels of biological hierarchy

Genes that are potentially capable of forming the CNS and brain are terribly old in term of evolutionary emergence!! For formation of the CNS and brain, Gene network must be a key.

Acknowledgements

/ Genome Network Project Consortium
- RIKEN (Hayashizaki's group)
- Tokyo U. (Sugano's group)
- Hitachi Co. Ltd.
- Kieo U. (Yanagawa's group)
- Saitama Medical College (Okazaki's group)

/ H-Invitational Consortium

/ BIRC at AIST, Tokyo, Japan

/ NIG – Genome Network Platform group