

STRUCTURE - ORGANISATION - DYNAMIQUE ET POLYMORPHISME DU GENOME HUMAIN

Encadrement réglementaire des activités de génétique en France

Introduction

Convention d'Oviedo (Conseil de l'Europe, STCE n°164) : seul instrument juridique international contraignant en matière de bioéthique. Inspirée très largement des **lois françaises** de bioéthique.

Convention pour la protection des Droits de l'Homme et de la dignité de l'être humain à l'égard des applications de la biologie et de la médecine: Convention sur les Droits de l'Homme et la biomédecine

Oviedo, 4.IV.1997



Article 11	Article 12	Article 13	Article 14
Non-discrimination « Toute discrimination à l'encontre d'une personne en raison de son patrimoine génétique est interdite. »	Tests génétiques prédictifs « Il ne pourra être procédé à des tests prédictifs de maladies génétiques ou permettant soit d'identifier le sujet comme porteur d'un gène responsable d'une maladie soit de détecter une prédisposition ou une susceptibilité génétique à une maladie qu'à des fins médicales ou de recherche médicale, et sous réserve d'un conseil génétique approprié . »	Interventions sur le génome humain « Une intervention ayant pour objet de modifier le génome humain ne peut être entreprise que pour des raisons préventives, diagnostiques ou thérapeutiques et seulement si elle n'a pas pour but d'introduire une modification dans le génome de la descendance . »	Non-sélection du sexe « L'utilisation des techniques d'assistance médicale à la procréation n'est pas admise pour choisir le sexe de l'enfant à naître, sauf en vue d'une maladie héréditaire grave liée au sexe. »

Protocole additionnel à la convention sur les Droits de l'Homme et la biomédecine relatif aux tests génétiques à des fins médicales (STCE n°203) :

Article 3
Primauté de l'être humain « L'intérêt et le bien de l'être humain concerné par les tests génétiques visés par le présent Protocole doivent prévaloir sur le seul intérêt de la société ou de la science. »

Fondements : Loi n°94-654 du 29 juillet 1994 Relative au **don** et à l'**utilisation des éléments et produits du corps humain**, à l'**assistance médicale à la procréation** et au **diagnostic prénatal**.

Bioéthique : bio « vivant » et éthique « ce qui est bon pour l'homme »

Révisions en 2004 par J-F Mattei (prof de génétique et ministre de la Santé), en 2011 et en 2018 (actuellement) grâce à :

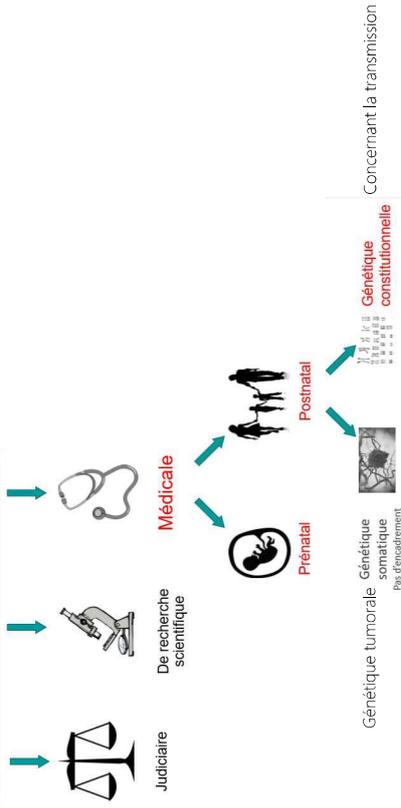


- ♣ **CCNE** : **Avis 129**, contribuant à la révision des lois de bioéthique et l'organisation d'Etats généraux (avis des citoyens)
- ♣ **Conseil d'Etat**
- ♣ **OPECI** (Office Parlementaire d'Evaluation des Choix scientifiques et Technologiques)
- ♣ **ABM**

→ **Projet de loi** relatif à la bioéthique adopté par Assemblée Nationale en 1^{ère} lecture le 15 octobre 2019

Examens de génétique

Trois finalités aux examens génétiques



Droit de bioéthique s'appliquant aux tests figure dans 3 codes :

- × **Code de Santé Publique**
- × **Code Civil**
- × **Code Pénal**

Art. L. 1131-1 du code de S publique : Examen des **caractéristiques génétiques d'une** personne à des fins médicales consiste à analyser ses caractéristiques génétiques **héritées ou acquises** à un stade précoce du développement prénatal (cancer non concerné). A pour objet :

- ❖ Poser, confirmer ou infirmer le **diagnostic** d'une maladie à caractère génétique chez une personne
- ❖ Rechercher caractéristiques génétiques de **gènes susceptibles** d'être à l'origine du **dévt** d'une maladie chez une personne ou les membres de sa famille potentiellement concernés
- ❖ Adapter la **prise en charge** médicale d'une personne selon ses caractéristiques génétiques

Art. L.1131-2 : Analyses aux fins de détermination des caractéristiques génétiques d'une personne à des fins médicales :

- ❖ Analyses de **cytogénétique** (moléculaire), càd chromosomes
- ❖ Analyses de **génétiq ue moléculaire**, soit molécule ADN
- ❖ Te autre **analyse de bio méd** prescrite ds intention obtenir des infos pr détermination des caractéristiques génétiques d'une personne équivalentes à celles obtenues par les analyses mentionnées ci-dessus. Ex : diagnostic de drépanocytose par électrophorèse (test sur une protéine)

Art. L.1131-3 : Seuls st habilités à procéder à des examens des caractéristiques génétiques d'une pers à des fins médicales les **praticiens agréés** à cet effet par l'ABM.



Art. L.1131-2-1 : Les examens des caractéristiques génétiques d'une pers à des fins médicales ne peuvent é pratiqués que ds des **labos de bio méd autorisés** à cet effet par l'ARS.



Plan France Médecine Génomique 2025

Pédicatif de la génomique en 2025. Objectifs :

- intégrer **séquencage** du génome **en routine** clinique
- développer une **filère nationale** de médecine génomique



Enjeux de S publique :

- Accès **égal** au séquençage
- Modification du **parcours et organisation des soins** (prise en charge, délais)

Enjeux scientifiques, technologiques

- Meilleure **compréhension** des pathologies
- **Recherche** clinique
- Dévts technologiq et d'une **expertise** en sciences du calcul et des données

Enjeux économiques :

- **Réduction des coûts** du système de soin
- Dév't d'une **filière industrielle**

Acteurs : Aviesan, Seqoia, CCNE, ABM, Orpha.net



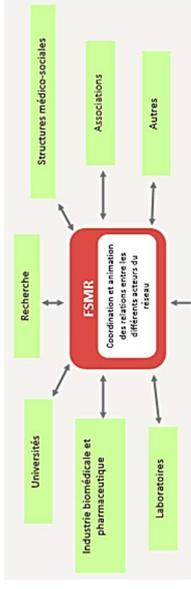
Maladies génétiques

Regroupent les pathologies qui résultent d'une **altération des gènes**. Font partie des **maladies rares** : **moins d'une personne sur 2000** (soit 30 000 personnes en France), dot 80% seraient d'origine génétique. En F : 8000 maladies rares identifiées, touchant plus de **3 millions de personnes** (4,5% de la population). Ds moitié des cas : enfants de moins de 5 ans ; resp de 10% des décès entre 1 et 5 ans ; ¼ apparaissent après 40 ans. Un grd nb de ces maladies = également maladies orphelines, càd sans traitement.

PNMR (Plans Nationaux Maladies Rares) : 1 (2005-2008), 2 (2011-2015) et 3 (2018-2022). + actions spécifiques au niveau européen.

PNMR2 a mis en place **CRMR** (centres de réf maladies rares), structurés par des **FSMR** (filières de S maladies rares).

23 FSMR (Filières de Santé Maladies Rares)
 COMPOSÉS DE :
 - 109 CRMR (Centres de Référence Maladies Rares)
 - 778 sites constitutifs
 - 1844 centres de compétence



Structure de l'ADN, taille des génomes et valeur C

Génome : ensemble de l'info héréditaire d'un organisme, présente en totalité dans chaque cellule.

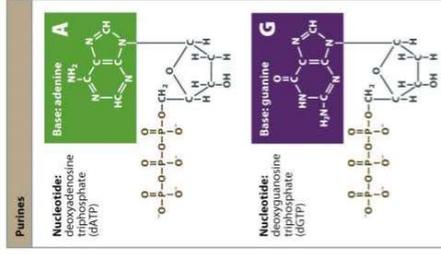


Figure 1. Molecular Biology (© Garland Science 2016)

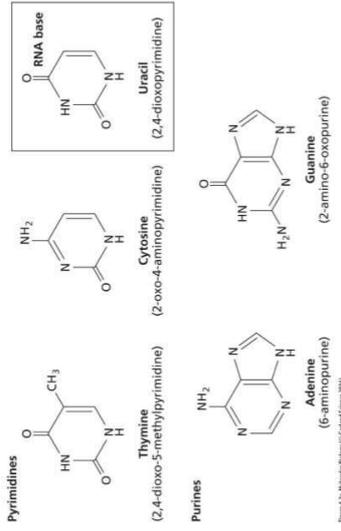
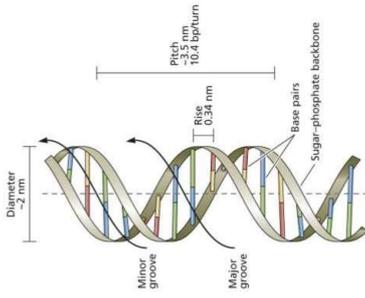


Figure 2.3 Molecular Biology (© Garland Science 2016)



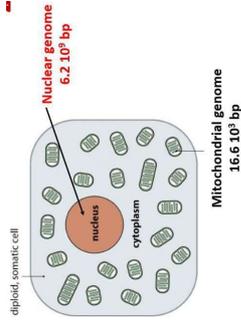
Brins d'ADN orientés : 5' (phosphate) →
3' (OH du sucre)
10 bases par tour de spire

1 pg ⇔ 978 Mb

Valeur C : taille d'un génome (spécifique d'espèce). Exprimée en Mb (millions de pnb) ou en pg (10^{-12}). Très variable en fonction des espèces : qq milliers pr virus, qq millions pr bact, qq milliards pr mammifères.

Taille des génomes ne corrèle pas avec la complexité des organismes = **paradoxe de la valeur C.**

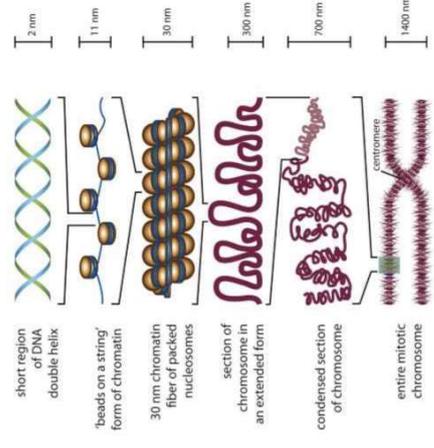
Présentation générale des génomes humains nucléaire et mitochondrial



2 copies par cell diploïde, 46 molécules linéaires, 23 chromosomes
Le + petit : 21 (50 Mb), Le + grd : 1 (250 Mb)
→ Ratio de 1 à 5

2-10 copies par mitoch, 100 à 1000 par cell, molécule circulaire.
Protéome mitoch : ~ 1000 prot, dt 1% codées par génome mitoch (13). Très autres st importées (codées par génome nucléaire).

	NUCLEAIRE	MITOCHONDRIAL
Taille	3,1 Gb (2,9 d'euchromatine)	16,6 Kb
Nb molécules d'ADN différentes	23 (XX) 24 (XY)	1
Nb total molécules d'ADN par cell	46 (diploïdes) 23 (haploïdes)	100 à 1000 (nb de copies variant entre types cell)
Protéines associées	Plusieurs classes d'histones et non-histones	Largement libre
Nb gènes codant protéines	19 965	13 (→ enz chaîne resp)
Nb gènes codant ARN	25 486	24
Densité gènes codant protéines	1/100 kb	1/0,45 kb
ADN répété	50%	Très peu
Transcription	Chaque gène a son/ses propre(s) transcrit(s)	Un par brin
Introns	Dans la plupart des gènes	-
ADN codant	12 %	66%
Code génétique	61 codant AA, 3 STOP	60 codant AA, 4 STOP
Recombinaison	Au moins pour chaque paire d'homologues pdt méiose	Non évidente
Héritage	Mendélienne (autosomes + X) Paternelle (Y)	Exclusivement maternelle



Chromatine : Association ADN et histones (très riches en AA basiques), sièges de modifications biochimiques.

Ds noy eucaryotes, organisation ADN en chromatine = essentielle pr **intégrité génomique**. Rôles :

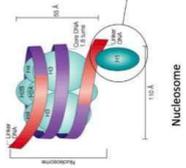
- ♣ Physique, compaction et protection ADN
- ♣ Etablissement et maintien des états d'expression des gènes à travers les divisions cell + identité cell

Noyau caract par domaines distincts de chromatine :

EUCHROMATINE	HETEROCHROMATINE
Décondensée	Très compactée
Transcription active	Régions non transcrites

Constitutive : portions inactives de chromosomes ds ttes les cell (centromères et télomères). Généralement répliquées tardivement

Facultative : état variable selon type de cell ou état de différenciation



Complexe ADN-Histones = **nucléosome**. unités fondam de chromatine. Comprend : région centrale + région inter-nucléosomique caract par présence de H1.

Région centrale : 146 pdb enroulées autour octamère protéique, **H2A, H2B, H3, H4** (x2). Extrém N-Term dépassent de struct du nucléosome. → **modif post-traductionnelles** (acteurs expression gènes)

Code génétique

Genome nucléaire et genome ribosomialisé - 4 codons interrompés - entièrement

AAA	Lys	AAA	Glu	AAA	STOP
AAG	Lys	AAG	Glu	AUA	Met
AAC	Asn	AAC	Asp	UAC	Tyr
AAT	Asn	AAT	Asp	UAA	STOP
ACA	Thr	ACA	Pro	UCA	Ser
ACC	Thr	ACC	Pro	UCC	Ser
ACG	Arg	ACG	Arg	UCU	Ser
ACU	Arg	ACU	Arg	UCG	Ser
AGU	Ser	AGU	Gly	UUA	Leu
AGC	Ser	AGC	Gly	UUG	Leu
AUA	Met	AUA	Glu	UUA	Leu
AUG	Met	AUG	Glu	UUG	Leu
AUC	Ile	AUC	Val	UUC	Phe
AUU	Ile	AUU	Val	UUG	Phe

Code génétique

→ nécessité ARNt corresp aux spécificités de son code génét + machinerie ribosomique spécifique

The Human Genome Project

Generating a Reference Genome Sequence (e.g., Human Genome Project)



Break genome into large fragments and insert into clones

Order clones

Break individual clones into small pieces

Generate thousands of overlapping and assemble sequence of clone

Assemble sequences of overlapping clones to establish reference sequence

Projet **séquençage génome** (1990–2003) : F. Sanger, L. Hood, J. Weissenbach, C. Ventier, F. Collins

1968 : 1^{er} séquençage de 12 nt (bactériophage)

Nouvelle méth séquençage (Gilbert) → 1979 : 1^{er} séquençage complet, génome du virus hépatite B

1977 : **Fred Sanger** (Prix Nobel de chimie 1955 -Insuline- et 1980 avec Gilbert -détermination des séquences des bases des acides nucléiques)

1986 : détection de fluorescence dans séquençage automatisé (**Leroy Hood**).

2001 : (*Science, Nature*) séquençage complet génome. A fait l'objet d'une annonce officielle par Maison Blanche.

John Craig Venter : 1^{er} hô dt génome a été séq (via méth Sanger).

« We predict that in the long view of history, the impact of DNA sequencing will be on a par with that of the microscope »

Genome sequenced	HGP (2003)	Watson (2008)	2019
Time taken (start to finish)	13 years	4.5 months	# days
Number of scientists listed as authors	> 2 800	26	-
Cost of sequencing (start to finish)	2.7 billion \$	# 1.5 million \$	< 10 000 \$
Number of institute involved	16	2	-
Number of countries involved	6	1	-

PREFIX	SYMBOL	Puissance de 10
EKA	E	10 ¹⁸
PETA	P	10 ¹⁵
TERA	T	10 ¹²

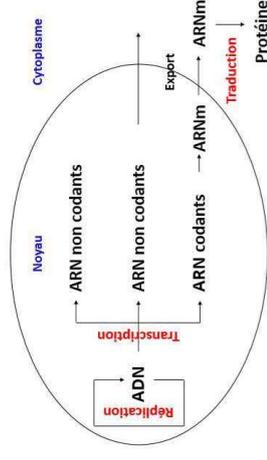
Génome humain : 60 000 gènes, dt 20 000 codant pr prot (2/3 ne codent pas pr prot), 25 000 codant pr ARN et 15 000 pseudogènes. 400 codent pr lg et segments du TCR.

The Genome Aggregation Database (gnomAD v2.1) → 15 000 génomes répertoriés.

Présentation des gènes codant les protéines

Gène : ensemble des **séquences génomiques** (ADN ou ARN pr les virus) codant un **ensemble cohérent** de produits fonctionnels (ARN ou chaîne polypep) potentiellement chevauchant.

(Ensemble cohérent : union de ts les produits ayant au moins une région commune en considérant le cas des produits ARN et polypep séparément)



Mise en évic **complexité organisationnelle** par proportion ARN non-codant (pas de paradoxe, cō pr valeur C).

Caract génome humain : 20 000 gènes codant prot dt exons occupent 1,2% du génome.

POLYMERASES	RNA
RNA polymérase I	28S, 5.8S, 18S rRNA
RNA polymérase II	mRNA, snRNA, miRNA, siRNA, piRNA, snoRNA, snRNA, lncRNA, LINEs
RNA polymérase III	tRNA, 5S rRNA, 7SL RNA, SINEs, ...

Promoteur : région de ADN située habituellement en amont (5') gène et **indisp à transcription**. Comporte site d'**initiation transcription** (+1) et sites fixation de facteurs de transcription et de l'ARN pol. Cz hô, interaction indirecte entre ARN pol et ADN (plateforme). Par convention, 1^{er} nt transcription appelé +1 (-1 celui qui le précède).

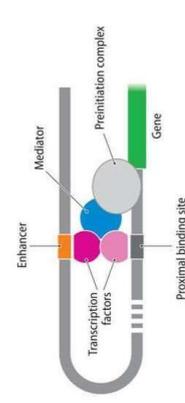
Contenu ADN cell somatiques (sauf lymphos et neurones) est identiq à celui œuf fécondé. Différenciation cell = due au profil gènes exprimés. On distingue :

- **gènes domestiques** (housekeeping genes), exprimés ds ttes cell, essentiels aux fonctions générales de cell
- **gènes tissu-spcifiques** (tissue-specific genes) exprimés ds cert cell ou tissu seulement

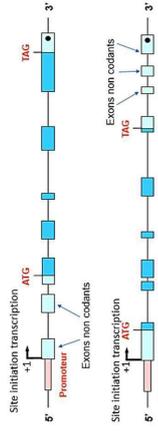
Enhancers : sites fixation fact de transcription pr **activer** transcription gènes. Agissent en cis et st situés, jusqu'à 1 Mb en amont/aval des gènes régulés.

Silencers : sites fixation fact de transcription pr **inhiber** transcription gènes. Agissent en cis et ... [ibid].

Insulators : sites fixation fact empêchant un enhancer/silencer d'activer/inhiber transcription d'un gène voisin (**régulation**).



Exons non codants



Exon : partie du gène persistant pdt maturation (épissage) du transcrit primaire en ARNm. Ne st pas ts codants, ni nécessairement multiples de 3 bases.

Intron : partie gène située entre 2 exons, excisée pdt épissage en ARNm.

UTR (UnTranslated Region) : régions transcrites non traduites (5'UTR et 3'UTR)

Séquences consensus : séquences nt impliquées ds fonctions semblables et ne présentant entre elles que qq variations de séq nt. Ex : gt—intron—ag

Séquence consensus du site 5' d'épissage chez les vertébrés

Base	-3	-2	-1	1	2	3	DEUT DIVERG	4	5	6	7
G	42	1542	1533	1	4	26	50	71	1056		
A	1269	2296	236	1	2353	2034	291	647	949		
T	182	104	10	1	10	10	10	10	10		
C	185	459	352	0	393	305	315	138	877	736	
Consensus	G	A	G	G	T	ANG	A	G	T	N	

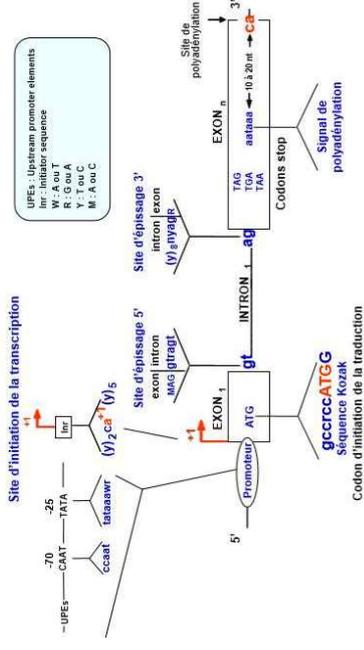
Séquence consensus du site 3' d'épissage chez les vertébrés

Base	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	EXON 2	
G	364	304	397	458	303	311	249	253	885	25	1	5843	1678	811
A	1269	1044	810	1036	1022	1015	1732	1024	939	880	1	404	1295	
T	182	104	10	10	10	10	10	10	10	10	1	10	10	
C	1080	1043	1158	1295	1418	1538	1571	1574	1137	3028	0	2	534	727
Consensus	TIC	N	CT	A	G	G	A							

Site donneur corresp à la charnière exon/intron.

Site de branchement, riche en pyrimidines et qui contient un A, = située à une trentaine de bases en amont du **site accepteur**, qui corresp à charnière intron/exon

Caractéristiques principales d'un gène codant protéine



Promoteur : TATA box (-20 ou -25), riche en T et en A

+1 : 1^{ère} base de lexon (pas forcément codant), initiation transcription. Puis : AUG, initiation traduction, précédé de **séq consensus Kozak** (gccrccatgg, avec r = g ou a).

Intron : gt...ag ac séq consensus.

Codons stop : UAA, UGA, UAG

a : site de **polyadénylation**

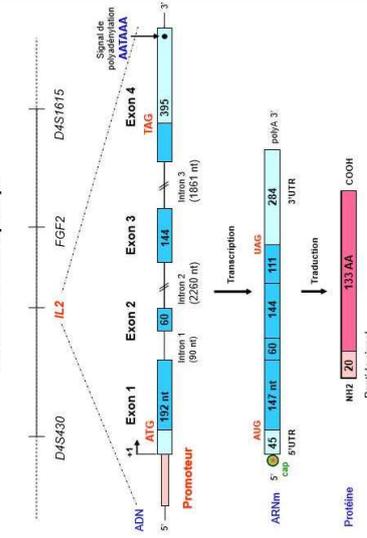
Interleukine 2 (IL2) : glycoprot codée par gène tissu-spécifique (cell T activées), expansion clonale des T immatures.

Chromosome 4q26-q27

Gène localisé sur **bras long chromosome 4, région 2, bande 6 (4q26-27)**. Promoteur (sites de fixation fact de transcription) = régulatrices (entre bases -548 et +39), 5 kb, 4 exons codants.

ARNm : présent seulement ds cell T activées. 800 nt, 5'UTR de 45 nt et 3'UTR de 284 nt.

Protéine : précurseur de 153 AA + modifications post-traductionnelles : clivage peptide signal (forme mature 133 AA) et réactions glycosylation et sialylation.



Traduction : Codon AUG reconnu par sous-u 40s → Initiation, définit ordre de lect. 60s recrutée. Puis : élongation (ARNt = conversation codon-anticodon) et terminaison (codons STOP : UAA, UAG et UGA)

Code génétique :

- dégénéré : plusieurs codons correspondent à 1 AA
- non ambigu : 1 codon → 1 AA

Codons synonymes non employés ac même fréq : biais d'usage des codons propre à chaque espèce. Gènes les + exprimés, càd ceux pr lesquels ARNt st les + nb, st les + biaisés

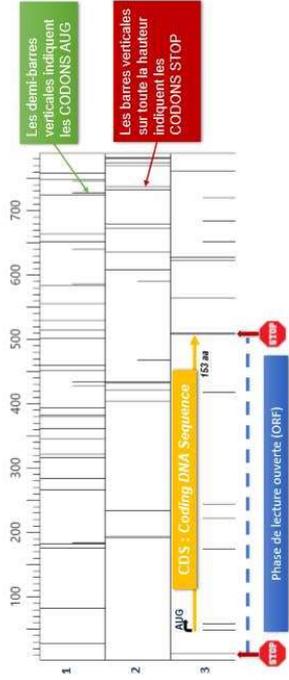
Définitions

Phase de lect ouverte (ORF, *Open Reading Frame*) : séquence de nt multiple de 3 bases qui épارة deux codons STOP (TAA, TAG, TGA).

Codons STOP très fréquents (3/64 codons), ms si fréq diminue ds une séq, signifie présence gène.

DS phase ouverte de lect, on retrouve une **séq codante** (CDS, *Coding Sequence*) qui débute par codon d'initiation de la trad (ATG) et se termine par codon STOP. Par abus de langage, CDS parfois appelée ORF.

En théorie, chaq séq nt a 6 phases de lect (3 pr chaq brn) dont 1 seule, sauf exception, est ouverte et contient une séq codante. ORF significative ↔ capacité à coder protéine. Potentiellement si mutation non multiple de 3 bases : cadre de lect peut devenir fermé.



Caractéristiques génome humain

20 000 gènes codant protéines, dt exons occupent 1,2% du génome. Taille très variable : moins de 1 kb pour gène SRY à + de 2 Mb pr DMD.

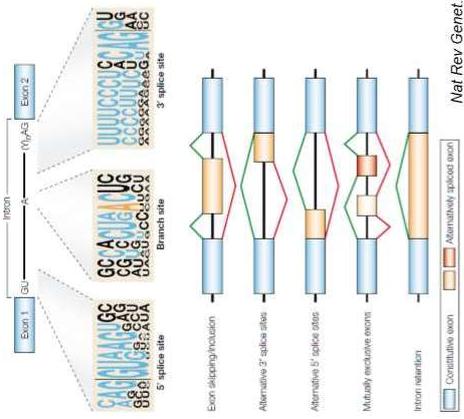
Gène humain moyen :

Taille	Nb exons	Taille exon	Taille intron	5'UTR	3'UTR	Taille séq codante
27 kb	9	145 pnb	3 365 pnb	300 pnb	770 pnb	1341 pnb (447 AA)

Plupart des gènes sont morcelés en exons (220 000 exons) séparés par des introns, dt nb très variable d'un gène à l'autre. Ex : Titine, 363 exons (+ de 35 000 AA).

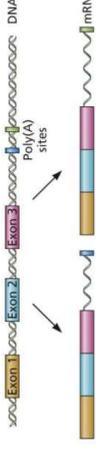
+ de 95% des gènes st soumis à transcription, épissage, maturation 3' alternatifs.

EPISSAGE ALTERNATIFS



Epissage alternatif

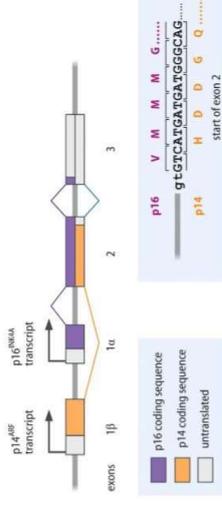
Si en retirant: exon, cadre de lect inchangé, p é retiré et donner prot. Expression gènes, exons conservés = interaction fact transcription (protéiques) – promoteur(s) (acides nucléiques). Chromatine doit é ouverte → séq active.



→ Exon maintenu ou non ds séq pot épissage du à interaction ac protéines.

Epissage lié à **protéines spécifiques** (si présence, exon maintenu, retiré sinon) + enhancers, silencers d'épissage. → Fonctionnalité séq exons va au-delà traduction.

Ex : **CDKN2A** : 3 exons, 2 promoteurs (1β et 1α) exon 2 = 1 des seuls pouvant é lu ds plusieurs cadres de lect → 2 prot (p14-ARF et p16-INK4A)



Gènes peuvent é répétés en **cluster**. +/- dispersés ds génome.

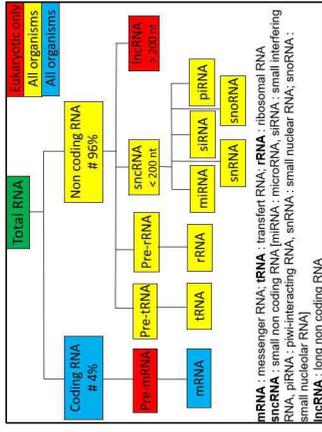
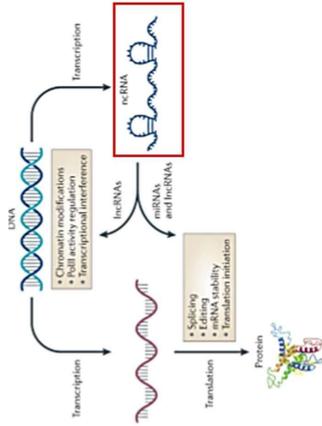
Gènes homologues : même origine **phylogénétique**, càd issus séq ancestrale commue. Notion d'homologie non quantifiable (contrairement à l'identité ou similarité de séq nt).

Gènes orthologues : gènes ayant évolué à partir même gène ancestral suite événements de **spéciation**. Fonction des gènes orthologues = svt conservée au cours évolution.

Gènes paralogues : gènes issus d'événements de **duplication** au sein génome. Un ou plusieurs paralogues peuvent subir modif de séq, et fonction peut devenir différente. Peut également perdre fonction et devenir pseudogène inactif.

Présentation des gènes produisant des ncRNA

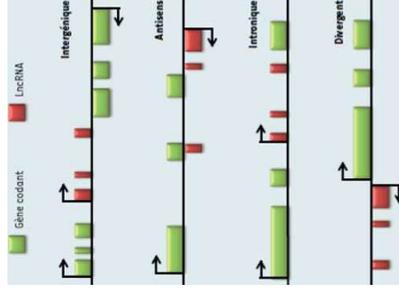
PROTEIN-CODING GENES	19 965
RNA GENES	25 485
long ncRNA	17 910
small ncRNA	7 576



miRNA : messenger RNA ; rRNA : ribosomal RNA ; tRNA : transfer RNA ; snRNA : small nuclear RNA ; snoRNA : small nucleolar RNA ; piRNA : piwi-interaction RNA ; siRNA : small interfering RNA ; pre-tRNA : pre-transfer RNA ; pre-rRNA : pre-ribosomal RNA ; lincRNA : long non coding RNA

Qd gène codant : ARNm non structuré car constitue un interm, passage, dt obj = protéine (struct II,...). Si non codant : ARN a fonction, définie par sa struct.

Découverte fondam **ARN non codant** ds expression gènes → prix Nobel 2006 *Découverte de l'extinction de l'expression des gènes par des ARN interférents* (Andrew Fire et Craig Mello).



Grds ARNnc (lincRNA) st définis par position relative aux gènes codants à proximité.

- **lincRNA intergéniques** (lincRNA) = localisés ds réion non annotée du génome.
- **lincRNA antisens** = transcrits ds direction opposée d'un gène codant lui étant associé.
- **lincRNA introniques** = contenus ds intron d'un gène codant.
- **lincRNA divergents** = transcrits de façon divergente au promoteur d'un gène codant.

Caractéristiques génome humain

25 000 gènes produisant ARN non codants.

- **NCRNA classiques** : tRNA, rRNA
- **Small ncRNA** (< 200 nt) : **snRNA** (small nuclear RNA), **snoRNA** (small nucleolar RNA), **siRNA** (small interfering RNA), **miRNA** (microRNA), **piRNA** (piwi-interaction RNA)
- **Long ncRNA** (> 200 nt) : intergéniques, introniques, antisens, divergents.

Présentation des pseudogènes

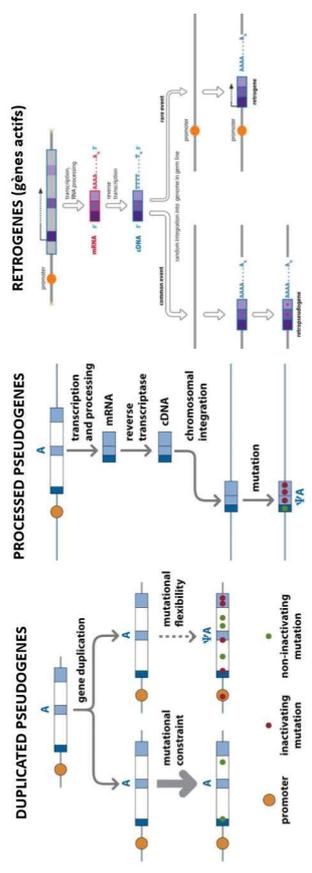
PSEUDOGENES	14 749
Processed	10 668
Unprocessed	3 556
Other	525

Environ 15 000.

Duplicated pseudogenes : gènes formés par **duplication** d'un gène parent fonctionnel codant une protéine, suivi d'une accumulation de mutations inactivatrices.

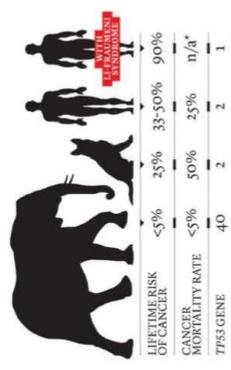
Processed pseudogenes : pseudogène créé par **rétrotransposition** de l'ARNm d'un gène parent fonctionnel codant protéine, suite d'une accumulation de mutations inactivatrices.

→ variant ds promoteur, accumulation mutation : gène devient non fonctionnel. Peut perturber expression d'un autre gène. Pseudogène peut 'intégrer ds génome, ms généralement pas le promoteur → devient inactif. Si s'insère à proxim d'un promoteur, peut devenir actif.

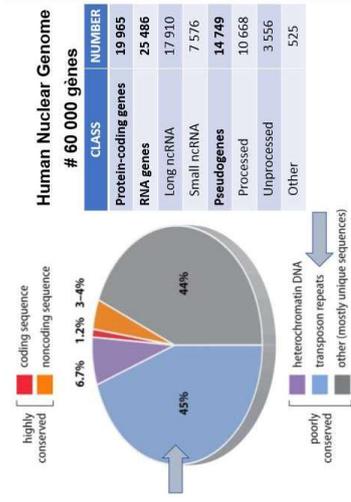


Paradoxe de Peto

Au niv des espèce, incidence du cancer non corrélée ac nb de cell de l'organisme (taille). Or, si proba de carcinogénèse était cste pr 1 cell donnée, on devrait observer incidence cancer en corrélation ac taille animal.



Présentation des séquences répétées dispersées : The Human Mobilome



TRANSPOSONS ADN

1ère fê prix Nobel seul, pour découverte des éléments mobiles : Barbara McClintock.

- Environ 3% du génome humain
- Inactifs ds génome humain depuis # 40 M d'années (n'existent plus fonctionnellement ds génome hum).
- Code une transposase (cap reconnaître séq répétées particulières : ITR pr transporter transposon)
- Encadrés par des séquences répétées inversées (ITR : « Inverted Terminal Repeats »)
- Éléments mobiles via un intermédiaire ADN selon deux mécanismes possibles : « COUPER/COLLER » (Transposition conservative, la + fréq) ou « COPIER/COLLER » (Transposition répliative). Selon les transposons, un mécanisme ou l'autre, ou les deux sont employés.

Rôle fondam ds évolution hum : **RAG1 et RAG2** issus transposon apparu ds génome il y a 500 M d'années. Réarrangement gènes Ig, aussi import ds évolution et complexification organismes, que endosymbiose bact (mitochondries, chloroplastes).

Sleeping beauty : transposon tel qu'il était il y a 40 M d'années, prod par synth chimique puis introduit ds cell humaine. (de même que piggy back).

RETROTRANSPOSONS

Mobiles par un intermédiaire ARN selon mécanisme « COPIER (transcription)/COLLER ». Éléments d'ADN reste où il est ; très invasif.

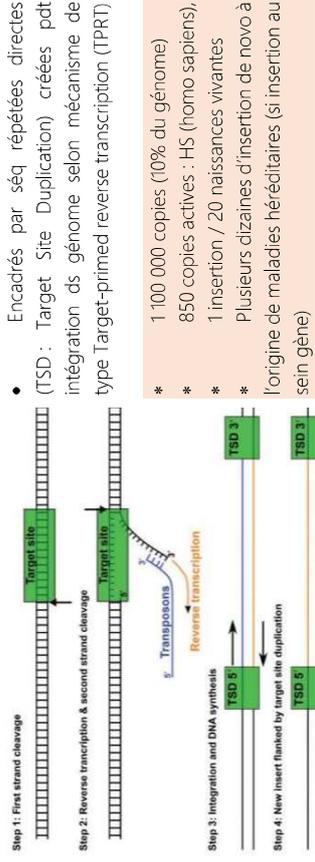
« Rétro » = rétrotransposase (ARN → cDNA)

Non autonomes, passifs

→ SINEs (Short (100 nt) Interspersed Elements)

Séquences Alu : dérivant du gène produisant ARN 7SL, cad ARN constitutif de ribonucléoprot SRP (Signal Recognition Particule), impliquée ds adressage prot destinées à è sécrétées.

- Gène ayant fait objet de réarrangements, duplications.
- Petit AARN transcrit par ARN pol III (promoteur interne) et pas d'ORF
- + de 1 M de copies (10% du génome) – régions GC riches du génome humain – bandes chromosomiques R
- Séq consensus de 280 pdb (200 familles)
- Spécifiques des primates et actifs depuis + de 65 M d'années



- * 1 100 000 copies (10% du génome)
- * 850 copies actives : HS (homo sapiens),
- * 1 insertion / 20 naissances vivantes
- * Plusieurs dizaines d'insertion de novo à l'origine de maladies héréditaires (si insertion au sein gène)

	14 749
PSEUDOGENES	
Processed	10 668
Unprocessed	3 556
Other	525

Rétroépélogènes

Autonomes

Non LTR

LTR ; séq nt caractéristique des extrémités des rétrovirus. Produite par transcription inverse de l'ARN du virus et se retrouve dupliquée aux deux extrémités de l'ADN ainsi synthétisé. Elle participe à l'intégration des éléments viraux ds génome de l'hôte. Après intégration dans l'ADN, le génome proviral se retrouve ainsi encadré par deux LTR, en 5' et en 3'. Le LTR 5' contient le promoteur qui assure la transcription de l'ARN rétrovirus.

→ LINE (Long Interspersed Element)

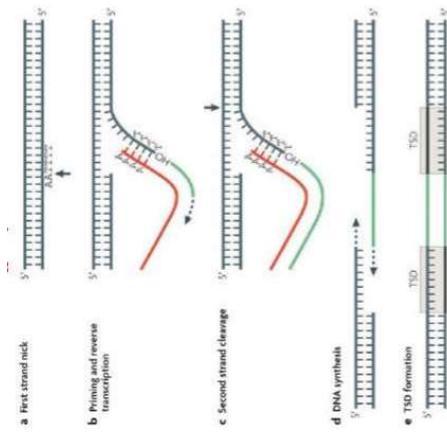
Taille de l'ordre de plusieurs milliers de paires de bases (~ 21% du génome humain)
Au sein des LINEs, il est possible d'individualiser les séquences L1s.

L1s :

- 6.1 kb (élément complet)
- 500 000 copies, la plupart incomplètes et tronquées en 5' (17% du génome humain)
- Régions AT riche – bandes G
- 40 à 50 éléments actifs (6 hot L1s)
- Promoteur Pol II

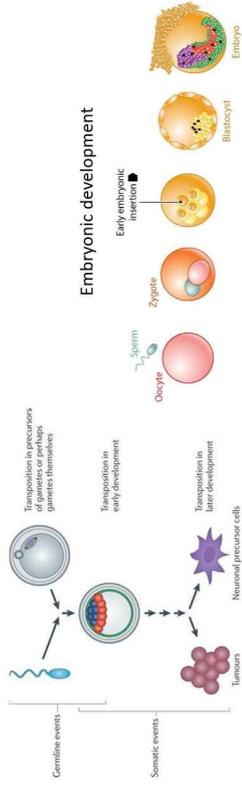
- Spécifiques des mammifères et actifs depuis ~150 millions d'années
- Trois ORFs :
 - ORF1 (endonuclease) et 2 (reverse transcriptase) communes aux mammifères
 - ORF0 antisens spécifique des primates
- Encadrées par des séquences répétées directes TSD (« Target Site Duplication »)
- 1 insertion / 100 naissances vivantes (insertions de novo à orig. maladies héréditaires)

Target-primed reverse transcription (TPRT)

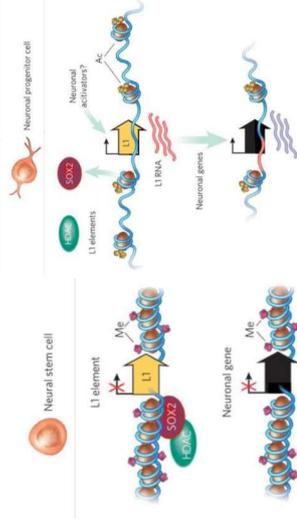


(a) The L1 endonuclease (EN) cleaves the first strand of target DNA
 (b) The free 3' hydroxyl (OH) generated by the nick is used to prime reverse transcription of retrotransposon RNA (red) by the L1 reverse transcriptase
 (c) The second strand of the target DNA is cleaved
 (d) and used to prime second strand synthesis
 (e) Hallmarks of the integration process include frequent 5' truncations, the presence of an oligo(dA)-rich tail at the 3' end and target site duplications (TSDs) of less than 20 base pairs in length
 → Il existe tjrs seq répétées ds transposons (mécanisme de réparation implique duplications de bases)

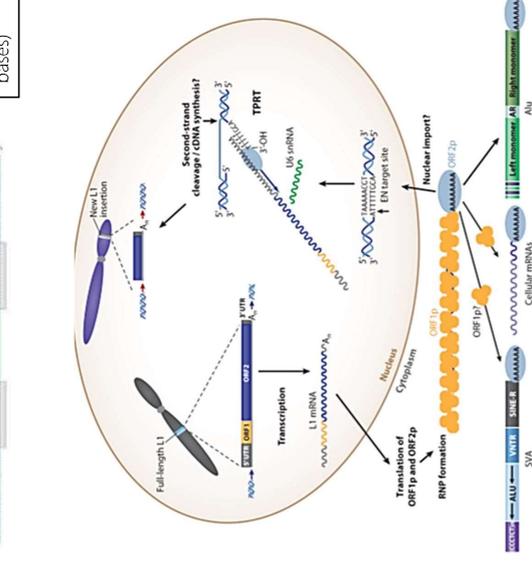
GERMLINE AND SOMATIC INSERTIONS



Chercheurs ont montré que plasticité neuronale était due à capacité à faire bouger spécifiquement ds neurones les séquences LINE



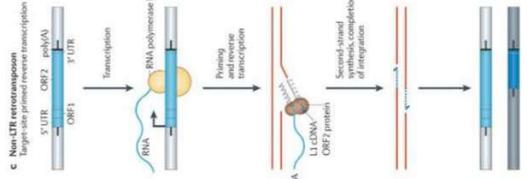
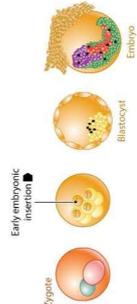
Neurone dt seq LINE bloquées en fermant chromatine (élément LINE non transcrits)
 Puis différenciation : ouvert chromatine → transcription seq LINE → insertion gènes ds génome
 → transposition éléments LINE = moteur ds diversité neurones



Réverse transcriptase : enzyme pas très processive, s'arrête svr encours de route → la plupart des L1 non fonctionnels car tronqués en 5' → Séquences LINE svr éteintes.

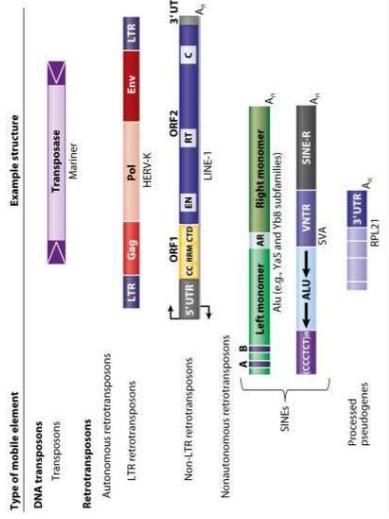
Éléments pouvant bouger ds cellules germinales :

Embryonic development



* **L1s et Alu** : # 750 millions bases.
 * Au cours des 6 derniers millions d'années, le génome humain a accumulé 2000 L1s, 7000 Alu et 1000 SVAs soit + de 8 millions de bases.
 * Plus d'une centaine de maladies héréditaires pour lesquelles le variant pathogène est une insertion de L1, Alu ou SVA ont été identifiées.
 * Les éléments transposables :

- contribuent au polymorphisme du génome humain (présence ou absence) et font partie des variants structuraux (MEI : Mobile Element Insertion).
- constituent un élément majeur de l'évolution des génomes.
- participent à la régulation expression des gènes.



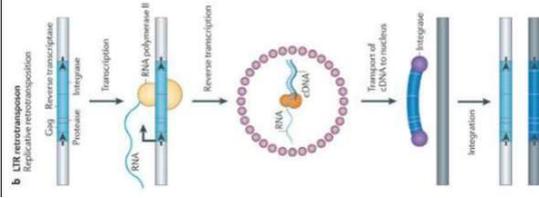
LTR rétrotransposons

HERV Human Endogenous RetroVirus

- * 8% du génome humain
- * Le plus souvent LTR isolés
- * Vestiges d'infections rétrovirales anciennes des **gamètes** → potentielle transmission à descendance
- * Inactifs en tant que rétrovirus

Caractéristiques du génome humain

50% du génome est constitué de séquences répétées représentées en majorité par des éléments transposables (SINES, LINES, HERV)



Présentation des séquences répétées en tandem

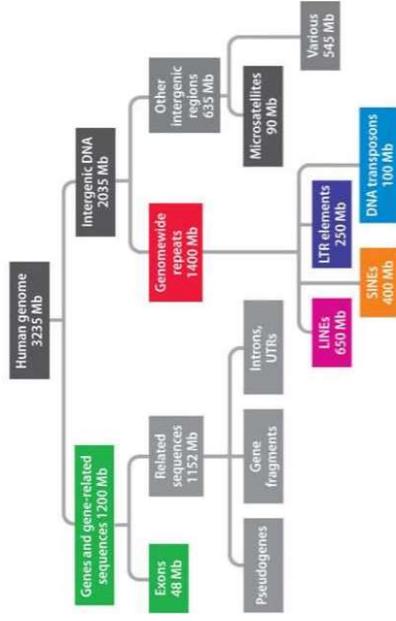
Séquences répétées le + svt concentrées dans des régions chromosomiques (télomères, centromères et régions subtélomériques).

Selon la taille de l'unité de répétition et la taille de la région concernée on distingue : ADN satellite très grds, svt centromères), ADN minisatellites (VNTRs), ADN microsatellites (STRs et Télomères : (TTAGGG)_n)

Hétérochromatine (6-7%) svt faite de séq répétées en tandem.

Certaines séquences intergéniques st très conservées (CNG, conserved non-genic sequences) et représentent 3-4% génome.

Composition of the human genome



Présentation des Polymorphismes

Polymorphisme génétique (du grec « poly » plusieurs et « morphé » forme) : coexistence de plusieurs allèles pour un gène ou un locus donné, dans 1 pop donnée.

Peuvent être explorés par analyse :

- des caractères (polymorphismes phénotypiques). Ex : couleur des yeux
- du produit protéique du gène (polymorphismes protéiques). Ex : groupes sanguins. Gène ABO : 5 exons non codants, 2 exons codants. Groupe sanguin O : délétion d'un G, motif cadre de lect → codon STOP.

Prix Nobel de Physiologie et de Médecine en 1980 : Découverte du Complexe Majeur d'histocompatibilité en

1958, Jean Dausset. Complexe HLA, polymorphisme tissulaire, sur chromosome 6.

- des chromosomes (polymorphismes chromosomiques) ex : tailles différentes
- de l'ADN (polymorphismes nucléotidiques)

Les polymorphismes chromosomiques et nucléotidiques n'ont pas besoin de séger dans des séquences codantes pour être détectés.

POLYMORPHISMES	
PROTEIQUES	
Groupes sanguins	1910
HLA	1970
NUCLEOTIDIQUES	
RFLPs	1975
VNTRs	1985
Minisatellites	1990
STRs	
Microsatellites	
SNPs/SNVs	# 2000
CNPs/CNVs	# 2000

1^{er} polymorphisme découvert : RFLPs (Restriction Fragment Length Polymorphisms)

Par Yuet Wai KAN. Utilisation d'enzymes de restriction (reconnaissant de courtes séquences et les coupant)

ACGCTGCCA **G** ACGTAGT

ACGCTGCCA **C** ACGTAGT

↓ PstI

ACGCTGCCA **G** ACGTAGT

ACGCTGCCA **C** ACGTAGT

Sir Alec Jeffreys (père des empreintes génétiques)

Séquences répétées en tandem d'une dizaine ou quelques dizaines de nucléotides et s'étendant sur une distance comprise entre 100 bp à 20 Kb. Principalement localisées dans les régions télomériques et subtélomériques.

Microsatellites ou STR (Short Tandem Repeats) découverts en 1989 par Michael Litt et Jeffrey Luty

- Séq répétées en tandem d'une unité de petite taille (1 à qq nt) s'étendant sur une centaine de nt
- Régulièrement répartis au niveau de l'euchromatine (# 1 STR tous les 30 Kb)
- Polymorphismes multi-alléliques (STRPs) en fonction du nb de répétitions
- Les STRPs dont l'unité de répétition n'est pas un multiple de 3 nucléotides sont absents des exons codants

Exemples de microsatellites :

- Répétitions mononucléotidiques : (A)_n, (T)_n, (G)_n ou (C)_n : 36.7 par Mpb
- Répétitions dinucléotidiques : (CA)_n : 27.7 par Mpb
- Répétitions trinucléotidiques : (AAT)_n : 4.1 par Mpb

(A) slippage causes insertion

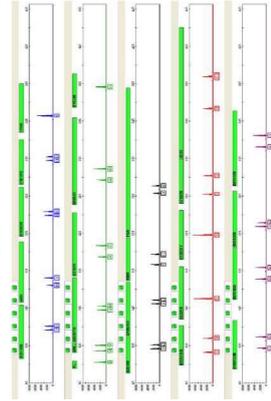


△ échelle européenne : définition d'un nombre précis et limité de polymorphismes multialléliques, pr identification (police scientifique).

→ Profils génétiques. Empreinte génétique obtenue = comparée au FNAEG (Fichier National Empreintes Génétiques), contenant 2,9 M d'empreintes en 2018. Mais, l'identification d'une pers par empreintes génétiques ne p é recherchée que :

1. Dans cadre de mesures d'enquêtes ou instructions diligentes pot procédure judiciaire
2. A des fins médicales ou de recherche scientifique
3. Aux fins d'établir, lorsqu'elle est inconnue, l'identité de personnes décédées

The probability of identity of GlobalFiler™ is 7.12x10⁻²⁶ (1 in 1.4x10²⁵)



FNAEG : Fichier national des empreintes génétiques

15 novembre 2006

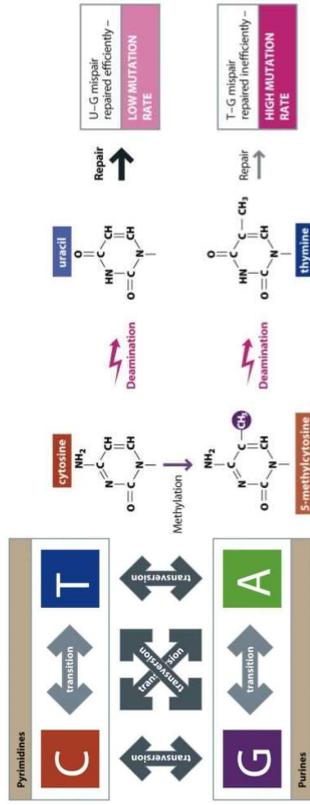
Le FNAEG sert à faciliter l'identification et la recherche des auteurs d'infractions à l'aide de leur profil génétique, et de personnes disparues à l'aide du profil génétique de leurs descendants ou de leurs ascendants.

Chiffres clés

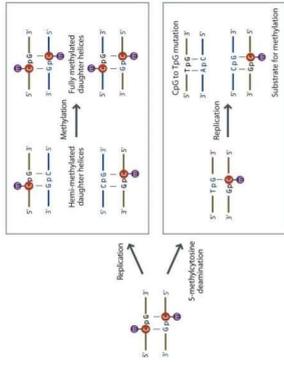
12 200 8
 2,9 millions de profils génétiques
 480 000 traces non identifiées

SNPs (Single Nucleotide Polymorphisms)

- Polymorphismes par substitutions nt : transitions sont plus fréquentes que les transversions



- Polymorphismes le plus souvent **bi-alléliques** présents dans au moins 1% de la population
- MAF : Fréquence de l'allèle le plus rare (Minor Allele Frequency) ≥ 1%
- Analysables individuellement (PCR) ou globalement (puces à ADN, Séquençage Sanger ou NGS)



SNP (Single Nucleotide Polymorphisms)

Altèle 1

5' ...TGCTGGGGCTA**T**CGCACATGATCTGATGTGCC...3'
 3' ...ACGACGGGAT**C**AGCGTGTACTAGACTACAGGG...5'

Altèle 2

5' ...TGCTGGGGCTA**T**CGCACATGATCTGATGTGCC...3'
 3' ...ACGACGGGAT**A**AGCGTGTACTAGACTACAGGG...5'

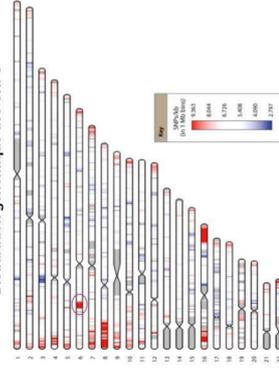
C se désamine en T

refSNP : rs1205

rs1205 [Homo sapiens]

ACCTCCAGTTTGGCTTCTCTCCA(C/T)AGTCTCTCCATGTCGACACAG
 Chromosome: 11
 Gene: GRP (Gene View)
 Functional Consequence: ur-variant 3 prime
 Validated by: 1000G
 Global MAF: T=0.3850

Distribution génomique des SNPs



HAPLOTYPE : succession des allèles d'un gène ou d'un locus sur une région chromosomique de petite taille.
 Du fait de l'absence de recombinaison, les associations d'allèles de plusieurs polymorphismes très proches sont transmis généralement en bloc, comme un seul marqueur.

SV (Structural Variants)

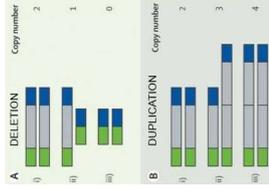
En fonction de la taille de la variation structurale on distingue :

- Les INDELS qui correspondent à des **insertions/délétions** de moins de 50 nucléotides (1 à 49)
- Les autres SV correspondent à des variations de plus de 50 nucléotides
 - ❖ CNVs
 - ❖ Grandes délétions
 - ❖ Inversions
 - ❖ MEI

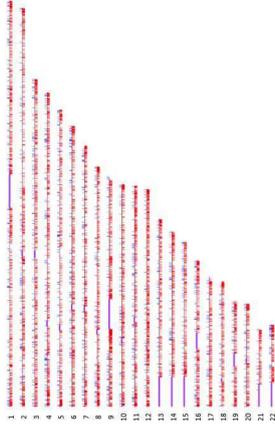
CNVs

- Variation du nombre de copies de segments d'ADN d'une taille de 50 à plus de 3 M nt et pouvant inclure des gènes (nb gènes variable d'un individu à l'autre).
- Polymorphismes bi-alléliques ou multialléliques
- Analysables par CGH array (ACPA pour Analyse chromosomique sur Puce à ADN) et par NGS (Next Generation Sequencing)

Diallelic locus (grey) and flanking loci (green and blue) with variation caused by (A) deletion and (B) duplication, each showing the locus with :
 (i) normal copy number,
 (ii) heterozygous modification,
 (iii) homozygous modification.



Distribution génomique des CNVs



25 ans après annonce séquençage génome (2015), article paru dans Nature : carte mondiale du génome, 2500 individus, 26 ethnies.

Un génome typique contient, par / à un génome de référence, 4 à 5 M. de variations. La plupart est des SNV et INDELS (99.9%). 2000 à 2500 SV (+ de 50 bases) affectant 20 M de bases :

- * 1 000 large deletions
- * 10 inversions
- * 160 Copy Number Variants (CNVs)
- * 1 094 Mobile Elements Insertions (MEIs) : 915 Alu, 128 L1 and 51 SVA

Median autosomal variant sites per human nuclear genome (Europe - 503 samples)

Sequence variations	Variations sites
SNVs	3,63 M
INDELS	546 K
Large deletions	939
Inversions	10
CNVs	157
MEI Alu	919
MEI L1	123
MEI SVA	53

Comparaison génome Nord-Européen de réf avec les génomes de 500 nord-européens.
 → où sont ces variations :

Sequence variations	Variations sites
Exon	non synonymous 10 200
	synonymous 11 200
	TOTAL 21 400
Intron	1 680 000
Known regulatory regions	UTR 30 000
	Promoter 82 200
	Insulator 57 700
	Enhancer 288 000
TFBSs	749
TOTAL	458 649

Caractéristiques du génome humain

20 000 gènes codant des protéines dont les exons occupent 1.2 % du génome

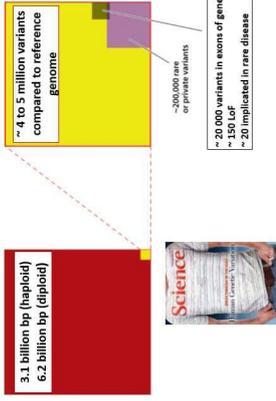
- ✓ Taille très variable : moins de 1 kb pour le gène SRY à plus de 2 Mb pour le gène DMD
- ✓ Morcelés en exons (# 220 000) séparés par des introns dont le nombre est très variable d'un gène à l'autre
- ✓ Plus de 95% des gènes sont soumis à transcription, épissage ou maturation 3' alternatifs.

25 000 gènes produisant des ARN non codants

- ✓ ncRNA classiques : tRNA, rRNA
- ✓ Small ncRNA (< 200 nt) : snRNA (small nuclear RNA), snoRNA (small nucleolar RNA), siRNA (small interfering RNA), miRNA (microRNA), piRNA (piwi-interacting RNA)
- ✓ Long ncRNA (> 200 nt) : intronic, exonic, overlapping and intergenic

50% du génome est constitué de séquences répétées représentées en majorité par des éléments transposables (SINEs, LINEs, HERV)

4 à 5 millions de variations comparativement à un génome de référence : en majorité des SNVs (3.5 millions) et des INDELS (500 000) et 2 000 à 2 500 SVs ≥ 50 bp affectant plus de 20 millions bp



Variabilité ++ du génome

GENETIQUE CLINIQUE	GENETIQUE DES POPULATIONS
- MUTATION : variation génomique associée à un effet pathogène - POLYMORPHISME : variation génomique a priori NON associée à un pathogène	- POLYMORPHISME : variation génétique présente à une fréquence $\geq 1\%$ dans la population générale. A noter que certaines variations à effet pathogène sont présentes à l'état hétérozygote dans la population à une fréquence $\geq 1\%$

MUTATION et POLYMORPHISME : Signification différente en fonction du contexte

→ on ne parle plus de mutation ou polymorphisme, mais de variation, selon l'échelle suivante :

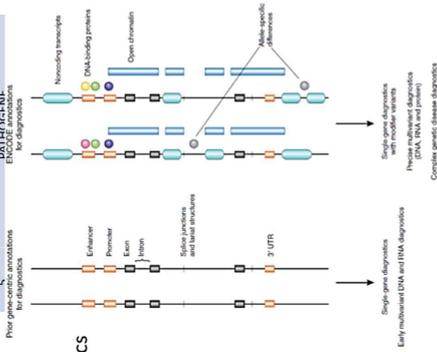
CLASSIFICATION DES VARIATIONS DE SEQUENCE

CLASSES	DESCRIPTION
1	BENIN BENIGN (B)
2	PROBABLEMENT BENIN LIKELY BENIGN (LB)
3	SIGNIFICATION INCONNUE UNCERTAIN SIGNIFICANCE (VUS)
4	PROBABLEMENT PATHOGENE LIKELY PATHOGENIC (LP)
5	PATHOGENE ENCODING AMINOACIDS for diagnostics

Encode Project : ENCYclopedia Of DNA Elements

The implications of :NCODE for diagnostics

Projet international d'étude du génome humain : cartographier séq génomiques + déf rôle de chacun de ces éléments.



lat Biotechnol 2012, 30:1064-5

Classification des maladies génétiques

Maladies héréditaires à transmission mendélienne : maladies dont l'apparition conditionnée par la mutation d'un seul gène. On parle également de « **maladie monogénique** ».

Maladies multifactorielles : maladies dont l'apparition est conditionnée par la présence de certaines combinaisons d'allèles de plusieurs gènes situés sur des locus différents, aucun n'étant indispensable par lui-même à l'apparition de la maladie. On parle d'hérédité complexe ou de « **maladie polygénique** ».

Maladies par aberration chromosomique : dues à une anomalie de nombre ou de structure des chromosomes.

Maladies « mitochondriales » : dues à une mutation génome mitochondrial (hérédité maternelle). Cytopathies mitochondriales peuvent être la conséq de mutations du génome nucléaire et du génome mitochondrial.

Maladie congénitale : présente à la naissance, cause génétique ou non. Par exemple, les malformations dues à la rubéole sont présentes à la naissance mais d'origine virale (virus de la rubéole).

Les maladies génétiques **peuvent être létales** avant la naissance, présentes à la naissance (congénitales) ou se déclarer plus tard. Beaucoup de maladies génétiques ne sont pas congénitales (absentes à la naissance).

Phénotopie : phénotype dû à une cause non héréditaire imitant un phénotype d'origine génétique.
Exemple : un cancer du sein sporadique survenant dans une famille prédisposée au cancer du sein, possible du fait de la grande fréquence de ces cas sporadiques dans la population générale (une femme sur dix développera un cancer du sein au cours de sa vie).

Incidence : nombre de nouveaux cas d'une maladie par unité de temps dans une population.
Prévalence : rapport du nombre de personnes atteintes à la population considérée.

Le **nom d'un gène codant une protéine** dérive le plus souvent de sa fonction biochimique ou du phénotype de la maladie. L'allèle responsable de la maladie est appelé délétère, morbide ou pathologique.

Le **symbole d'un gène** est indiqué sans indice ni exposant, en majuscules, en italique, sans ponctuation. Ex : HBB, gène codant la sous-unité (chaîne) β de l'hémoglobine humaine.

Les **fragments d'ADN** en copie unique ne correspondant pas à un gène sont désignés par la lettre D (pour DNA) suivi du numéro de chromosome de provenance puis d'un S (Single locus) et du numéro d'ordre de la séquence assignée au chromosome (numéro d'enregistrement). Ex : D4S20 correspond à la 20ème séquence attribuée au chromosome 4.

2 types d'hétérogénéité génétique :

L'**hétérogénéité allélique ou intralocus** : une maladie, quel que soit son mode héréditaire, peut être due à des mutations délétères différentes d'un même gène (mutations alléliques) → 1 maladie/plusieurs mutations délétères différentes d'un même gène.

Habituellement : **maladies récessives** et **dominantes** dites par haplo-insuffisance. On rapproche de cette hétérogénéité allélique le cas de mutations délétères d'un même gène (mutations alléliques), qui, selon qu'il s'agit d'une perte de fonction ou d'un gain de fonction, peuvent conduire à des phénotypes différents : un gène/plusieurs maladies différentes.

Il est à noter que dans ce cas, il s'agit de maladies différentes et non de formes cliniques différentes d'une même maladie (hétérogénéité clinique).

L'**hétérogénéité interfocus** se traduit par le fait qu'un phénotype apparemment identique peut être la conséquence de mutations délétères ségeant dans des gènes différents non alléliques : une maladie/plusieurs gènes différents (non alléliques).

Exemple : Hémophilie A (gène F8C – Xq28) et Hémophilie B (gène F9 Xq27)

Fréquence des cancers chez l'homme		Fréquence des cancers chez la femme	
1) prostate	50 430	1) sein	58 459
2) poumon	31 231	2) colorectal	20 120
3) colorectal	23 216	3) poumon	15 132
Mortalité par cancer chez l'homme		Fréquence des cancers chez la femme	
1) poumon	22 761	1) sein	12 146
2) colorectal	9 209	2) poumon	10 356
3) prostate	8 115	3) colorectal	7 908

En 2018, le nombre de personnes de 15 ans et plus vivantes et ayant eu un cancer au cours de leur vie est de l'ordre de **3,8 millions**.

Il existe une **prédispo** au cancer sein et /ou ovaire (syndrome sein/ovaire), de l'ordre de **5%**.

Transmission sur un mode **autosomique dominant** à **pénétrance incomplète**. Liée à un variant pathogène constitutionnel de **BRCA1** (chromosome 17) ou **BRCA2** (chromosome 13).

BRCA1 : identifié en 1994. Bras long du chromosome 7 (17q21), s'étend sur ~ 80 kb, 23 exons (22 codants) et code une protéine de 1 863 AA, impliquée ds mécanismes réparation ADN.

Fréquence variants pathogènes de **BRCA** = variable en fonction des origines géographiques. En F : **2 femmes sur 1 000** (1 f. sur 500 = hétérozygote [Nd]).

En F : 20 millions de femmes âgées de 20 à 70 ans, **40 000** environ porteuses de variants pathogènes de **BRCA**, à haut risque de développer la maladie.

Formes héréditaires familiales à distinguer des concentrations familiales fortuites (**phénotopies**).

Prédispo héréditaire suspectée :

- si fréquence ++ dans une même famille,
 - si touche des femmes jeunes (15% des cancers du sein de la femme de moins de 30 ans sont liés à une mutation du gène BRCA1),
 - si bilatéralité des tumeurs et association à un cancer de l'ovaire.
- Un cas de cancer du sein chez un homme est également très évocateur d'une prédisposition héréditaire.

→ Orientation vers consultation d'oncogénétique pr recueillir infos médicales, reconstituer hist perso et fam, construire arbre généalogique, estimer proba prédispo et prescrire ou non un test génétique BRCA. Pers porteuses d'un variant pathogène BRCA = prise en charge spécifique, surveillance et/ou chirurgie prophylactique, adaptée aux différents risques tumoraux associés à l'altération génétique identifiée.

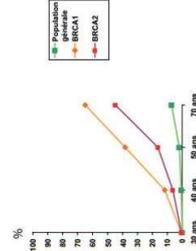
Présence variant pathogène constitutionnel de BRCA ↗ risque dév :

- * Cancer du sein à un âge précoce
- * Cancer sur sein controlatéral après diag du 1er cancer
- * Cancer de l'ovaire, essentiellement après 40 ans (risque variant selon gène touché et hist fam).

Risque cumulé de cancer du sein à 70 ans = 65 % pr femme [Nd] **BRCA1** (50 % si **BRCA2**) → risque relatif par / pop générale= 6 (5 pour gène **BRCA2**).

Risque cumulé de cancer des ovaires à 70 ans = 40% pr femme [Nd] **BRCA1** (15% si **BRCA2**) → risque relatif par / pop générale = 40 (18 pour le gène **BRCA2**).

H0 : variant pathogène de **BRCA1** = risque cumulé de cancer du sein de 5% à 70 ans (pénétrance très faible), contre 0% ds pop générale ; + sur risque cancer prostate (en particulier pour **BRCA2**).

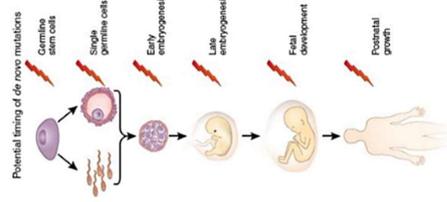


Variants pathogènes de novo

Patient ac maladie AD (P=100%) peut naître de parents sains et non porteurs d'un variant pathogène. Proportion variants pathogènes de novo = variable ms p é très élevée pr cert maladies.

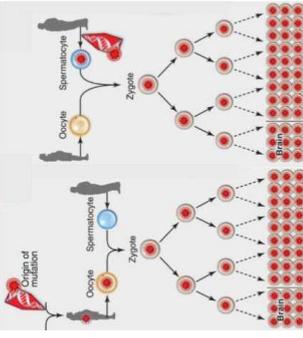
Si variant pathogène létal, tous les cas résultent de mutations de novo. Ds descendance du patient porteur du variant de novo, on retrouve les caractéristiques de l'hérédité AD.

→ Distinction **formes familiales** (+ d1 patient par fam) / **formes sporadiques** (1 seul patient ds fam).



Examen des caractéristiques génétiques d'une pers à des fins médicales consiste à analyser ses caractéristiques génétiques héritées (constitutionnelles) ou acquises à un stade précoce du développement prénatal.

Hypermutabilité du génome humain : tx de mutation = **10⁻⁸** par nt et par génération.



Variants **pathogènes de novo** apparus ds **gam parentaux**

Ex : Achondroplasie, nanisme disharmonieux (MIM 100800), 1 naiss sur 25 000, P = 100%.

100810
ACHONDROPLASIA; ACH

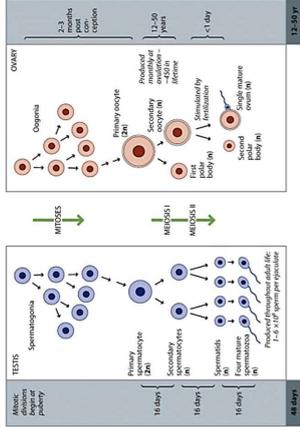
Phenotype-Genes Relationships			
Location	Phenotype	Gene/Allele	Gene/Locus
4p16.3	Achondroplasia	FGFR3	16p11.2
	Hyperostosis hyperostiformis	FGFR3	16p11.2
	Hyperostosis hyperostiformis	FGFR3	16p11.2

+ 80% des patients ont parents de taille normale (cas sporadiques) et descendance présente caractères habituels de la transmission AD. → **Majo des cas = sporadiques** (mutations de novo)

Gène identifié en 1994 : bras court du chromosome 4 (4p16.3) et code un récepteur pour un facteur de croissance des fibroblastes (FGFR3) exprimé dans le cartilage en croissance.

Pratiquement tous hétérozygotes : mutation faux sens p.Gly380Arg.

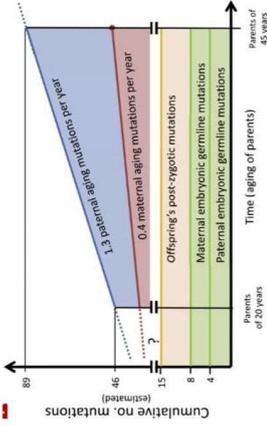
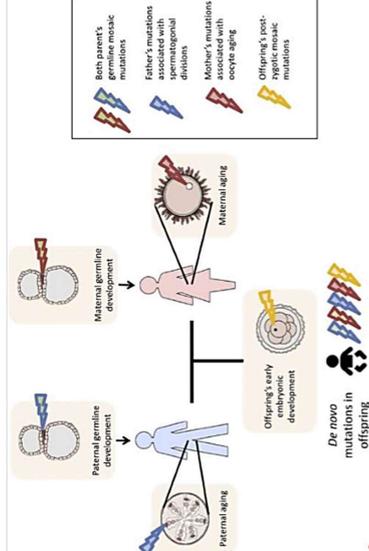
Variant pathogène = transmis par père dt âge au moment de la procréation est élevé, apparu ds cell germinales du père (gamètes mutés ont un avantage sélectif). Le risque x10 pour un père de 50 ans.



Nb réplifications chromos ds cell germ très différent entre hō et fē. Fē = 23 réplifications (quelque soit âge) car ovogénèse = uniquement pdt vie foetale. Au contraire : spermatogonies souches = 30 réplifications (du stade embryo jusqu'à puberté), puis 5 pdt cyclés spermiogénèse (23 cycles par an).

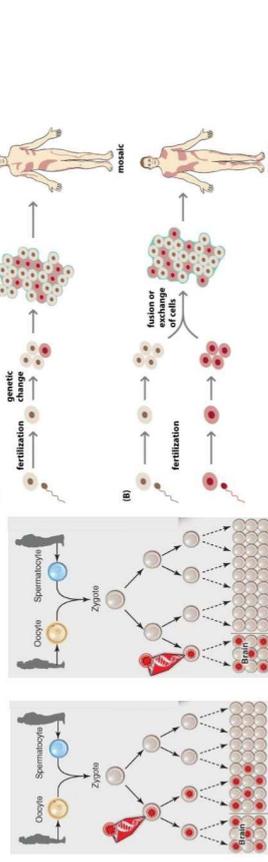
→ Nb réplication \propto ac âge cz hō et fav survvenue mutation de novo (âge moy père en F = 33,6 en 2018 contre 30,6 pr fē).

Overview of Distinct Classes of Germine De Novo Mutations (DNMs)



Variants pathogènes de novo post-zygotiques

→ 1ers stades développement zygote. Influence différente en fonction caract. précoc de apparition du variant. Obtention mosaïques cell saines/mutées. 3 types : somatic, germline et gonosomal (les 2).



Mosaïcisme

→ présence, cz même individu, de pop cell génétiquement distinctes, dérivant d'un même zygote.

- Mosaïcisme germlinal : double pop cell germinales, cert porteuses d'un variant pathogène. Parent porteur peut transmettre variant à sa descendance. Si absent des cell somatiques : maladie non exprimée cz parent porteur, mais pourra être transmise. → Parents indemnes peuvent avoir enfant malade.
- Mosaïcisme somatique : conséq variant pathogène ds cell somatique. Elle se divise ensuite → cell filles identiques, porteuses de la même mutation. Variants somatiques non hérités des parents, ni transmis à la descendance (sauf variants pathogènes post-zygotiques touchant lignées germ et somatique).

Les données récentes du NGS (WES) : ts nos tissus = mosaïques pour différents variants pathogènes.

Elephant Man (Joseph Merrick) : syndrome de Protée (MIM 176920 et ORPHA 744), maladie génétique exceptionnelle due à une mutation somatique en mosaïque activatrice de l'oncogène AKT1 qui est responsable d'une croissance excessive, asymétrique et déformante de multiples tissus notamment osseux, conjonctif et adipeux.

Expressivité clinique variable

Ds cert maladies dominantes, degré d'atteinte des malades varie (variabilité interfamiliale) et même au sein d'une même famille (variabilité intrafamiliale).

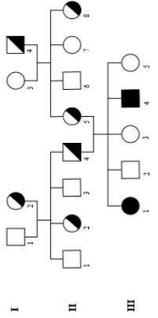
Dépend de fact :

- génétiques propres à l'individu : gènes modificateurs
- environnement.

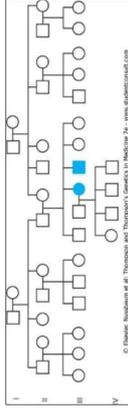
Anticipation = cas particulier d'expressivité variable : au cours des générations successives, l'âge d'apparition de la maladie est de plus en plus précoce et s'accompagne d'une accentuation de la gravité.

Maladies autosomiques récessives

- * Héritéité liée aux **autosomes (transmission horizontale)**
- * Sujets atteints **homozygotes** ou **hétérozygotes composites** pour l'allèle délétère. Parents porteurs sains; hétérozygotes pour allèle délétère
- * Hô et fé : même proba é atteints
- * Risque de récurrence après naiss d'un enfant atteint = 1/4.
- * La descendance d'un sujet atteint est habituellement normale. Cependant, la **consanguinité** > incidence de la maladie d'autant plus que la fréquence de l'allèle délétère est faible.



Abre généalogique d'une famille atteinte de mucoviscidose



Modèle de Hardy-Weinberg

Fréq alléliques et génotypiques d'une pop théorique idéale (unions au hasard, panmixie) restent stables de génération en génération → **équilibre de Hardy-Weinberg** et il existe une relation simple entre fréq alléliques et génotypiques.

Pour un locus autosomique à 2 allèles N (sauvage) et d (délétère), de fréquence p et q respectivement (avec p+q=1), les fréquences génotypiques sont égales à :

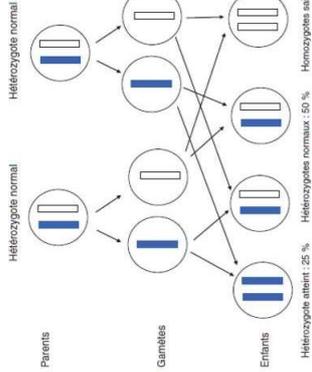
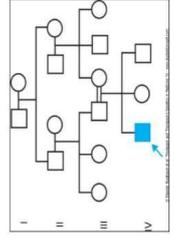
$f(NN) = p^2$	$f(Nd) = 2pq \approx 2q$	$f(dd) = q^2$
Homozygotes sains	Hétérozygotes sains	Homozygotes malades

Avec $p^2 + q^2 + 2pq = 1$ et $p \approx 1$

La fréquence des couples à risque est égale à $2pq \approx 2pq (4q^2 p^2)$

Maladies	Fréquence de la maladie (q²)	Fréquence de l'allèle délétère (q)	Fréquence des hétérozygotes sains
Mucoviscidose	~ 1/4 600	~ 1/68	~ 1/34
Galactosémie congénitale	1/40 000	1/200	1/100

Distrib des all peut s'écarter de manière significative du modèle de H-W qd il existe en particulier une proportion significative d'unions consanguines.



Abre généalogique d'une famille atteinte de mucoviscidose

Notion de consanguinité

Coefficient de parenté entre deux individus : proportion des gènes qu'ils partagent du fait de l'existence d'ancêtres communs identifiés.

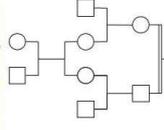
Le **coefficient de consanguinité f** d'un individu = proportion de ses loci portés à l'état homozygote du fait du lien de parenté de ses parents. Il est égal à la moitié du coefficient de parenté des parents.
 Coefficient de consanguinité correspond à la probabilité que les deux allèles d'un locus quelconque soient identiques par ascendance, c'est à dire occupés par deux allèles identiques provenant d'un même allèle d'un ancêtre commun à ses deux parents.

$f = (1/2)^{i+j} [(1/2) + (f_w/2)]$
 i et j : nb de génération séparant respectivement le père et la mère de leur ancêtre commun.
 f_w est l'éventuel coefficient de consanguinité de l'ancêtre

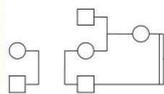
→ $f = (1/2)^{i+j+1}$

UNIONS

Entre cousins germains



Entre un oncle et sa nièce



$f = (1/2)^{2+2+1} = 1/32$
 $f = (1/2)^{2+2+1} = 1/32$

Les unions entre apparentés fav apparition de maladies récessives si ancêtres communs sont [Nd]. **L'effet de la consanguinité** est d'autant plus important, en fréquences relatives, que les **maladies** récessives sont **rares** (fréquence de l'allèle délétère q faible).

Union entre	Nbr apparentés (géméologiques)	Cousins germains	Accroissement du risque relatif
Coefficient de consanguinité f	0	1/16	
Mucoviscidose q = ~ 1/68	1/34 x 1/34 x 1/4 ~ 1/4 600	q² + fq 1/4 600 + 1/1 088 = 1/860	x ~ 2.5
Galactosémie congénitale q = 1/200	1/100 x 1/100 x 1/4 1/40 000	q² + fq 1/40 000 + 1/3 200 = 1/3 000	x 14

Notion d'effet fondateur

Création d'une **nvelle pop** à partir d'un **nombre restreint d'individus** (fondateurs) provenant d'une population existante. Par simple effet d'échantillonnage, la composition génétique de la nouvelle population peut fortement différer de celle de la population d'origine. Cette variation d'échantillonnage induite par la scission de population est appelée « effet fondateur ».

Les variants pathogènes à l'origine de nb maladies mendéliennes ont été identifiés ds pop hum dites à effet fondateur car fondées par un grpe d'individ d'effectif restreint et restés **isolés** pendant plusieurs générations. Différents arguments suggèrent l'existence, dans ces populations, d'un ancêtre commun à la majorité voire l'ensemble des malades, ancêtre qui portait une mutation délétère et la transmise à sa descendance.

Les malades partagent alors une région du génome autour du variant pathogène ce qui facilite l'**identification des gènes impliqués** dans ces maladies.

Ex : mucoviscidose : + de 1700 variants pathogènes différents, 1 variant = 70% (délétion d'une Phe en 508). Ce variant est apparu 1 seule fois il y a 20 000 à 50 000 ans, ds pop mère à orig du N-Europe. Cette mutation de CFTR a conféré avantage face au choléra → fréq a augmenté.

Application de la loi de Hardy-Weinberg aux maladies autosomiques dominantes

Pour une maladie autosomique dominante déterminée par un allèle a de fréquence q, la probabilité qu'un individu soit atteint par cette maladie correspond à la fréquence de ce phénotype dans la population, soit $q^2 + 2pq$.

Maladies	Fréquences de la maladie (q²+2pq)	Fréquences de l'allèle délétère (q)	Nombre d'individus atteints pour un homozygote malade (2pq/2)
Hypercholestérolémie familiale	1/500	1/1000	2 000
Maladie de Huntington	1/15 000	1/30 000	60 000

Si les **homozygotes** sont **rare** voire **absents (allèles létaux)**, la fréquence de l'allèle délétère est égale à la moitié de la fréquence de la maladie.

Dépistage Néonatal

« Dépistage » svt utilisé à tort dans le sens de « diagnostic ».
Dépistage : action de **S public** organisée par une instance et proposée à une **large cohorte d'individus**, définie selon le type de dépistage.
 Test diagnostique est prescrit à un patient présentant un pb de S par un médecin au cours d'une consultation **individuelle**.
Dépistage néonatal = dépistage de masse destiné à toucher **tous les nouveau-nés** d'un pays dans le but de détecter affections, le + svt héréditaires, à des fins de **prévention II** et avec bénéfice direct.

Critères OMS de Wilson et Jungner

- 1) Pb important de **S publique**
- 2) **Histoire** de la maladie = comprise
- 3) **Traitement** efficace
- 4) Effectué à un stade **pré-symptomatique**
- 5) Réalisable par une **méthode fiable**
- 6) **Accepté** de la population
- 7) Accompagné d'un **protocole thérapeutique** précis
- 8) Organisation du **diag et traitement des malades**
- 9) Bon rapport **coût-bénéfice**
- 10) Être **pérenne**



Arrêté du 22/02/2018 relatif à l'organisation du programme national de dépistage néonatal recourant à des examens de biologie médicale

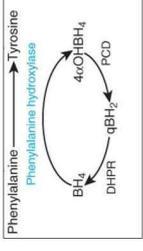
Article 1	Article 2	Article 3	Article 4
Le dépistage néonatal recourant à des examens de bio méd constitue un prog de S national . Objectif : la prévention secondaire de maladies à forte morbimortalité , dt les manifestations et complications surviennent dès l'âge des 1ères sem de vie et peuvent être prévenues ou minimisées par un traitement adapté si ce dernier est débuté très précocement.	Les examens de biologie médicale du programme de dépistage néonatal sont réalisés sur un échantillon de sang total prélevé sur brûlard au plus tôt 48 heures après la naissance, au mieux à 72 heures .	La réalisation du dépistage néonatal est proposée à titre gratuit pour tous les nouveau-nés .	Le programme de dépistage néonatal est mis en œuvre dans chaque région par le centre régional de dépistage néonatal (CRDN) .

Le programme de dépistage néonatal concerne :

- Pour l'ensemble des nouveau-nés :
 - la **phénylcétonurie** ;
 - l'**hypothyroïdie congénitale** ;
 - la **mucoviscidose**.
- Pour les nouveau-nés nés à partir de 32 semaines d'aménorrhée : l'**hyperplasie congénitale des surrénales**.
- Pour les nouveau-nés présentant un **risque particulier** de développer la maladie : la **drépanocytose**.

❖ **Phénylcétonurie**

Autosomique récessive, incidence à la naissance est ~ 1/15.000.
 Déficit en **Phe hydroxylase (PAH)**, assurant ds foie l'hydroxylation de la Phe en Tyr. Chez l'enfant non traité dans la période néonatale, l'hyperphénylalaninémie entraîne une encéphalopathie progressive sévère.



Dépistage néonatal (mesure de la phénylalaninémie - **test de Guthrie**) → régime appauvri en Phe maintenu pdt maturation cérébrale (jusqu'à 8 ans).

Fé atteintes doivent en pré-conceptionnel et pdt tte la grossesse reprendre ce régime car risque majeur d'embryo-foetopathie (retard mental, microcéphalie et cardiopathie congénitale) liée à l'effet tératogène de l'hyperphénylalaninémie.

❖ **Hypothyroïdie congénitale**

Principale cause de déficience intellectuelle évitable à la naiss. Prévalence à naiss = 1/3 500.
 Etiologies variées (rarement génétiques) et dépistage néonatal (mesure de la TSH) permet d'instaurer **traitement préventif** (administration à vie de L-thyroxine) qui évite l'installation des lésions neurologiques et permet dévt normal à condition d'être commencé très tôt dès le résultat du dépistage.

❖ **Mucoviscidose**

Dépistage néonatal : dosage de la trypsine immunoréactive. Si valeurs élevées : R mutations fréq du gène CFTR. Le diagnostic confirmé par un test de la sueur.

En cas de diagnostic +, l'enfant est adressé à un **Centre de Ressources et de Compétences pour la Mucoviscidose (CRCM)** pour prise en charge globale multidisciplinaire (pédiatre spécialisé, infirmière puéricultrice, kinésithérapeute, diététicienne, psychologue et biologiste généticien).

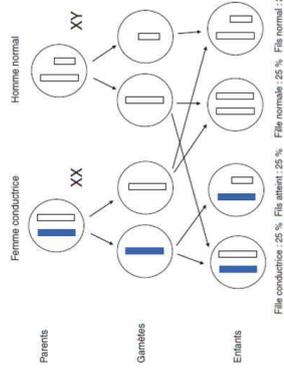
1 des caract dépistage néonatal = recours à un **examen des caractéristiques génétiques**. → CCNE a émis un avis sur pb que posent les applications des progrès de l'étude du génome en matière de dépistage (<http://www.ccne-ethique.fr/francais/pdf/avis097.pdf>).

❖ **Hyperplasie congénitale des surrénales** : déficit en 21-hydroxylase

❖ **Drépanocytose**

Maladies transmises selon mode récessif lié à l'X

- * Gènes impliqués dans les maladies RLX sur le chromos X.
- * Sujets atteints le + svt des garçons (**hémizygotés**).
- * **Fé hétérozygotés** habituellement **asymptomatiques** ms cert peuvent exprimer la maladie avec sévérité variable (inactivation du chromosome X).
- * Titres les **filles d'un homme atteint** sont **vectrices**, tous les garçons sont **sains** (pas de transmission père/fils)
- * **Fils d'une fé hétérozygote** risque = **50%** d'hériter de la maladie.



<p>Vectrice obligatoire</p> <ul style="list-style-type: none"> - fille d'un père malade - mère d'un seul garçon malade avec antécédents familiaux 	<p>Vectrice potentielle</p> <ul style="list-style-type: none"> - mère d'un garçon malade sans antécédents familiaux - femme ayant des antécédents familiaux mais n'ayant pas encore donné naissance à un garçon malade.
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Mutations de novo ou Mutations récentes ou Néomutations

Maladies RLX : mutations ont la particularité d'avoir **effet récessif** **cz** **fé dominant** **cz** **h** **hémizygotés**.
 Fréquence des individus atteints = **q** chez **l'homme** et à **q²** chez les **femmes**. Comme q très petit, filles rarement atteintes → surtt: vectrices ac fréq 2pq soit 2q si q est petit.

Haldane (1935) : **proportion des mutations récentes** parmi garçons atteints; m, calculable selon l'équation suivante : **m = μs / (2μ + v)**
s = désavantage sélectif (perte de fécondité) induit chez l'individu atteint. Si s=0, il n'y a pas de sélection ; pour une maladie létale dans l'enfance s=1.
μ = tx mutation vers l'allèle délétère chez la femme
v = taux de mutation vers l'allèle délétère chez l'homme. Pour simplifier, on considère que les taux de mutation sont identiques dans les deux sexes alors m = s/β, soit, pour un gène létal m = 1/3

Pour les **maladies RLX létales** ou entraînant une **stérilité totale** des garçons, comme la **myopathie de Duchenne**, **1/3 des garçons atteints sont porteurs de mutations récentes**.
 En conséquence, les meilleures stratégies de dépistage et de diagnostic prénatal ne permettront jamais de prévenir la naissance d'au moins 1/3 des garçons atteints, ceux porteurs de mutations récentes.

Inactivation du chromosome X - Théorie de Mary Lyon (1925-2014), 1961

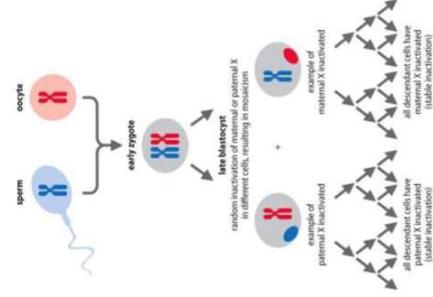
1 des 2 X des cell somatiques d'une fé (46,XX) = inactivé (extinction transcriptionnelle).

Inactivation (lyonisation) **au hasard** → X maternel (Xm), ou X paternel (Xp), **stade précoce dévt embryonnaire** et se transmet de façon **stable et irréversible** au cours des divisions cellulaires.

↳ nb chromosomes X ds cell somatiques (47,XXX), 1 seul chromosome X actif. Fé hétérozygotes Aa habituellement asymptom. Cpdt ds cert cas (inactivation X normal), maladie exprimée ac sévérité variable (dépg degré inactivation).

Zygote : 2 X actifs, puis blastocyste : inactivation.

Processus compliqué, dépendante lncRNA (Xist, 14 kb). *Xist RNA covers the entire chromosome and silences gene expression through epigenetic modification of histones and DNA.*



Hémophilie(s)

Maladies hémorragiques héréditaires, RLX dues au déficit prot coagulation : Facteur VIII (hémophilie A), fact IX (hémophilie B), Facteurs VIII et IX codés par gènes différents (→ même maladie = conség mutation **gènes non alléliques**) :

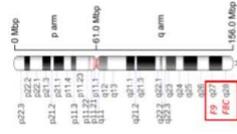
- F8C, localisé en Xq28 (facteur VIII)
- F9 localisé en Xq27 (facteur IX)

Incidence à naiss : 1 à 2/10 000 garçons. Hémophilie A = 85% des cas (séqu codante gène F8C environ 4 fois + grde que celle de F9).

Gravité maladie dépend tx facteur coagulant mesuré ds plasma, exprimé en % par / valeur normale théorique (100%). Formes sévères (environ 50% des cas) : < 2%.

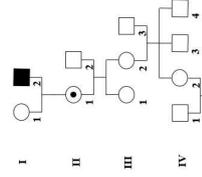
Détermination activité coagulante ds plasma (test de coagulation) → déterminer si vectrice ou non (en théorie fé normale = 100% d'activité, vectrice = 50%). Mais ceci est vrai ds 70% des cas seulement.

Non détection de ttes vectrices obligatoires = liée à inactivation chromos X (fé vectrice ayant inactivé préférentiellement X délétère = test normal) → on ne peut jamais affirmer que vectrice potentielle ayant test normal n'est pas vectrice (soit non vectrice, soit fait partie des 30% vectrices non détectées).



Le risque conditionnel

Obj : pondérer risque a priori (risque mendélien) en y intégrant données généalogiq ou biologiq.
 → théorème de Bayes : calcul proba réalisation événement sachant qu'un autre s'est déjà produit.



$$P(B/A) = \frac{P(A/B) \times P(B)}{P(A/B) \times P(B) + P(A/\bar{B}) \times P(\bar{B})}$$

$$P(\text{III-2 conductrice}/\text{tests normaux}) = \frac{3/10 \times 1/2}{3/10 \times 1/2 + 1 \times 1/2} = 3/13$$

$$P(\text{III-2 conductrice}/\text{2 garçons normaux}) = \frac{1/4 \times 3/13}{1/4 \times 3/13 + 1 \times 10/13} = 3/43$$

$$P(\text{IV-2 conductrice}) = 3/86 \text{ à priori}$$

$$P(\text{IV-2 conductrice}/\text{tests normaux}) = \frac{3/10 \times 3/86}{3/10 \times 3/86 + 1 \times 83/86} = 1/93$$

Quel est le risque pour IV-2 d'avoir un enfant hémophile ? (III-2 et IV-2 présentés des tests de la coagulation normale)

$$\text{Risque pour IV-2 d'avoir un enfant hémophile} = 1/93 \times 1/4 = 1/372$$

Maladies dominantes liées à l'X (DXL)

- * Proportion enf atteints en cas union fê atteinte – hô normal = celle caractère autosomique dominant.
- * Ms, si union hô atteint - fê normale : ttes filles atteintes et ts garçons normaux.
- * SvT - **grave chez fille** que chez garçon (p é létale pdt vie embryon) → ds ce cas, ne semble toucher que les filles.
- * **Pénétrance parfois incomplète et expressivité variable** (cô pr AD).
- * Mode de transmission + rare que ceux précédemment décrits.

Ex : Syndrome de Rett (OMIM 312750 - ORPHA 778)

Rare (1/15 000 filles) DXL, létal pr garçons. Trouble mental grave et global dévt SNC. Dévt quasi normal pdt lère année, régression rapide entre 1-3 ans ac perte utilisation volontaire des mains + retrait social. Tableau clinique dominé par stéréotypies manuelles très évocatrices ; langage absent ou rudimentaire. Marche très instable (« apraxique ») ou jamais acquise. Décélération croiss périmètre crânien → atrophie cérébrale diffuse, affectant substance grise. Evolution = polyhandicap souvent compliqué d'épilepsie.

Cas particulier des disomies uniparentales (DUP)

→ présence chez même individu diploïde de 2 chromosomes homologues hérités d'un même parent. 2 types de DUP :

- ❖ **Isodisomie** : présence de deux copies d'un même chromosome hérité du même parent (non disjonction méiotique pdt **méiose II**).
- ❖ **Hétérodisomie** : présence des deux chromosomes différents (homologues) hérités du même parent (non disjonction méiotique pdt **méiose I**).

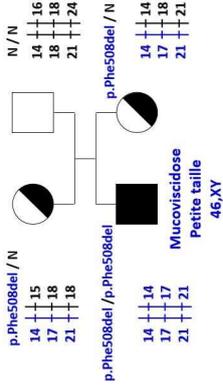
Mécanismes

- **Correction de trisomie** : zygote trisomique perd 1 des chromosomes intéressés par la trisomie.
- **Complémentation gamétique** : fécondation gam disomique pour 1 paire donnée par gam nullisomique pour le même chromosome.
- Duplication chromosomique : fécondation gam nullisomique pour un chromosome par gam monosomique pour le même chromosome, suivie par duplication de ce chromosome.

Mise en évid fondée sur **étude des polymorphismes génétiques** (microsatellites) puisque **canotype normal**.

Comparaison à un locus donné du génotype du propositus ac parents; → mise en évidence de double contribution d'1 parent et non-contribution de l'autre. Conséq des DUP = variées et en rapport ac mécanismes divers :

1) **Isodisomie uniparentale** peut faire apparaître maladie récessive, alors qu'1 seul parent hétérozygote pour l'allèle délétère. → Isodisomie peut révéler maladie autosomique récessive.



II-1 : Pas d'haplotype du père (a hérité de 2 all délétères par sa mère)
Petite taille : DUP peut porter sur empreinte parentale (suite)

2) Une DUP, lorsque porte sur territoire soumis à **empreinte parentale**, peut engendrer une pathologie si elle exclut le gène parental actif.

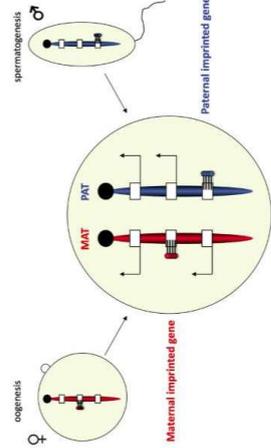
Empreinte parentale

Majo des gènes = expression bi-allélique : 2 allèles (paternel et maternel) s'expriment. Mais gènes soumis à empreinte parentale : **expression monoallélique** et spécifique de l'origine parentale.

→ Modification épigénétique de certains gènes en fonction de leur origine parentale.

S'accompagne d'une inactivation transcriptionnelle spécifique de orig parentale du gène et conduit à **hémizygotie fonctionnelle** (expression monoallélique).

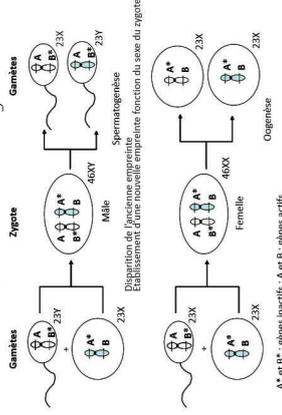
Gène soumis à **empreinte maternelle** → **inactif qd transmis par mère** (ibid pr père)



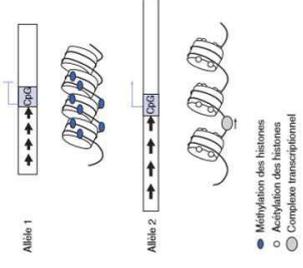
On peut suspecter implication gène soumis à empreinte ds maladie lorsque :

- DUP entraînent phénotypes particuliers en fonction orig parentale
- Manifestations cliniques de la maladie différent en fonction du sexe du parent transmetteur (sexe du parent transmetteur définit type de pathologie)

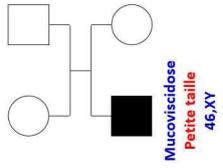
Chez hô : centaine de gènes soumis à empreinte parentale.



- * Doit é initialisée **avant fécondation** (stade génomes séparés).
- * **Effaçable** pcd gamétogénèse pr qu'un chromosome transmis par père à fille (portant empreinte paternelle) puisse, au niv des gam de cette fille, porter empreinte maternelle et inversement.
- * Transmise de façon **stable** lors des divisions cellulaires.
- * Doit pouvoir **réguler expression des gènes** afin d'aboutir à une expression monoallélique (méthylation, condensation chromatine).



Petite taille - conséquence de la disomie uniparentale avec absence de contribution du génome paternel pour le chromosome 7 qui porte un gène soumis à empreinte maternelle et impliqué dans la croissance.



Stratégies d'identification gènes impliqués ds maladies mendéliennes

Gène : ensemble des séquences génomiques codant un ensemble cohérent de produits fonctionnels potentiellement chevauchant.

1 – Stratégies fondées sur la connaissance de la protéine et indépendantes de la position du gène

Protéine codée par le gène connue. Fonction de la protéine codée par le gène connue. Modèle animal de la maladie mendélienne. Cas particulier des maladies avec anticipation associées à des expansions instables de microsatellites

- Gène HBB codant la sous-unité de l'hémoglobine. 1978 : 1^{ère} séq nt gène humain (β-globine)
- Gène F8C codant le FVIII de la coagulation (1980 : VIH, hémophiles traités par don de facteur VIII → contamination de tous les hémophiles par VIH). 2004 : identification gène codant fact VIII.

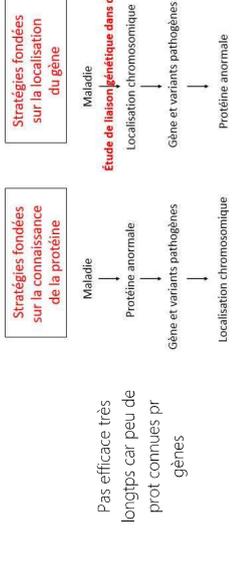
Maladies avec anticipation

Cas particulier d'expressivité variable : au cours générations successives, âge d'apparition maladie de + en + précoce et accentuation gravité.

Base moléculaire : instabilité des répétitions tri nt (microsatellites).

- ❖ Maladies associées à expansions massives de répétitions tri nt localisées en dehors des régions codantes.
 - 1991 : J-L Mandel, syndrome X fragile lié à instabilité triplet CCG ds exon 1 du gène FMR1, =5'UTR (CG = svf méthylés dc extinction promoteur).
 - En 3'UTR : maladie de Steinert
- ❖ Maladies associées à expansions modérées de répétitions CAG localisées dans séq codantes et codant résidus glutamine (Q).

Maladie de Huntington : expansion de CAG (triplet polymorphe ds région codante → pas changement phase de lecture).



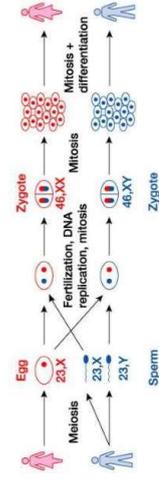
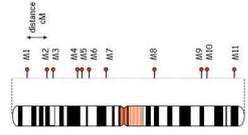
2 – Stratégies fondées sur localisation gène, indép connaissance fonction du gène

→ Clonage positionnel

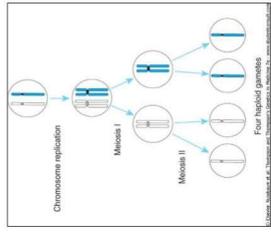
Carte génétique : représentation du génome sur laquelle sont placés des marqueurs génétiques, pour se repérer le long des chromosomes.

Echelle génétique : distance en cM

Construire carte génétique = positionner marqueurs génétiques les uns par / aux autres le long chromos et déterminer dist relatives déterminées par proba recombinaison à la méiose exprimée en cM (1% de recombinaison).

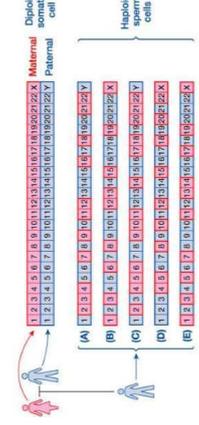


L'établissement d'une carte génétique utilise les propriétés de la recombinaison à la méiose



BRASSAGE ENTRE CHROMOSOMES NON HOMOLOGUES

BRASSAGE ENTRE CHROMOSOMES HOMOLOGUES (crossing-overs)

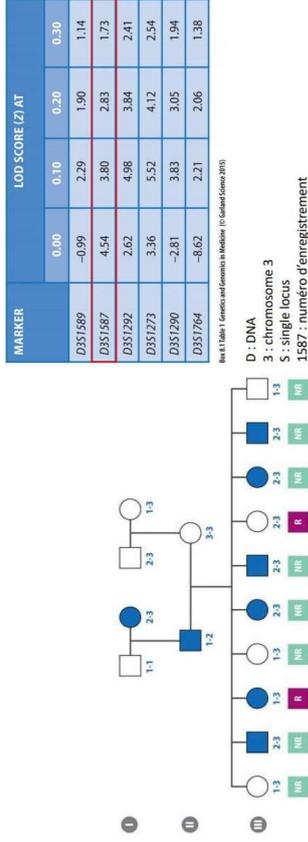


Cartographie de gènes de maladies mendéliennes → familles où ségrège la maladie

Supposons maladie mendélienne AD à pénétrance complète, déterminée par variant pathogène d'1 seul gène.

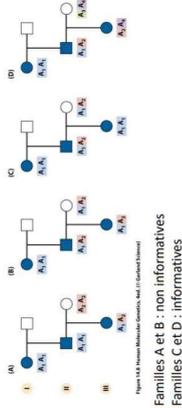
- Cartographie consiste à **localiser gène** de la maladie **par / à** série de **marqueurs moléculaires** de la carte génétique dont l'emplacement est connu sur le génome.
- Nécessité de disposer de familles ayant un ou plusieurs individus atteints et de connaître génotype de chaque membre des familles pour les marqueurs moléculaires considérés.
- Localisation du gène de la maladie : 1ère étape vers son identification.

→ Objectif : $\theta = 0$



$Z(\theta) = 4,5 > 3$ → marqueur lié au gène de la maladie
 Nb gènes identifiés par cette méthode ds années 2000 (1500 gènes identifiés au total en 2002).

Les **méioses informatives** correspondent à des individus chez lesquels il est possible d'identifier l'origine parentale – maternelle ou paternelle – des allèles, ce qui sous-impose l'existence d'une **hétérozygotie** (nécessaire mais pas toujours suffisante) des polymorphismes portés par les autosomes dans les deux sexes et par le chromosome X chez la femme.



- * **Maladie de Huntington** : 1^{er} marqueur lié à gène impliqué ds maladie (bras court chromosome 4), 1983. Puis identification précise gène : 1993.
- * **BRCA1** : bataille féroce pr localisation gène, Marie-Claire King.
- * **F5HD** (myopathie facio-scapulo-humérale) : 1 seule famille a permis localisation gène

Difficultés propres aux différents modes de transmission

- ❖ **Maladies autosomiques dominantes (AD)**
 Nd = malades ; NN = sains. Phénotype renseigne sur génotype, à condition que P = 100%.
 Etudes de liaison ne permettent pas de localiser gène en cause ds maladies AD létales car ts cas observés sont sporadiques (**mutations de novo**).

❖ **Maladies autosomiques récessives (RA)**

Seuls les sujets malades st susceptibles d'être inclus car homozygotes dd ou hétérozygotes composites dd alors que sujet sain = homozygote NN ou hétérozygote Nd.

Etude de fratries seulement si au moins 2 enfants atteints, car on doit comparer les allèles des marqueurs étudiés entre 2 sujets de génotype connu et par conséquent les fratries informatives sont rares dans les pays à faible natalité → nécessité de plusieurs familles, avec plusieurs frères et sœurs atteints.

❖ Difficultés en rapport avec la **structure de la famille ou de la pathologie**

Certains sujets = inaccessibles à l'examen clinique et/ou au prélèvement sanguin pour raisons matérielles ou psychologiques ou décédés au moment de l'étude.

Maladie se manifestant tardivement (pénétrance incomplète en fonction de l'âge) → impossible d'inclure dans l'étude des sujets jeunes (statut non défini).

❖ **Défaut d'informativité**

Habituel qu'un ou plusieurs sujets non hétérozygotes pour un ou plusieurs marqueurs

❖ **Hétérogénéité génétique**

1 maladie peut être conség de mutations de gènes différents non alléliques

❖ **Fréquence maladie**

Si peu fréquente : calcul LOD score impossible (non significatif)

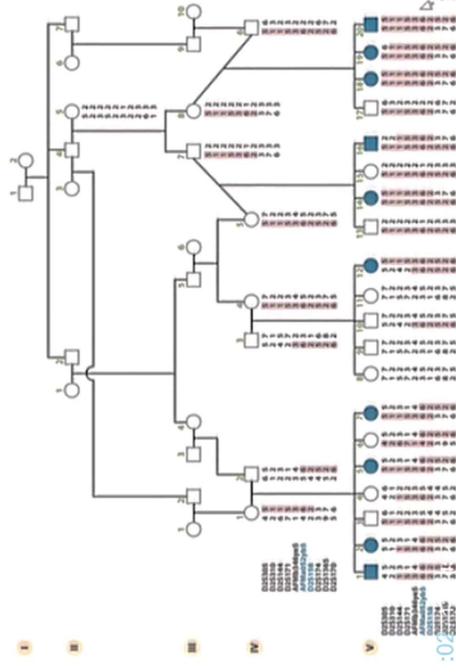
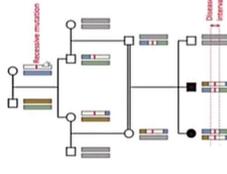
Cas particulier des maladies héréditaires rares et transmises sur mode récessif

Consanguinité favorise arrivée maladies AR (si 1 individu atteint).

→ Cartographie par **autozygotie** (marqueurs homozygotes) = Autozygoty (Homozygosity) mapping

AUTOZYGOSITY MAPPING

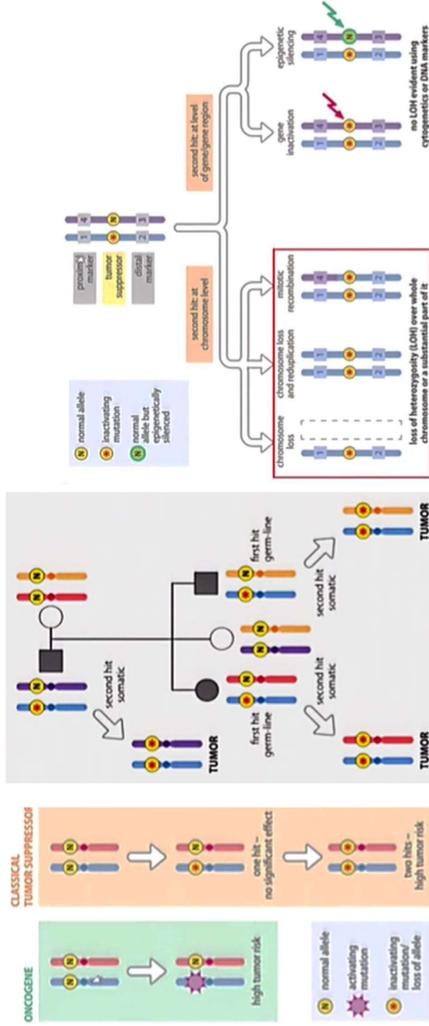
- **Affected individuals will be homozygous (autozygous) for microsatellite marker/SNP alleles around disease locus**
 - **Unaffected individuals not homozygous for the same marker alleles in the gene region = (heterozygous or homozygous for another allele) – this information can be used to exclude candidate regions**
- Avec consanguinité : territoire dans lequel se situe allèle délétère = le même entre individus partageant mutation ancestrale.
 → recherche de **territoires où gène est homozygote** ; territoires d'homozygotie peu fréquents (proba homozygotie ds microsat multi alléliques = faible)



Cas particulier des prédispos héréditaires au cancer

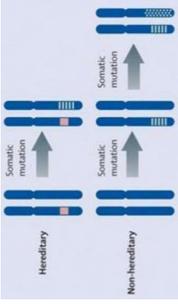
→ Localisation des gènes suppresseurs de tumeurs (GST) par analyse des pertes d'hétérozygotie (LOH : Loss Of Heterozygosity)

Cancer : driver mutation, puis accumulation de mutations



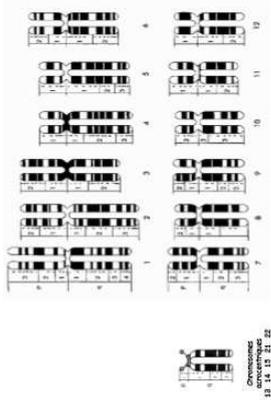
Formes héréditaires : individu hétérozygote (ds ttes cell : all délétère pr gène suppresseur de tumeur) → 1 seule mutation somatique nécessaire du 2^e allèle pr que tumeur survienne

→ Survient + facilement que pr individu bien portant (2 mutations somatiques pr que cancer apparaisse)
→ Individus jeunes car 1 seul événement nécessaire.



Caryotype

Recherche d'anomalies chromosomiques. Maladies rares : caryotype le + svt normal



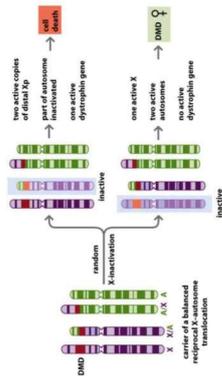
Exceptions :

Myopathie de Duchenne (RX)

Dystrophie musculaire de Duchenne (MIM 310200/ORPHA 98896

46,XX, (X:5)(p21;q35)

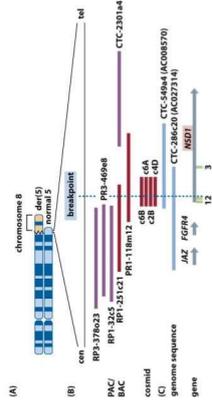
Due à translocation équilibrée X-5, notée (X;5)(p21;q35).
 Fê portant cette translocation = malades. Pourquoi ?
 Hypothèse : translocation a cassé en deux le gène DMD
 Fê hétérozygote ne devrait pas être malade. Mais inactivation chromosome X délétère = fréquente (inactivation autosomes non réalisées, = létales, haploinsuffisance non observée cz autosomes).
 → localisation du gène



Syndrôme de SOTOS (MIM 117550/ORPHA821)

Génotype, macrocéphalie et difficultés d'apprentissage

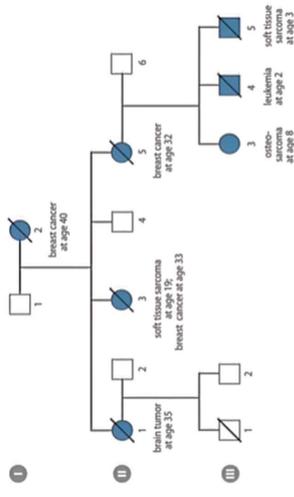
46,XX,45,X(9p5;q24.1)



LFS (Syndrome de Li-Fraumeni) : affection rare du sujet jeune, **prédispo à divers tumeurs**. Définition basée sur critères familiaux : observation d'un sarcome cz sujet atteint de moins de 45 ans apparenté au 1^{er} degré à une personne ayant eu un cancer de n'importe quel type avant 45 ans, ou au 2^e degré à une personne ayant eu un cancer ou un sarcome à moins de 45 ans.

Se transmet sur mode AD. **Mutation germinale de TP53** retrouvée ds 70% des familles LFS. Risque de dév cancer pr sujet porteur mutation délétère de TP53 est de 15% à 15 ans, 80% pr fê à 50 ans et 40% hô à 50 ans

Syndrôme de Li-Fraumeni (MIM 151623 – ORPHA524)



Mais : Patiente sans séquence répétée = saine. Donc problème repose dans séquence D4Z4.
 Etat normal : beaucoup de séquences répétées → chromatine fermée, méthylation, D4Z4 non exprimée.

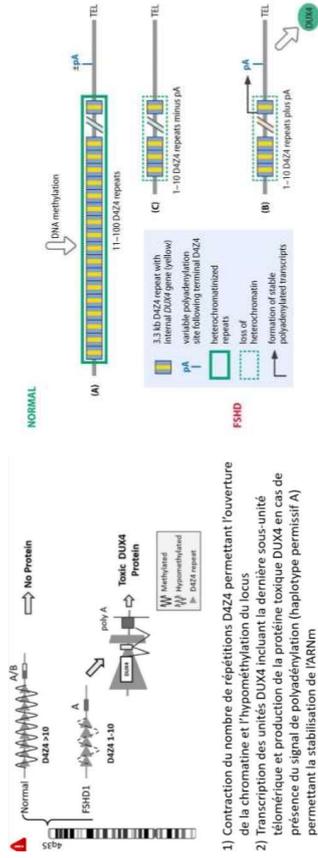


→ Diminution nb répétitions, contraction de D4Z4 → ouverture chromatine → D4Z4 transcrite, puis traduite → protéine toxique pour le muscle (**gain de fonction**)

A partir transcription : ARN messenger, doit être mûruré en 3' par **signal de polyadénylation** (stabilisation)

Haplotype A (permissif) : signal de polyadénylation

Haplotype B (non permissif) : pas de poly-A, messenger instable → pas de protéine traduite → sain



- 1) Contraction du nombre de répétitions D4Z4 permettant l'ouverture de la chromatine et l'hyperméthylation du locus
- 2) Transcription des unités DUX4 incluant la dernière sous-unité télomérique et production de la protéine toxique DUX4 en cas de présence du signal de polyadénylation (haplotype permissif A) permettant la stabilisation de l'ARNm

Approches pangénomiques utilisées pour identifier les gènes impliqués dans les maladies mendéliennes

CGH-array (Comparative Genome Hybridization)

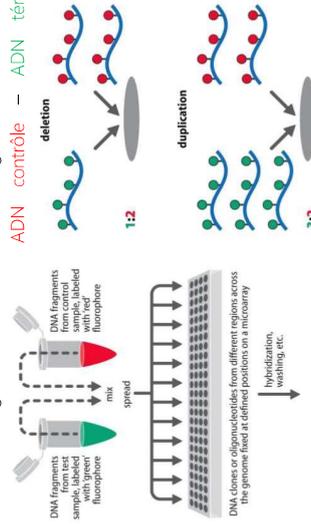
= ACPA (Analyse Chromosomique sur Puce à ADN)

→ Résolution nettement supérieure : 50-100 kb

Puce à ADN = réseau ; surface (lame de verre) pr fixation d'oligonucléotides de 60-mers (=60 bases), représentatifs du génome humain. Nombre d'oligonucléotides définit la résolution de la puce.

ADN contrôle – ADN témoin (fluorophores différents). Puis hybridation :

hybridation :
 Délétion : rouge ++ (contrôle) ;
 duplication : vert ++ (témoin)
 → observation des CNVs (+ de 50 pnb), pertes/gains de matériel de façon pangénomique

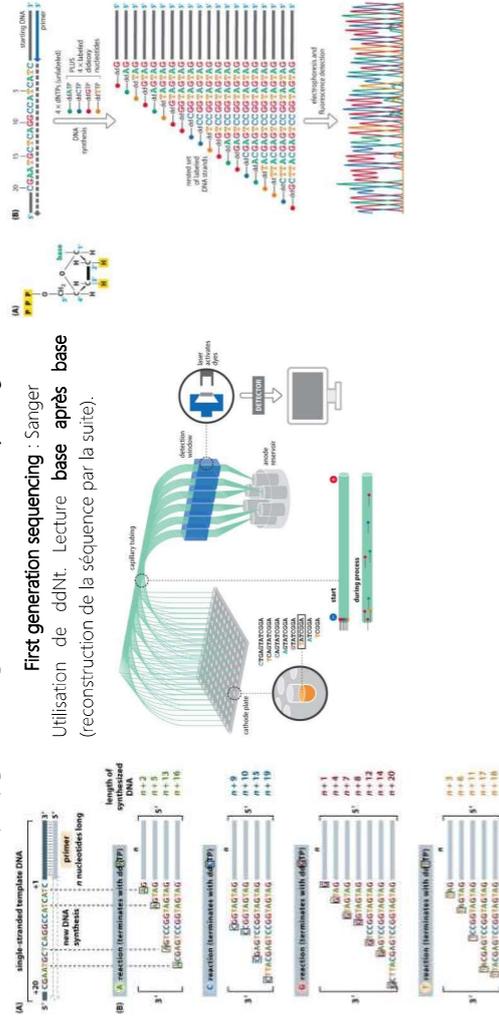


NGS (Next Generation Sequencing)

- Séquençage de tous les exons et introns flanquants du génome (**Whole Exome Sequencing - WES**)
- Séquençage de tout le génome (**Whole Genome Sequencing WGS**)

First generation sequencing : Sanger

Utilisation de ddNt. Lecture **base après base** (reconstruction de la séquence par la suite).

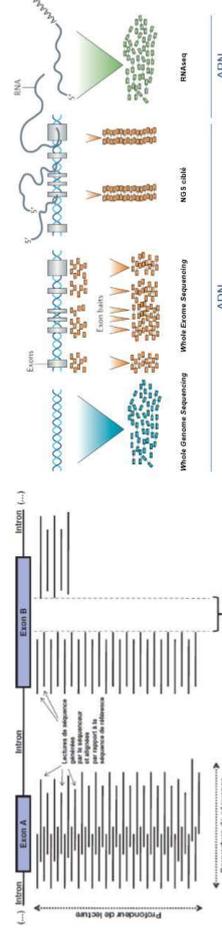


NGS : principe de séquençage par protection/déprotection, on parle de **séquençage par synthèse**. Lecture continue des bases.

Séquençage de **READS** (morceaux d'ADN de 151 pnb). Chaque read lu 2 fois (sur les 2 brins → Paired-End (pour être sûr de ne pas se tromper)).

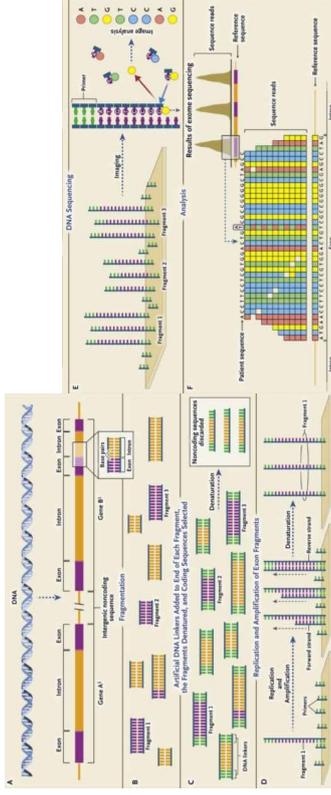
1 flowcell = 14 milliards de reads

Profondeur : nb de lectures indépendantes d'une base par séquençage.
Couverture : par rapport à une cible, pourcentage de bases séquencées.



WES # 50 millions nucléotides
Couverture des exons # 98%
Profondeur > 150X

Séquencer seulement les exons = possible en utilisant comme oligonucléotides (sondes) les séquences des exons → **Capture des exons par hybridation** ac sonde.
 Comparaison exome entre deux personnes : 20 000 variants.



Plus aucune capture d'exons : WGS (3 milliards nucléotides)

Profondeur # 30X

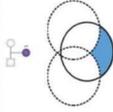
Couverture plus homogène des exons que le WES

→ Par syst de lecture, possibilité de lire seulement les exons (exome in silico)

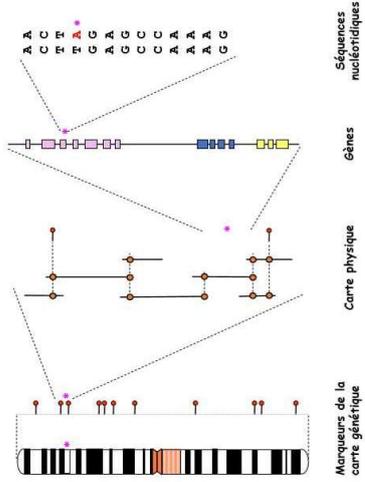
Mutations héritées – mutations de novo

Approche du trio : papa/maman/enfant

Mutations de novo : séquençage de trios



WGS ou WES SOUSTRACTIFS



Manquiers de la carte génétique

Carte physique

gènes

Séquences nucléotidiques

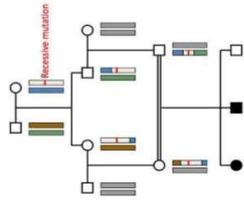
Mutations héritées	Mutations de novo
<p>Si consanguinité : indiv malade = homozygote pr tout un territoire génétique → WES ou WGS à recherche de ROH (Run Of Homozygosity) pr trouver régions candidates pr accueillir mutation</p> <p>Si maladie récessive : perte de fonction → recherche mutations pLoF (predicted Loss of Function) homozygotes → 2/3 des maladies identifiables</p>	<p>Approche trio</p> <p>Pas de forme familiale, maladie vient d'apparaître (Jusqu'au NGS, impossible)</p> <p>→ isoler variants spécifiques de l'enfant = candidats (pb devient seulement informatique)</p>

Mais, limites de l'approche du trio :

- * Pb peut ne pas é génétique
- * Si P < 100% : père ou mère peut posséder la variation, mais celle-ci sera potentiellement exclue
- * Père doit é dispo

AUTOZYGOSITY MAPPING

Affected individuals will be homozygous (autozygous) for the same marker alleles in the region – (heterozygous or homozygous for another allele) – this information can be used to exclude candidate regions

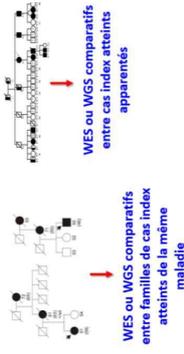


Ostéogénèse imparfaite (AD) : collagène non stable, mutation gain de fonction.

Arabie Saoudite : argent + consanguinité (10%) → séquençage, WGS pr recherche mutations pLoF homozygotes. Hô devient modèle naturel de génétique (test capacité d'un gène à subir des mutations).

Autosomiques dominantes

FORMES FAMILIALES



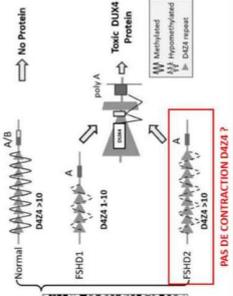
WES ou WGS comparatifs entre cas index d'une même famille ou entre cas index de familles différentes atteints de la même maladie (sujets atteints qui se ressemblent) → ont-ils des variants pathogènes en commun ? (Attention : derrière une maladie, il peut y avoir différents gènes impliqués ; cf Hémophilie)

FSHD : rétrogène D4Z4 ne doit pas être exprimé, car code protéine toxique pour muscle (DUX4).

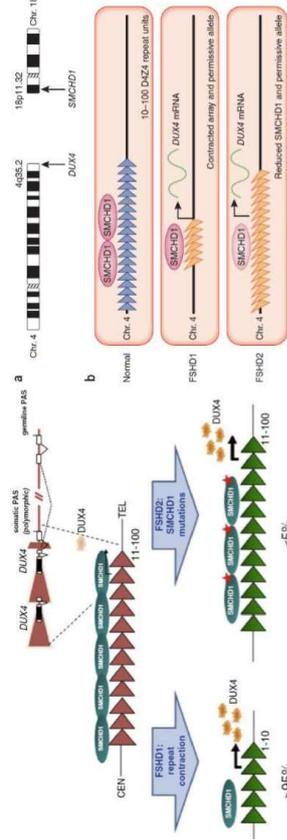
Comptage nb répétitions impossible par PCR (fait + de 3 kb, max = 1 kb) → Southern blot

Mais, exception : patient ac D4Z4 > 10 répétitions, mais pas de contraction → machinerie nécessaire à contraction en cause.

→ Exome comparatif ou soustractif pr mise en évidence gène.



→ Gène SMCHD1 : impliqué ds inactivation de l'X et fermeture chromatine, situé sur chromosome 18, possède allèle permissif : si SMCHD1 muté → FSHD (même si nombreuses répétitions de D4Z4).

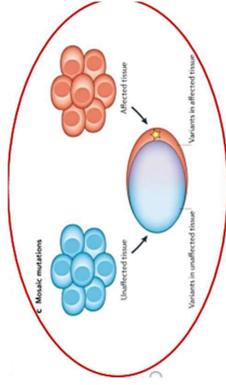


→ Apparition maladie = conditionnée par mutation de 2 gènes = **DIGENISME**

- Mutation à l'état hétérozygote du gène SMCHD1
- Présence d'un allèle permissif (PAS+) au locus D4Z4

Expression clinique maladie dépend nb répétitions de D4Z4 + allèle permissif ou non ds gène SMCHD1.

Exomes sous-structurés



Problèmes spécifiques posés par les approches « pangénomiques »

Séquençage génome : pb = tout le génome, càd partie concernée par maladie + le reste → mutation qui pourrait intervenir ds maladie

→ NOTION DE DONNEES ADDITIONNELLES, SECONDAIRES ET INCIDENTES

Données additionnelles : résultat d'un examen **sans relation directe** avec l'indication initiale.

- **Donnée secondaire** : variation pathogène sans relation directe avec indication initiale et **recherchée activement** en analysant une liste de gènes préétablie.
- **Donnée incidente** : variation pathogène sans relation directe avec l'indication initiale et de **découverte fortuite**.

→ Réflexion du CCNE sur l'évolution des tests génétiques liée au séquençage ADN humain à très haut débit (avis 124) : droit de savoir et de ne pas savoir.

Hypercholestérolémie familiale

Maladie AD fréquente (**30% ds pop franç**). Constitue, ac obésité, diabète et hypertension artérielle, un des facteurs de risque majeurs d'**athérosclérose** générant maladies cardiovasculaires :

- * Cardiopathies coronariennes
- * Maladies cérébrovasculaires
- * Artériopathies périph
- + cardiopathies rhumatismales, cardiopathies congénitales, thromboses veineuses profondes et embolie pulmonaire.

OMS, 2012 : **17,5 millions de personnes décédées** de maladie cardiovasculaire ds monde et 1/3 des maladies ischémiques cardiaques dû à l'x cholestérol élevé.

Maladies cardiovasculaires = **1^{ère} cause mortalité ds monde**, 2^e en F (1^{ère} pr fr), juste après cancers et st à orig de + de 140 000 décès par an + une des principales causes morbidité en F ac 3,5 millions de personnes (assurés du régime général) traitées en 2012.

Coût important pr société : en 2009= 106 milliards d'euros (maladies cardio-neurovasculaires), soit 9% des dépenses totales de santé en Europe et USA. Coûts directs annuels devaient tripler entre 2010 et 2030.

Touche **13 millions d'individ** ds monde. Fréquence estimée : 1/200

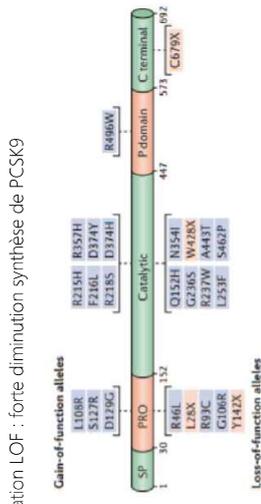
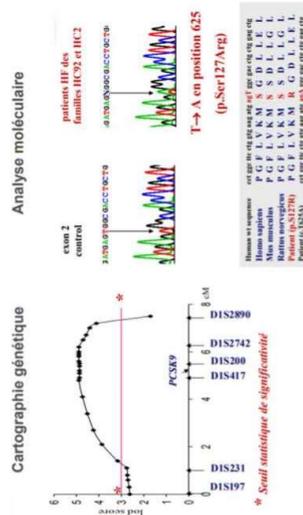
Retrouvée ds art : auge de Delphes, La Joconde (xanthome tendineux, xanthelasma)

Années 90 : HF surtout due à mutations gène LDLR (récepteur cholestérol)

Puis : **analyse liaison génétique** pr trouver 3^e gène impliqué → **PCSK9** (Catherine Boileau). Mise en évidence de variants faux-sens de PCSK9

→ Rupture de paradigme : « trio » (et non duo) LDLR interagit ac ligand apo B et PCSK9 → complexe LDL internalisé. PCSK9 **régule nb de LDLR à surf cellule** : + il y a de PCSK9, + il y a de récepteur et - il y a de cholestérol circulant.

- * Mutation GOF (gain de fonction) de PCSK9 : demi-vie de PCSK9 + longue et affinité + forte pr LDLR.
- * Mutation LOF : forte diminution synthèse de PCSK9



PHYLOGENIE

Phylogénie : représentation sous forme de **struct arborescente**, des **relations de parenté** entre organismes vivants/disparus.

Raconte hist émergences successives des gpes d'organismes au cours tps.

Permet de classer tout ce qui évolue au cours tps, pas seulement organismes vivants : langues, comportements...

Taxon : terme générique; regroupement organismes reconnu en tant qu'unité formelle (pop, espèces, genres, fam, indiv, ...), céd ce qui peut faire objet classification.

→ « Qui est le plus proche parent de qui ? » (≠ généalogie: « Qui descend de qui ? »)

= recherche parents (non d'ancêtres). Lect arbre en termes de groupe-parents (non de groupes ancêtres à descendants).

Phylogénie = fondée sur concept de « **descendance avec modification** » énoncé par Darwin :



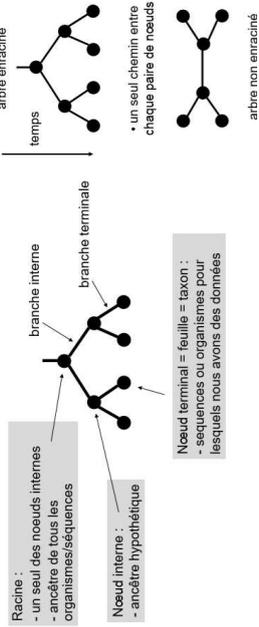
Arbres phylogénétiques

→ Inférence sur le passé (ce qui est le + probable).

Arbre (et non réseau) : 1 seul chemin relie deux taxons. Suppose modèle d'évolution par **divisions dichotomiques**, sans échange génétique entre taxons divergents.

Racine = pt origine arbre, détermine sens lect (arbre p è enraciné ou non). → 1er ancêtre commun.

Arbre phylogénétique = tirs enraciné ; orienté ds tps et raconte une histoire → succession des émergences gpes d'organismes vivants/ayant vécu au cours tps.

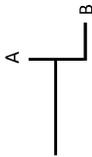


Clade : regroupement de taxons partageant **ancêtre commun exclusif** (= ancêtre commun + tous ses descendants).

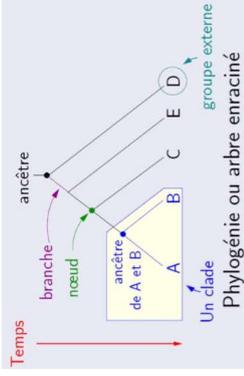
Cladogramme : infos sur relation de parenté

Phylogramme : infos sur relations parenté+ vitesse d'évolution des taxons.

Ex :



Ici : même tps écoulé depuis émergence, ms B a accumulé + de mutation (évolution + rapide)



Reconstruction phylogénétique

- 1) Choix **unités évolutives**
- 2) Obtention de **caractères comparables** pr objets étudiés
- 3) Application **méthode de reconstruction phylogénétique** pr meilleure arborescence possible

Objets : espèces, pop, indiv, fossiles, gènes, langues

Caractères : morpho, linguistiques, patronymiques, ethnologiques, génétiques (localisation : autosomes, X, Y, ADN nucléaire/ mitochondrial ; struct moléculaire : SNP, INDELS, polymorphisme de répétition (STR), CNV).

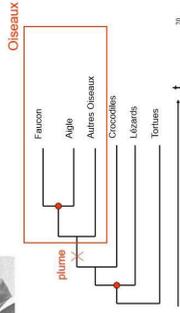
Méthodes :

- ❖ **Phénétique (distance)** : qui ressemble le + à qui ? → regroupement sur base de la + grde **similitude globale**
- ❖ **Cladistique (parcimonie)** : regroupement sur base **caract nouveaux/dérivés**. Grâce à évolution : passage état primitif à dérivé (pr 1 caract). Taxons regroupés doivent avoir caract dérivés en commun.
- ❖ **Probabiliste (max de vraisemblance)** : modèle d'évolution le + **vraisemblable** sachant l'observation.

Méthode cladistique (parcimonie)



Hennig (1913-1976) : La cladistique



Ancêtres communs les + récents signifie taxons les + proches.

2 grds principes :

- Regroupements exprimant liens de parenté sur base du **partage d'innovations évolutives** (= état dérivé d'un caractère) → **gpes monophylétiques** = clades (seuls reconnus ds classifications modernes), doivent regrrper ancêtre commun + ts descendants (sinon = gpe paraphylétique, non reconnu).

- **Parcimonie** (= principe d'économie d'hypothèses) : arbre le + parcimonieux, c-à-d celui qui supposera le moins de transformations évolutives des caractères (minimise changements évolutifs, = le + probable).

→ Maximise ressemblances dues à ascendance commune, partage ancêtre commun (= homologies)

Polarisation des caractères

Au préalable : identification états évolués des caractères. Plusieurs critères utilisés pr préciser états primitif/dérivé :



Critère paléontologique

État primitif = apparu le + anciennement au cours tps géologiques ; état évolué = apparu le + récemment : + profondeur ds sol = import, + fossile est vieux.



Critère ontogénétique (expérimental et peu utilisé)

Loi de Haeckel : « l'ontogénie récapitule la phylogénie » : pdt dév embryonnaire, caractères généraux apparaissent avant caractères + particuliers à l'espèce. Etat évolué apparaît le + tardivement.

Critère de l'extra-groupe (le plus utilisé)

Choix d'un ou plusieurs taxons dt on n'est sûr qu'ils ne font pas partis grpe d'étude, et qu'ils ne viendraient pas se brancher à l'intérieur de « l'ingroup ». Enracinement du grpe d'étude, ancêtre commun + ancien que ceux partagés entre taxons du grpe int. Etat du caractère observé ds l'extra-groupe = primitif, ts autres caract (observés ds l'ingrp) = dérivés.

Homologie = ressemblance héritée d'un **ancêtre commun** (détection par application du principe des connexions).

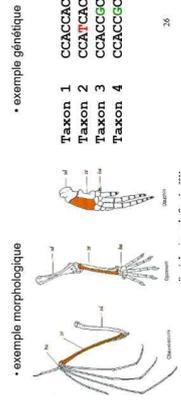
Homoplasie = ressemblance non héritée d'un ancêtre commun, due à phénomènes de convergence ou de réversion :

- Convergence = acquisition d'un même état de caractère plusieurs fois indépendamment.
- Réversion = retour à l'état de caractère primitif.

Principe des connexions

Struct = homologues si mêmes connexions ac struct voisines cz organismes de même plan organisation (V fonction et forme).

Ex : radius cz chauve-souris, opossum, dauphin



→ 7 caractères (nt), homologie car entourés par états de caractères similaires

Question initiale : **qui est le + proche de qui ?**

- Définition pb phylogénétique et délimitation groupe d'intérêt
- **Echantillonnage taxonomique** : choix taxons représentant groupe intérieur (ingroup). Choix également d'un ou plusieurs extra-groupe(s) → enrancement + polarisation des caractères.
- **Echantillonnage des caractères** : comparables entre taxons et informatifs.
- **Codage et construction matrice taxon-caractères** : établissement hypothèses d'homologies (principe des connexions) et **polarisation des caractères**.
- **Exploration de tous les arbres possibles** : Chaque arbre est évalué au regard matrice.
- **Critère de choix arbre** : principe de parcimonie dans le cas de la cladistique → on choisit l'arbre qui pose le minimum d'hypothèses de transformation des caractères.

L'arbre choisi permet :

- identifier a posteriori **homologies et homoplasies** (et éventuellement cd sur mode évolution caractères).
- obtention **classification phylogénétique** : groupes monophylétiques fiables se voient attribuer un nom.

Exemple

Les chauve-souris sont-elles des oiseaux ou des mammifères volants ?

Caractères	Homme	Chauve-souris	Traite
Poumons O	1	1	0
Pattes ●	1	1	0
Mamelles □	1	1	0
Poils ■	1	1	0
Ailes ∇	0	1	0
Plumes —	0	0	1
Mâchoires x	0	0	0

7 caractères macroscopiques choisis

Codage et construction **matrice taxon-caractères**

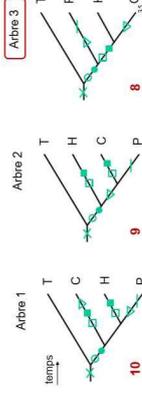
Caractères	Taxons		
	Groupe d'étude		Extra-groupe
	Homme	Chauve-souris	Pigeon
Poumons O	1	1	1
Pattes ●	1	1	1
Mamelles □	1	1	0
Poils ■	1	1	0
Ailes ∇	0	1	1
Plumes —	0	0	1
Mâchoires x	0	0	0

0 : caractère retrouvé cz extra-groupe = état primitif (ex : absence de poumons)

1 : état dérivé

- Exploration des arbres possibles : 3 arbres racinés possibles.
- Evaluation de chacun au regard de la matrice
- Choix de l'arbre le **plus parcimonieux**

Arbre 3 : 8 transformations → le + parcimonieux



Arbre 1 : convergence évolutive pour poils et mamelles, mais on aurait pu proposer réversion (même nb transformations)

Arbre 3 : 8 transformations → le + parcimonieux → Chauve-souris + proche de Hô, dc mammifère volant (ailes) = convergence évolutive, homoplasie

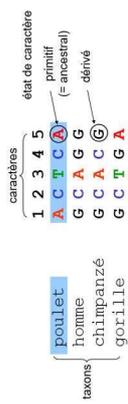
- Caractères variables ? Tous, sauf mâchoires
- Caractère invariant = non informatif. Mais caract non informatif n'est pas forcément invariant. Ex : poumons, pattes et plumes = non informatifs.
- Ici, caract informatifs = poils, mamelles, ailes
- Homologies révélées par l'analyse cladistique : Hô-chaue-souris pr poils, mamelles. Homoplasie : Chauve-souris-pigeon pr ailes.

Traitement de données moléculaires

- * Séquences **homologues** ↔ héritées d'un **ancêtre commun** → organismes ac origine commune.
- * **Matrice taxons/caractères** = **alignement multiple** de séquences nucléiques ou protéiques.
- * Alignement des séquences = cruciale, permet obtenir caractères comparables entre taxons et poser hypothèses d'homologie.
- * Chaque **position de l'alignement** = **1 caractère**. **Bases/AA** = **états des caractères**.
- * **Polarisation** des caractères par le critère extra-groupe

Alignement multiple des séquences de protéines codées par le gène CDC2 chez diverses espèces

	1	2	3	4	5	10	15	20	25	30
ab015	M	E	D	T	K	E	K	I	E	K
ab044	M	E	D	V	I	K	E	K	I	E
ab259	M	E	D	V	I	K	E	K	I	E
ab260	M	E	D	F	E	K	E	K	I	E
ab264	M	E	D	F	E	K	E	K	I	E
ab265	M	E	D	F	E	K	E	K	I	E
ab266	M	E	D	F	E	K	E	K	I	E
ab267	M	E	D	F	E	K	E	K	I	E
ab268	M	E	D	F	E	K	E	K	I	E
ab269	M	E	D	F	E	K	E	K	I	E
ab270	M	E	D	F	E	K	E	K	I	E
ab271	M	E	D	F	E	K	E	K	I	E
ab272	M	E	D	F	E	K	E	K	I	E
ab273	M	E	D	F	E	K	E	K	I	E
ab274	M	E	D	F	E	K	E	K	I	E
ab275	M	E	D	F	E	K	E	K	I	E
ab276	M	E	D	F	E	K	E	K	I	E
ab277	M	E	D	F	E	K	E	K	I	E
ab278	M	E	D	F	E	K	E	K	I	E
ab279	M	E	D	F	E	K	E	K	I	E
ab280	M	E	D	F	E	K	E	K	I	E
ab281	M	E	D	F	E	K	E	K	I	E
ab282	M	E	D	F	E	K	E	K	I	E
ab283	M	E	D	F	E	K	E	K	I	E
ab284	M	E	D	F	E	K	E	K	I	E
ab285	M	E	D	F	E	K	E	K	I	E
ab286	M	E	D	F	E	K	E	K	I	E
ab287	M	E	D	F	E	K	E	K	I	E
ab288	M	E	D	F	E	K	E	K	I	E
ab289	M	E	D	F	E	K	E	K	I	E
ab290	M	E	D	F	E	K	E	K	I	E
ab291	M	E	D	F	E	K	E	K	I	E
ab292	M	E	D	F	E	K	E	K	I	E
ab293	M	E	D	F	E	K	E	K	I	E
ab294	M	E	D	F	E	K	E	K	I	E
ab295	M	E	D	F	E	K	E	K	I	E
ab296	M	E	D	F	E	K	E	K	I	E
ab297	M	E	D	F	E	K	E	K	I	E
ab298	M	E	D	F	E	K	E	K	I	E
ab299	M	E	D	F	E	K	E	K	I	E
ab300	M	E	D	F	E	K	E	K	I	E

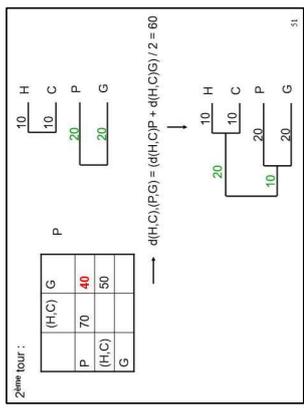
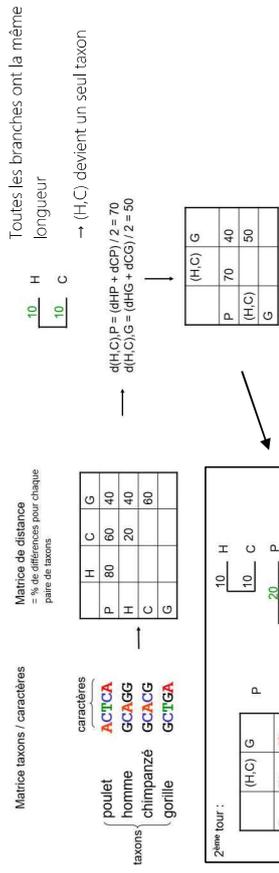


- **synapomorphie** : état de caractère dérivé (= innovation) partagé
- **symplesiomorphie** : état de caractère primitif (= ancestral) partagé
- **autapomorphie** : état de caractère dérivé propre à un seul taxon

Méthode phénétique (distance)

- Quantifie **ressemblance globale** entre taxons et les regroupe en fonction de cette similitude globale.
- Calcul pour chaque couple de taxons d'un **indice de similitude globale**
- Construction d'un arbre à partir de la **matrice de distances**
- ❖ Méthode **UPGMA** (Unweighted Pair-Group Method using arithmetic Averages) → construction par agglomération successive des taxons, des + proches aux + éloignés.

- Calculer la matrice de distance
- Choisir la paire la plus similaire
- Grouper les taxons correspondants
- Re-calculer la nouvelle matrice de distance

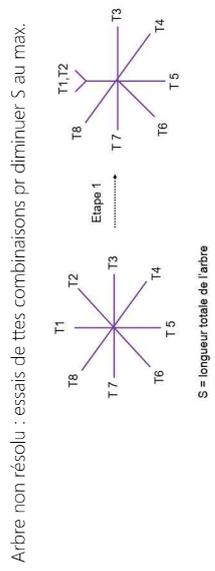


Distance ancêtre commun-taxon = tjrs la même

→ Hypothèse **d'horloge moléculaire** (fréquence mutations = tjrs la même). Mais rarement vérifiée → arbres parfois faux.

❖ Méthode du **Neighbor-joining**

Algorithme le + élaboré et le + rapide : le + utilisé. Regroupe les taxons de façon à **minimiser la longueur totale de l'arbre** (pas d'hypothèse d'horloge moléculaire).



Limites :

- Arbres obtenus = **non enracinés** ; peuvent é enracinés a posteriori avec méthode pt médian (racine placée à égale distance de ts taxons).
- Arbres de distances = **degrés de ressemblance entre taxons**. Pas de distinction entre ressemblance héritée/non héritée. Peut conduire à agglomération de taxons, sur base de caractères primitifs partagé → **pas de prise en compte des homoplasies** éventuelles.

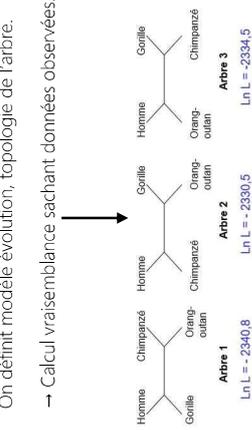
Méthode probabiliste (maximum de vraisemblance)

La + utilisée pr classifications moléculaires (séquences).

- Modèle**
 Paramètres θ
- arbre T
 - Probabilité de substitution
 - Proportion des différents états
 - Hétérogénéité du taux de substitution
 - Proportion de sites invariants

On définit modèle évolution, topologie de l'arbre.

→ Calcul vraisemblance sachant données observées.



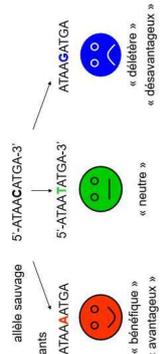
Méthode
 Probabilité des Données, sachant le Modèle et les θ
 $P(D|M, \theta) \propto L(M, \theta)$
 $\max_{\theta} P(D|M, \theta) \Rightarrow \hat{\theta}$

En résumé

Méthode cladistique	Méthode phénetique	Méthode probabiliste
<ul style="list-style-type: none"> - Analyse des caractères - Construction matrice taxons-caractères - Maximise nb caractères dus à l'ascendance commune 	<ul style="list-style-type: none"> - Analyse de distance - Construction matrice de distances - Regroupement des taxons sur base de similitude globale 	<ul style="list-style-type: none"> - Identification arbre le + vraisemblable sachant les données (matrice tax-caract ou dist) - Sachant modèle d'évolution des caractères, posé a priori

GENETIQUE DES POPULATIONS

Etude de variation ADN : évolution ds tps et espace, effets sur indiv qui les portent, transmission, évolution au sein et entre espèces, histoires passées dt elle est témoinnage.



Allèle : état du locus sur séquence donnée.

Il existe 2 éch évolution :

Microévolution	Macroévolution
Au sein de populations d'une espèce	Entre plusieurs espèces
Echelle : générations	Relations de parenté entre espèces
Mutation, migration, dérive génétique, sélection naturelle	Echelle : tps géologiques

Introduction et définitions

Microévolution : changement des fréquences alléliques au sein d'1 pop d'indiv de la même espèce.

Ex : allèles couleur des ailes (vert - jaune) → fréquence ds pop à 1 moment donné.

Evolution : accumulation ds tps de changements hérités au sein des pop.

Espèce : grpe génétiquement fermé au sein duquel organismes peuvent, par alternance méiose / fécondation, séparer ou réunir les divers all de chacun des gènes et concevoir combinai génétiques nvelles par recombinaison génétique.

⇒ Barrières peuvent lim possibilités de croisement entre indiv de même espèce :

- isolement géo (océan, ...)
- isolement éco (floraison décalée, ...)
- barrières cult, sociales, ethniques, ...

→ Nécessité de définir une unité locale de croisement des indiv au sein de l'espèce.

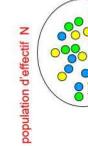
Population : ensemb d'indiv appartenant à même espèce, tel qu'1 de ses membres a proba très élevée, si se reproduit, d'y trouver ses partenaires et d'y laisser ses descendants, alors que proba + faible de trouver partenaire et laisser ses descendants ds autre pop (espèce supposée diploïde à reproduction sexuée).

- * Critère d'ordre systématique : appartenir à même espèce
 - * Critère d'ordre bio : possibilité de se croiser → brassage des gènes entre générations, pop= entités adaptables et évolutives, espèce= organisée en pop.
- Génomme collectif ou patrimoine/pool génétique = Σ génotypes indiv pr chacun des loci.

Variations génétiques dans les populations

Mesure variation génétique ds une pop :
 Co-dominance : chaque génotype a son propre phénotype.

- organisme diploïde
 - locus autosomique bi-allélique : B, J
 - transmission co-dominante
 - génotypes : (BB) (BJ) (JJ)
 - phénotypes :



1. Estimation des fréquences phénotypiques :

$$f(B) = \frac{n(B)}{N} \quad \Sigma = 1; \quad 0 \leq f \leq 1$$

$$f(J) = \frac{n(J)}{N}$$

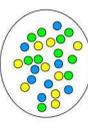
2. Estimation des fréquences génotypiques :

$$f(BB) = \frac{nb \text{ individus génotype } BB}{N}$$

$$f(BJ) = \frac{nb \text{ individus génotype } BJ}{N}$$

$$f(JJ) = \frac{nb \text{ individus génotype } JJ}{N} \quad \Sigma = 1$$

- génotypes : (BB) (BJ) (JJ)
 - phénotypes :



3. Estimation des fréquences alléliques :

$$f(B) = \frac{n^B(BB) + n^B(BJ)}{2 \cdot N}$$

$$f(J) = \frac{n^J(BJ) + n^J(JJ)}{2 \cdot N}$$

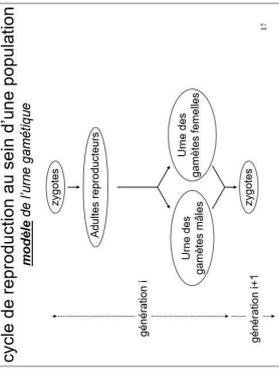
Σ fréq all pr 1 locus = 1

- Fréq génotypique** : mesure distribution de la variation génétique d'une pop parmi ses membres pr 1 locus donné
- Fréq allélique** : mesure tx variation génétique ds une pop pour un locus donné et détermine le polymorphisme du locus. Locus polymorphe si : ≥ 2 allèles dans la population et fréq de l'all minoritaire ≥ 0.01
4. **Taux d'hétérozygotie (Hz)** : somme des fréq des hétérozygotes, calculées sous loi de Hardy-Weinberg pour un locus donné.

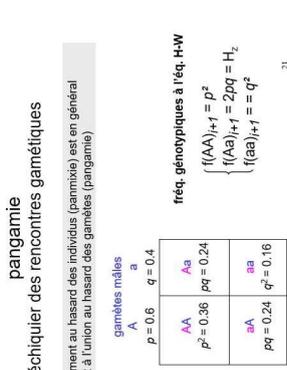
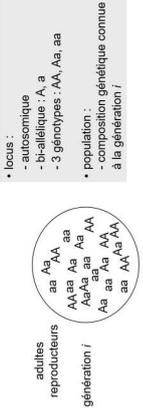
$$Hz = \Sigma 2 p_i q_i = 1 - \Sigma p_i^2 \quad (\text{organismes diploïdes})$$

Equilibre de Hardy-Weinberg

- Si conditions d'application st remplies, pop n'évolue pas d'un pt de vue génétique (fréq all inchangées) :
- Effectif infini
 - Pas de migration
 - Pas de mutation
 - Pas de sélection (fécondité, viabilité)
 - Panmixie, pangamie (rencontre aléatoire des gamètes).



- Population **théorique idéale** d'un pt de vue évolutif :
- Respecte conditions d'application de l'équilibre/principe de H-W
 - Fréq all stables au cours tps et d'une génération à suivante
 - Fréq génotypiq s'établissent à valeurs d'équilibre du fait de panmixie, et restent stables au cours tps.



- Si croisements entre indiv = aléatoires (**panmixie**), **fréq** des différents **génotypes** à génération i+1 peuvent être **prédites** à partir des seules fréq alléliques de la génération précédente i.
- Pop à l'équilibre et conditions idéales restent réunies : **fréq all et génotypiq st stables** au cours tps : **maintien du polymorphisme**.
- L'évolution étant définie par changement des fréq alléliques : **pop idéale n'évolue pas**. Pas de force évolutive faisant évoluer fréq des all (sélection nat, migration, mutation,...)
- L'équilibre H-W = équilibre **neutre**.
- Loi de H-W = **robuste** : conformité au modèle n'implique pas que ttes les conditions d'application de cette loi soient respectées (effectif infini, absence de mutation, de sélection, de migration, ...).
- Hypothèse la + import à respecter = **panmixie**.

Exemple : Population à l'équilibre ?



• 3 allèles : A_1, A_2, A_3
 • fréquences : p, q, r

gamètes mâles : A_1, A_2, A_3
 gamètes femelles : A_1, A_2, A_3

homozygotes : A_1A_1, A_2A_2, A_3A_3
 hétérozygotes : A_1A_2, A_1A_3, A_2A_3

$H_z = 1 - (p^2 + q^2 + r^2) = 2pq + 2pr + 2qr$

• 2 allèles, 3 phénotypes → phénotype à transmission co-dominante

Génotype	Effectifs génotypiques observés	Fréquences génotypiques observées
(BB)	200	0.2
(BJ)	600	0.6
(JJ)	200	0.2
N=1000		$\Sigma=1$

On dénombre dans un échantillon d'une population :

phénotype	génétype	effectifs
bleu	(BB)	200
vert	(BJ)	600
jaune	(JJ)	200

Déterminez :

- mode de transmission du phénotype
- fréquences phénotypiques
- fréquences génotypiques
- fréquences alléliques
- taux d'hétérozygotie

• $f(B) = 0.2 + \frac{1}{2} \times 0.6 = (2 \times 200 + 600) / (2 \times 1000) = 0.5$

• $f(J) = 1 - f(B) = 0.2 + \frac{1}{2} \times 0.6 = (2 \times 200 + 600) / (2 \times 1000) = 0.5$

• $H_z = 1 - [f(B)^2 + f(J)^2] = 2 \cdot f(B) \cdot f(J) = 0.5$

application (1) – correction sur l'intranet

Soit le locus bi-allélique « couleur » porté par un autosome et déterminant la couleur du pelage chez une espèce animale

Estimation des fréquences alléliques en cas de dominance d'un allèle

étude du groupe sanguin dans une population humaine du Pays Basque (Mourant 1976)

génétypes	(DD)	(Dd)	(dd)
phénotypes	[Rh-]	[Rh+]	[Rh-]
effectifs	230	230/400	170
fréquences	p^2	$2pq$	q^2

1. calculez les fréquences p et q des allèles D et d
 2. calculez les fréquences génotypiques

application (2)

Estimation des fréquences alléliques en cas de dominance d'un allèle

étude du groupe sanguin dans une population humaine du Pays Basque (Mourant 1976)

génétypes	(DD)	(Dd)	(dd)
phénotypes	[Rh-]	[Rh+]	[Rh-]
effectifs	230	170	

1. calculez les fréquences p et q des allèles D et d
 2. calculez les fréquences génotypiques

Test de conformité à H-W :

4 étapes :

1. Hypothèse (H₀) : **équilibre de H-W**
2. Calcul : **effectifs génotypiques observés (O_i)** et fréq all observées + **effectifs génotypiques attendus (C_i)** et fréq génotypiques attendues sous H₀
3. Comparaison effectifs observés / effectifs attendus pour chaque génotype → test stat χ^2
4. $\chi^2_{calculé}$ = comparé à valeur seuil (χ^2_{table}) en fonction de 2 paramètres :
 - un risque α choisi par l'utilisateur (5%)
 - **ddl = nb génotypes – nb allèles**

→ 2 allèles r et R, 3 génotypes rr, rR et RR ⇒ ddl = 3-2=1

Décision : Si $\chi^2_{calculé} < \chi^2_{table}$ ⇒ **Non rejet de H₀ → Equilibre de H-W**

Si $\chi^2_{calculé} \geq \chi^2_{table}$ ⇒ Rejet de H₀ (avec un risque α de se tromper)

Génotype	Effectifs génotypiques observés	Fréquences génotypiques attendues	Effectifs génotypiques attendus
(rr)	200	$p^2 = 0,25$	$N \cdot p^2 = 250$
(rR)	600	$2pq = 0,5$	$N \cdot 2pq = 500$
(RR)	200	$q^2 = 0,25$	$N \cdot q^2 = 250$
Total	N = 1000	$\Sigma = 1$	N = 1000

$$\chi^2_{calculé} = \frac{(200-250)^2}{250} + \frac{(600-500)^2}{500} + \frac{(200-250)^2}{250} \rightarrow \chi^2_{calculé} = 40$$

• ddl = 3 génotypes – 2 allèles = 1 ⇒ $\chi^2_{table} = 3,84$

$\alpha = 5\%$ (par convention)

d.d.l.	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0158	0,455	1,074	1,62	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	3,008	3,573	4,605	5,991	7,879	9,210	13,815

• $\chi^2_{calculé} (40) > \chi^2_{table} (3,84)$

• ⇒ on rejette l'hypothèse H₀ que la population est à l'équilibre de H-W (avec un risque α de se tromper)

• ≥ 1 des conditions d'application de la loi de H-W n'est pas respectée

Les forces évolutives

Facteurs **déterministes** : mutation, sélection, migration, non panmixie → **ft changer fréquences** all ds un sens donné, **prédictible**.

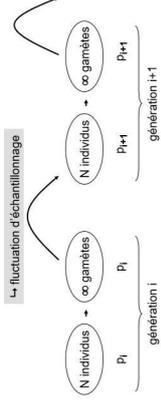
Facteurs **aléatoires** : dérive génétique → **impossible de prédire** sens variation fréquences.

→ Evolution des populations

Dérive génétique

Fluctuation aléatoire, d'une génération à autre, des fréq all. Origine : pop ≠ de grde taille / effectif non infini.

A partir d'une infinité de gam, échantillonnage de 2N gamètes



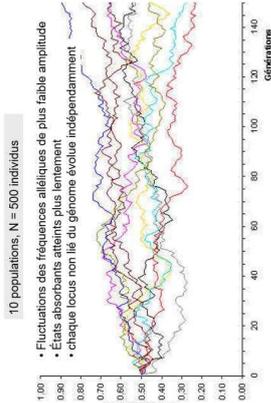
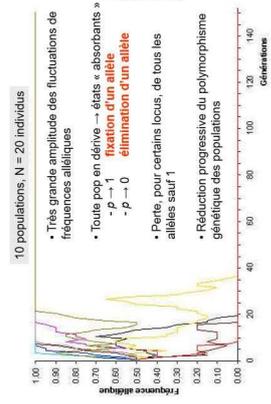
Un indiv peut : transmettre plusieurs copies du même gam à génération suivante ou ne transmettre aucune copie à génération suivante.

+ N petit, + proba d'exposition à fluctuations d'échantillonnage augmente.

Etat absorbant : fixation d'un all, élimination d'un autre.

dérive plus forte dans les petites populations

dérive plus faible dans les grandes populations



Conséquences et ptés :

Entre populations :

- * **Divergence génétique** sous l'effet du **hasard** uniquement (fluctuations d'échantillonnage)
- * **Populations séparées depuis longtps** : parfois jeux d'all ≠, au moins par leur fréq (différenciation génétique)

Au sein d'une population :

- * Fréq all changeant au cours tps sous effet du **hasard** uniquement
- * Changements ++ ds pop de **petits effectifs**
- * En absence de mutation / migration :
 - a. Ts all (d'un locus) sauf 1 seront perdus par dérive génétique
 - b. Diversité génétique de pop décroît
- * + grde diversité génétique ds grdes pop
- * Endogamie augmente ds pop, non par choix mais par hasard (à cause de petite taille pop)

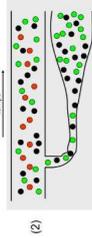
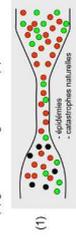
Effet fondateur

Changement aléat et soudain des fréq all, dû à **échantillonnage** d'indiv migrants.

⇒ Fondation d'une pop par petit échantillon de reproducteurs

Proba ++ que ces indiv ne soient pas représentatifs de pop d'orig. All rares = perdus le + svnt, les autres changent de fréq.

(1) goulet d'étranglement et (2) effet fondateur



Océanie.

Goulet d'étranglement démographique

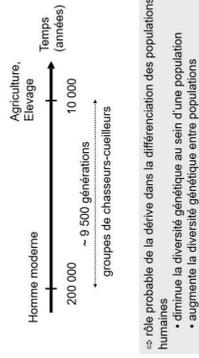
Ds Hist pop : événement peut provoquer décroiss de effectif de pop. Ex : épidémies, catastrophes nat. Réduction import pop, perte diversité génétique. Puis, crois → **redistribution fréq all**.

L'effet fondateur peut expliquer diffusion importante, au hasard, de cert all pourtant délétères, ds des pop particulières.

Exemple 1 : Amish d'Amérique du Nord

Pop constituée à partir ~30 individus fondateurs suisses-allemands, rejet progrès technologique, autarcie, pas de contact ac autres pop. 1 des fondateurs = syndrome de Ellis-van Creveld : maladie mendélienne AR (petite taille, doigts supplémentaires, pathologies cardiaques).

- Amish : fréq all délétère $q \sim 0,07$
- Autres pop : $q \sim 0,001$



→ Effet fondateur + consanguinité

Les mutations

Ti: **changement du matériel héréditaire**, sans aucune considération fonctionnelle. → **Evènement** ou processus (pas le produit)



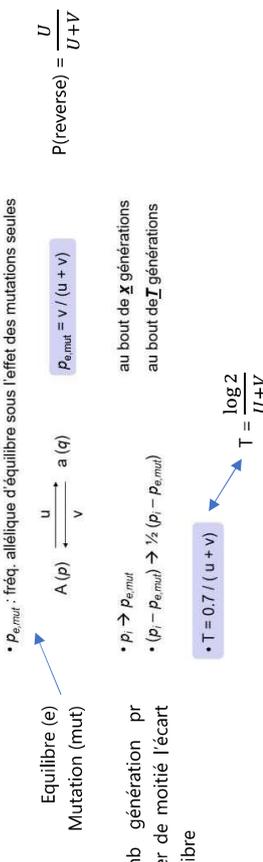
Tx mutation : nb mutations attendues/observées sur segment ADN (locus, gène étudié, site nt) pdt une période de tps (année, génération, réplication).

- Unique source de **nouvelles variations génétiques**, à l'éch de l'espèce. Autres forces : redistribution fréq d'all existants.
- Origines :
- **spontanées** (sauf exceptions : agents mutagènes)
 - **accidentelles** : erreurs de copie (ex : SNP, STR), tx mutation varie en fonction région.
 - **aléatoires** (= non dirigées) quant aux besoins de l'organisme → Triées ensuite par les autres forces

Vocabulaire : soit un SNP

- mutation : Evènement ou processus (pas le produit) ≠ *génétiq hum* : all délétère
- polymorphisme : + de 2 all ds pop, fréq all minoritaire $\geq 0,01 \neq$ *génétiq hum* : all sans effet fonctionnel

Mutations et vitesses d'évolution



T = nb génération pr diminuer de moitié l'écart à l'équilibre

Dans le vivant : $10^{-6} \leq$ tx mutation $\leq 10^{-5}$ (tx mutation SNP à échelle nt : 10^{-8})

Action très lente pr modifier les fréq all.

- Conséq faible tx de mutation pr substitutions nt :
- * Réversion : retour à l'allèle **ancestral** A → a
 - * Récurrence : le même évènement de mutation à un site donné se produit plusieurs fois indépendamment → négligeables à l'échelle des temps des populations (ça dépend des espèces ...)

• $\mu \times t$ (mutation)	2.10^{-5}	2.10^{-6}	2.10^{-7}
• $T(0.7^{(u+v)})$ (génération)	35.10^3	35.10^4	35.10^5
• Années (Drosophile)	35.10^2	35.10^3	35.10^4
• Années (Homme)	7.10^2	7.10^3	70.10^4

• *Drosophila simulans* : 1 génération = 52 jours
 • *Homo sapiens* : 1 génération = 20 ans

SNPs : considérés cō polymorphismes bi-all à l'échelle des pop d'une même espèce : A → T (€, €)

Résumé :

Mutations : unique source de **nouveauté génétique**

Tx de mutation si faibles que cette force a **peu d'effet** sur changement fréq all au cours des générations **à l'éch pop**

Sur de très longues périodes d'évolution, pop pourraient atteindre **équilibre** entre 'mutation' et 'mutation reverse'

Sélection naturelle

Théorie sélection naturelle (1859) : Charles Darwin - *formulation 'moderne'*

1. Ds ttes espèces, parents produisent en moy **+ de descendants** qu'il ne peut en survivre et se reproduire (Malthus)
2. Organismes diffèrent par **capacité à survivre** et se reproduire, en partie à cause **différences génotypiques** (phénotypiques au tps de Darwin).
3. A chaque génération, génotypes qui assurent survie ds envt du moment = présents en excès à l'âge reproducteur, et contribuent **proportionnellement plus** à descendance de la prochaine génération

→ Tri, au sein des pop, des indiv les mieux adaptés aux contraintes de l'environnement. Agit sur les phénotypes (indirectement sur les génotypes).

Sélection agit sur :

- * **Viabilité**
- * Sélection sexuelle : compét entre indiv pr obtenir partenaire accouplement
- * Sélection gamétique : agit sur phase haploïde du cycle de reproduction d'un organisme
- * **Fécondité**

Viabilité (V) : tx de survie d'un génotype, de sa naiss (zygote) à reproduction ; $0 \leq v \leq 1$

Fécondité (f) : prod moy de descendants / individu d'un génotype donné, à partir du moment où couple est formé (nb entier).

Valeurs adaptatives ou fitness

Fitness absolu (W) : **tx crois** d'un groupe *génétiq* déf d'individus au sein d'une pop au cours d'une ou plusieurs générations.

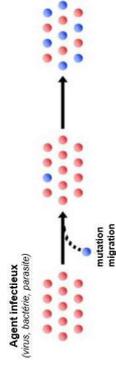
Nb **moyen de descendants fertiles** laissés à génération suivante par individu porteur d'un génotype donné.

$W_{\text{génotype}} = V_{\text{génotype}} \times f_{\text{génotype}}$

Ibid avec antibiotiques

Sélection fréquence-dépendante ou avantage du rare

Interaction syst immunitaire – agent infectieux. Arrivée niveau variant (mutation, migration) → moins bien reconnu par syst immunit : fréq augmente ds pop.



Avantage de l'hétérozygote – exemples

Mucoviscidose : mendélienne, AR. Locus impliqué : CFTR (ABCC7), all délétère ΔF508. Trouble fonctionnement des glandes exocrines. Prévalence : élevée ds pop européenne (F : ~1/4600)

CFTR = canal Cl à surf apicale des épithéliums. Personnes saines : sortie Cl- et eau par osmolarité → mucus (barrières, tapis roulant, élim poussières et pathog).

Si allèles délétères : protéine absente ou non fonctionnelle → mucus épais à la surface des épithéliums → fav infections pulm et viabilité diminuée. + stérilité « mécanique » (fécondité masculine diminuée) par obstruction et dégénérescence in utero des canaux déférents

ΔF508 = mutation délétère la + fréq (70%). Mutation survenue 1 seule fs il y a 20 000-50 000 ans, ds pop mère des pop europ. Fréq a augmenté par dérive génétique, maintenance par avantage hétérozygote.



Interaction drépanocytose / paludisme : mendélienne, AR. Locus impliqué : HBB (chaîne β de Hb), all délétère βS : p.Glu6Val. Anémie hémolytique

Fréq all βS = 20% ds cert région d'Afrique → très loin de l'équilibre attendu (devrait tendre vers 0).

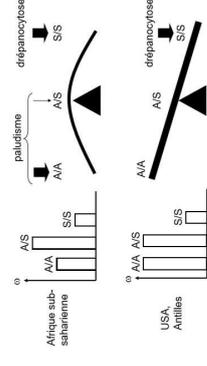
Années 50 : calcul fréq all cz enfants (gamètes) et adultes : O + effectifs attendus C. Enf = équilibre H-W, mais pas adultes → Force évolutive agit pdt dév't des zygotes jusqu'à âge adulte.

Forte concordance distribution all βS / paludisme à Plasmodium falciparum : en présence palu, hétérozygotes (AS) = avantages. Résistance « relative » au paludisme, mesurée à l'échelle « population »

- Hypotheses : 1. suppression crois ds hématies (A/S) par rate
- 2. Elimination accrue des hématies (A/S) parasitées
- 3. Reconnaiss + rapide des hématies (A/S) parasitées par syst immunitaire

Table 11.3 Frequency of hemoglobin genotypes in humans and calculation of their fitness

Genotype	Genotype		Total	Frequency of S allele
	AA	AS		
Number of individuals	18 (100)	8 (40)	26	0.31
Number of alleles	40 (400)	24 (480)	64	0.75

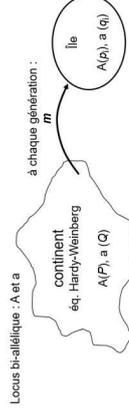


La migration

Mvt d'individus, ou parfois de pop entières, d'une région à une autre.

Flux génique : mvt de gènes d'une pop à une autre.

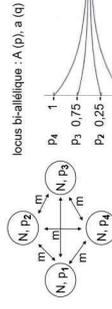
Modèle de l'île



$T = 0,7 / m$ $P_{eq, migration} = P$

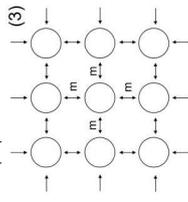
1. **Migrations unidirectionnelles**

Tx migration $m : 0 \leq m \leq 1$; proportion d'individus se reproduisant à chaque génération ds pop de l'île, mais qui sont issus de pop du continent.



2. **Mutations multidirectionnelles**

Uniformisation des fréq all qui convergent vers P_{moyen} (fréq moyenne des 4 pop).
Limite fortement différenciation génétique entre pop (s'oppose à dérive et sélection directionnelle) → introduction de nouveaux all dans une population



3. **Modèle en japonais** : migration + probable entre pop voisines → gradients de fréq all en fonction gradient de dist. Notion d'isolement par dist.

Bilan sur les migrations

Tx migration peuvent être élevés : **effets souvent import et rapides** sur fréq all (même à l'échelle d'1 ou qq générations !!)

Intensité des échanges migratoires = svr corrélée à **distance géo** entre pop.

Effet principal : uniformisation des fréquences alléliques des populations concernées. Limite fortement différenciation génétique entre populations, **s'oppose à dérive et sélection directionnelle** et s'applique à **tous les locus du génome**.

Introduction de **nouveaux allèles** dans une population

Exemple : dispersion des gènes de résistance aux agents anti-infectieux (paludisme – pyriméthamine).

Diffusion locale, régionale, continentale, intercontinentale du gène dhfr triple mutant de résistance à la pyriméthamine, sélection sur ensemble du continent African → abandon du traitement du paludisme simple par sulfadoxine-pyriméthamine.

Ecart à la panmixie

Reproduction **non aléatoire** : croisements privilégiés entre indiv → probablement lié à phénotype.

Type d'écart	Croisements privilégiés entre indiv	Fréquence génotypique	Fréquence allélique	Locus concernés
Homogamie	De même phénotype	f (Aa) < 2pq ↗ homozygotes	Inchangée	Contrôlant le phénotype (ne concernent que locus impliqués ds choix conjoint)
Hétérogamie	De phénotypes différents	f (Aa) > 2pq ↗ hétérozygotes	Inchangée	
Apparentement (endogamie)	Apparentés	f (Aa) < 2pq	Inchangée	Tout le génome
Exogamie	Non apparentés	f (Aa) > 2pq	Inchangée	

BILAN / comment les forces évolutives augmentent (+) ou diminuent (-) diversité génétique

	Au sein d'une population	Entre populations
Dérive génétique	-	+
Consanguinité	-	+
Mutation	+	+
Migration		-
Sélection directionnelle	- (si environnement cst)	+/- (envts différents entre pop)
Sélection balancée	+	+/- (intensité sélection varie)