



Génétique humaine : variants, transmission, méiose et méthodes

OSSIK • COURS 05-02 • LE CORRIGÉ • RÉVISION 0.5

PASS - majeure santé - classement d'après la maquette publique de Paris Cité, sans lien avec l'université. Cours original rédigé avec une IA. Le cours de ta faculté fait foi.

Sommaire du cours

Partie 1 • Variants, conséquences et interprétation	1
Partie 2 • Transmission, pedigrees et probabilités	8
Partie 3 • Méiose, recombinaison et anomalies chromosomiques	15
Partie 4 • PCR, séquençage et analyse expérimentale	23

Partie 1 • Variants, conséquences et interprétation

Prérequis et objectifs

Connaître gène, épissage et code génétique. Le cours distingue description d'un changement d'ADN, conséquence moléculaire et effet phénotypique. Une différence de séquence n'est pas automatiquement pathogène ; un mécanisme plausible n'est pas une preuve suffisante.

1. Décrire le changement

Une substitution remplace un nucléotide. Une insertion ajoute de la séquence ; une délétion en retire. Une duplication, une inversion ou une translocation concerne une organisation plus large. Le terme variant décrit une différence par rapport à une référence.

Une transition échange deux purines ou deux pyrimidines ; une transversion échange une purine et une pyrimidine. Cette classification renseigne sur le changement chimique, pas directement sur sa gravité.

La description exige une référence et une orientation. Une même variation peut être décrite différemment selon le transcrit choisi, notamment lorsque plusieurs isoformes existent. Un numéro sans référence n'est pas une identité complète du variant.

2. Variants codants

Un variant synonyme conserve l'acide aminé dans le code utilisé. Il peut néanmoins modifier l'épissage, la stabilité de l'ARN ou la cinétique de traduction selon son contexte. « Synonyme » ne signifie donc pas « biologiquement toujours neutre ».



Un faux-sens remplace un acide aminé. Son effet dépend de la position, du rôle du résidu, de la différence chimique et de l'environnement structural. Remplacer une glycine dans une zone contrainte peut avoir un effet différent du même remplacement dans une région flexible.

Un non-sens crée un codon stop prématuré. Il peut produire une protéine tronquée ou déclencher une dégradation de l'ARN via des mécanismes de surveillance, selon la position et le contexte. La longueur de protéine prédite n'est pas toujours celle réellement observée.

Une insertion ou délétion dont la longueur n'est pas multiple de trois peut décaler le cadre de lecture dans une séquence codante. Une variation en phase conserve le cadre, mais peut retirer un domaine essentiel et rester sévère.

3. Variants non codants

Un variant de promoteur peut changer l'expression ; un variant d'élément régulateur distal peut affecter un gène éloigné. Un variant intronique peut créer ou abolir un site d'épissage. Les effets non codants demandent donc de relier position et mécanisme.

Une variation dans une région non traduite peut modifier stabilité, localisation ou efficacité de traduction de l'ARNm. L'absence de changement d'acide aminé ne suffit pas à conclure à une absence d'effet.

Inversement, la proximité d'un gène associé à une fonction ne prouve pas que tout variant voisin agit sur lui. La régulation peut être spécifique d'un tissu, d'un stade ou d'une architecture chromatinienne.

4. Quantité et mode d'action

Une perte de fonction réduit l'activité ou la quantité utile du produit. Une haplo-insuffisance apparaît lorsqu'une copie fonctionnelle ne suffit pas pour la fonction normale dans le contexte considéré.

Un effet dominant négatif implique qu'un produit altéré perturbe le produit normal ou un complexe. Une simple absence de protéine ne produit pas nécessairement le même effet qu'une protéine inactive mais incorporable dans un assemblage.

Un gain de fonction peut augmenter une activité, la rendre constitutive ou lui donner une nouvelle propriété. Les modes dominant et récessif de transmission ne se déduisent donc pas uniquement du mot « mutation » ni de la taille du changement.

5. Constitutionnel, somatique et mosaïque

Un variant constitutionnel est présent largement depuis les premières étapes du développement, avec des nuances selon l'histoire biologique. Un variant somatique apparaît dans une lignée cellulaire et peut être absent d'autres tissus.

Un mosaïcisme signifie la coexistence de populations cellulaires génétiquement différentes issues d'un même zygote. La fraction d'allèle mesurée dans un prélèvement dépend de la proportion de cellules, de la ploïdie, du nombre de copies et de la technique.

Une fraction de variant à 20 % ne signifie pas automatiquement que 20 % des cellules sont hétérozygotes : dans un tissu diploïde simple sans autre anomalie, ce serait plutôt compatible avec environ 40 % de cellules hétérozygotes. Mais ce calcul change dès que les hypothèses changent.

6. Construire une interprétation

Les données utiles peuvent inclure fréquence dans des populations, ségrégation familiale, conséquence attendue, études fonctionnelles et cohérence avec le phénotype. Aucune ligne ne suffit dans tous les cas.

Une prédiction informatique fournit une hypothèse probabiliste ou un score selon son modèle ; elle ne démontre pas une pathogénicité. Une association statistique n'identifie pas automatiquement le variant causal ni son mécanisme.

Un test fonctionnel doit être pertinent pour le gène, le tissu et le mécanisme. Une surexpression artificielle peut masquer un défaut de stabilité ou créer un effet absent aux concentrations physiologiques. Une restauration de fonction peut renforcer l'argument si les contrôles sont adaptés.

Niveau	Exemple	Limite
Séquence	Substitution détectée	Ne donne pas l'effet
ARN	Épissage modifié	Ne donne pas toute la fonction
Protéine	Activité diminuée	Dépend du test
Organisme	Phénotype associé	Causalité à établir

Interpréter un variant

1. Décrire avec une référence
2. Identifier le transcrit
3. Prévoir les mécanismes
4. Chercher des preuves indépendantes
5. Conserver l'incertitude restante

Le schéma encadre un raisonnement pédagogique, pas une classification clinique personnelle.

Développement : causalité, phase allélique et mosaïcisme

Cis et trans

Deux variants détectés dans un même gène peuvent se trouver sur la même copie, en cis, ou sur les deux copies homologues, en trans. Dans une maladie récessive, cette distinction peut être décisive pour savoir si les deux allèles sont affectés.

Un séquençage listant deux variants sans information de phase ne répond pas automatiquement à cette question. Les données parentales, des lectures couvrant les deux positions ou d'autres méthodes peuvent aider à établir la phase.

Deux variants en cis peuvent laisser l'autre copie fonctionnelle. Deux variants en trans peuvent affecter chacune des copies, mais leur conséquence dépend encore de leur effet réel. Le nombre de variants ne suffit pas à prédire le génotype fonctionnel.

Perte de fonction et dosage

Une délétion d'un gène peut réduire sa quantité, mais certaines fonctions tolèrent une copie restante. Pour d'autres, la quantité devient insuffisante. La sensibilité au dosage appartient au gène et au contexte, pas à la seule taille de la délétion.

Un variant faux-sens peut rendre une protéine instable et se comporter comme une perte de fonction, ou produire une protéine stable perturbatrice. Deux modifications du même gène peuvent donc suivre des modes d'action différents.

Un variant non-sens tardif peut échapper à certaines surveillances de l'ARN et produire une protéine tronquée. Il ne faut pas affirmer que tout non-sens équivaut exactement à l'absence totale de produit.

Effet dominant négatif dans un dimère

Dans un modèle simplifié où des sous-unités normales et altérées s'assemblent aléatoirement en proportions égales, on obtient un quart de dimères normaux-normaux, une moitié de dimères mixtes et un quart altérés-altérés.

Si une seule sous-unité altérée suffit à inactiver le dimère, seulement un quart des complexes fonctionne, malgré une moitié de sous-unités normales. Ce calcul illustre le mécanisme dominant négatif.

Il repose sur des hypothèses fortes : expression égale, assemblage aléatoire, stabilité égale et inactivation par une seule sous-unité. Une expérience réelle doit les vérifier ; le chiffre 25 % n'est pas une règle universelle de dominance.

Fraction allélique et nombre de copies

Dans une cellule diploïde hétérozygote, la fraction théorique de lectures variantes est proche de 50 % si les deux allèles sont amplifiés et séquencés sans biais. Une mosaïque de cellules hétérozygotes à fraction f donne environ $f/2$.

Mais une duplication, une perte d'allèle ou un mélange de populations de ploïdies différentes modifie la relation. Une fraction à un tiers peut provenir de plusieurs configurations et ne doit pas être traduite automatiquement en un pourcentage de cellules.

Les biais d'amplification peuvent aussi favoriser un allèle. La profondeur et l'incertitude d'échantillonnage comptent : cinq lectures sur dix sont bien moins précises que cinq cents sur mille.

Ségrégation familiale

Un variant présent chez plusieurs personnes atteintes et absent chez des personnes non atteintes peut soutenir une hypothèse causale. La force de cet argument dépend de la taille de famille, du mode de transmission, de la pénétrance et des âges.

Un individu non atteint trop jeune pour exprimer un phénotype tardif ne constitue pas le même contre-argument qu'un individu âgé clairement évalué. La définition du phénotype doit être cohérente.

Une famille petite peut montrer une concordance par hasard. La ségrégation est une preuve parmi d'autres, pas un verdict automatique.

Tests fonctionnels

Un essai peut mesurer une activité catalytique, une localisation, un épissage ou une interaction. Il doit tester le mécanisme attendu. Une activité enzymatique normale dans un tube n'exclut pas un défaut de trafic ou de stabilité dans la cellule.

Une expérience de restauration doit comparer des niveaux d'expression adaptés. Une surexpression massive peut compenser artificiellement une faible activité et donner l'impression d'un variant sans effet.

L'utilisation de plusieurs lectures indépendantes réduit le risque de confondre un artefact de test avec une propriété du variant.

Incertitude explicite

Une interprétation peut rester indéterminée malgré une description précise de la séquence. Conserver cette incertitude est scientifiquement correct. Elle peut être réduite par nouvelles données, mais ne doit pas être remplacée par une conclusion catégorique pour rendre le dossier plus simple.

Cas d'intégration : fréquence allélique et mosaïcisme

Un variant autosomique est hétérozygote dans une fraction des cellules d'un prélèvement, tandis que les autres cellules ne le portent pas. Sous l'hypothèse de deux copies du locus dans toutes les cellules et d'une lecture sans biais, une fréquence allélique de 15 % est mesurée. Estimer la fraction de cellules porteuses et préciser les limites de cette inférence.

Résolution et discussion

Chaque cellule porteuse contribue une copie variante parmi deux copies, soit une fraction de 50 % dans cette sous-population. Si les cellules porteuses représentent une fraction f de toutes les cellules, la fréquence allélique globale vaut $f/2$. On obtient donc $f = 0,30$: environ 30 % de cellules porteuses dans le modèle annoncé.

Cette conversion ne fonctionne pas automatiquement en présence d'une variation du nombre de copies, d'une perte d'hétérozygotie ou d'un biais de séquençage. Une composition tissulaire hétérogène peut également modifier l'interprétation : la fraction observée décrit le prélèvement, pas nécessairement tous les organes de la personne. Le mosaïcisme est spatial autant que quantitatif.

Une fréquence faible doit aussi être distinguée d'une erreur technique. Profondeur de lecture, qualité des bases, équilibre des orientations et confirmation par une méthode adaptée participent à l'évaluation. Cent lectures avec quinze observations variantes et dix mille lectures avec mille cinq cents observations ont le même pourcentage, mais pas la même incertitude d'échantillonnage.

Application : une fraction allélique d'un tiers n'impose pas un mosaïcisme

Supposons que toutes les cellules d'un prélèvement portent trois copies d'un locus, dont une porte le variant et deux la séquence de référence. Si les copies sont lues avec la même efficacité, la fraction allélique attendue est $1/(1 + 2) = 1/3$, soit environ 33 %.

Dans un autre modèle diploïde, deux tiers des cellules sont hétérozygotes et un tiers ne portent pas le variant. La fraction variante attendue est alors $(2/3) \times (1/2) = 1/3$ également. Une même fraction peut



donc correspondre à deux organisations cellulaires très différentes.

La lecture de séquence seule ne départage pas ces modèles. Une information sur le nombre de copies, la composition du prélèvement et les biais d'amplification est nécessaire. Cet exemple porte sur le nombre de copies d'un locus ; il ne signifie pas obligatoirement une triploïdie de tout le génome.

Approfondissement v0.4 : pondérer les copies et tester un mécanisme dominant négatif

Compter les allèles avant de calculer leur fraction

Dans un mélange, une fraction f des cellules possède C copies du locus, dont m variantes ; les autres cellules sont diploïdes sans variant. Sous extraction et lecture équilibrées par copie, le numérateur est fm et le dénominateur $fC + 2(1-f)$. La fraction allélique vaut donc $V = fm/[fC + 2(1-f)]$. La formule $f/2$ n'est retrouvée que pour $C = 2$ et $m = 1$.

Avec $f = 0,40$, $C = 3$ et $m = 1$, le numérateur vaut $0,40$ et le dénominateur $2,40$: $V = 1/6$, soit $16,7\%$. Appliquer à tort $f = 2V$ donnerait $33,3\%$ de cellules au lieu des 40% imposés. Le nombre de copies change la contribution de chaque population au total des molécules.

Inverser seulement si les paramètres sont connus

La relation précédente conduit à $f = 2V/[m - V(C-2)]$. Avec $V = 0,20$, $C = 3$ et $m = 1$, elle donne $f = 0,40/0,80 = 0,50$. Mais sans connaître C et m , une même fraction V admet plusieurs explications. Une mesure très précise de V ne résout pas ce manque d'identifiabilité.

Cette écriture suppose une population porteuse homogène. Plusieurs sous-populations, des pertes d'allèle ou des différences d'extraction demandent une somme pondérée plus complète. On ne doit pas convertir un pourcentage de lectures en pourcentage de cellules avant d'annoncer le modèle biologique.

Étendre l'assemblage aléatoire à un oligomère

Si la fraction de sous-unités normales est p et si un complexe de n sous-unités n'est actif que lorsque toutes sont normales, la fraction active théorique est p^n . Pour $p = 1/2$ et un tétramère, elle vaut $1/16$, soit $6,25\%$. Ce chiffre ne découle pas du seul mot « hétérozygote » : il suppose proportions protéiques égales, assemblage indépendant, stabilité comparable et inactivation par une seule sous-unité altérée.

Si au moins trois sous-unités normales suffisent dans un tétramère, la fraction active devient $p^4 + 4p^3(1-p)$. Pour $p = 1/2$, elle vaut $5/16$, soit $31,25\%$. Changer la règle fonctionnelle change donc fortement la prédiction sans changer les quantités de sous-unités.

Utiliser le modèle pour concevoir une expérience

Mesurer l'expression de chaque version, leur incorporation au complexe et l'activité selon leur proportion permet de tester la règle proposée. Une baisse d'activité supérieure à 50% peut suggérer une interaction perturbatrice, mais ne la prouve pas seule : stabilité, localisation ou quantité totale peuvent également intervenir.



Exercices progressifs

1. Un variant synonyme est-il toujours sans effet ?
2. Une délétion de six nucléotides décale-t-elle nécessairement le cadre ?
3. Quelle différence entre haplo-insuffisance et dominant négatif ?
4. Dans un mélange diploïde simple, 30 % de cellules hétérozygotes donnent quelle fraction d'allèle variant attendue ?
5. Un score informatif défavorable suffit-il à classer un variant comme causal ?
6. Dans le modèle à trois copies dans toutes les cellules, quelle fraction attend-on si deux copies portent le variant ?
7. Une population contient 25 % de cellules avec quatre copies du locus, dont deux variantes, et 75 % de cellules diploïdes sans variant. Calculer V. Puis calculer la fraction de tétramères entièrement normaux si $p = 0,75$.

Corrections argumentées

1. Non. Il peut agir sur épissage ou régulation de l'ARN.
2. Non. Six est multiple de trois ; le cadre peut être conservé, sans garantir une protéine fonctionnelle.
3. La première manque de quantité fonctionnelle ; le second implique une perturbation active par le produit altéré.
4. Environ 15 %, car chaque cellule hétérozygote porte une copie variante sur deux.
5. Non. Il faut intégrer des preuves indépendantes et les limites du modèle.
6. $2/3$, soit environ 67 %, sous hypothèse de lecture équilibrée. Une fraction différente de 50 % ne suffit donc pas à conclure à un artefact ou à un mosaïcisme sans examiner les copies du locus.
7. Copies variantes moyennes : $0,25 \times 2 = 0,50$. Copies totales : $0,25 \times 4 + 0,75 \times 2 = 2,50$. $V = 0,20$. Pour l'assemblage indépendant, $p^4 = 0,75^4 = 0,3164$, soit 31,64 %. Les deux calculs portent sur des niveaux différents : copies d'ADN et assemblage protéique.

Références et périmètre

NHGRI, Mutation, glossaire génomique. Exemples pédagogiques ; aucune classification clinique individualisée n'est établie.

Références du complément v0.4, consultées le 2 octobre 2026 : NHGRI, génétique et génomique. Exemples et calculs originaux.

Partie 2 · Transmission, pedigrees et probabilités

Prérequis et objectifs

Connaître allèles, chromosomes homologues et probabilités simples. Le cours apprend à construire un risque sous hypothèses explicites, conditionner une probabilité et lire un arbre sans surinterpréter. Les exemples sont des modèles pédagogiques, non un conseil génétique.

1. Modèle autosomique récessif

Pour un locus autosomique à deux allèles A et a, un croisement $Aa \times Aa$ donne $1/4 AA$, $1/2 Aa$ et $1/4 aa$ si la ségrégation est mendélienne et les fécondations comparables. Dans le modèle récessif à pénétrance complète, seuls aa sont atteints.

Chaque grossesse est un nouvel événement sous les mêmes hypothèses. Avoir déjà eu un enfant atteint ne consomme pas le « quart de risque » et ne garantit pas que les suivants seront non atteints.

Un enfant non atteint de deux parents Aa n'a pas un risque de portage de $1/2$ après prise en compte de cette information. Parmi les trois quarts non atteints, deux quarts sont Aa : le risque conditionnel vaut $(1/2)/(3/4) = 2/3$.

2. Autosomique dominant

Dans un modèle simple Aa atteint $\times$ aa non atteint, chaque enfant reçoit a ou A du parent hétérozygote avec probabilité $1/2$. La probabilité de porter A est distincte de la probabilité de manifester le phénotype si la pénétrance n'est pas complète.

La pénétrance est la proportion de porteurs du génotype qui expriment le phénotype défini dans les conditions considérées. L'expressivité décrit la variation de manifestation parmi les personnes qui l'expriment. Une pénétrance incomplète peut faire paraître une génération « sautée ».

Un variant de novo peut apparaître sans antécédent familial apparent. Un arbre négatif ne prouve donc pas une origine non génétique.

3. Transmission liée au chromosome X

Dans le modèle récessif lié à X, un garçon reçoit son X de sa mère et son Y de son père. Un père ne transmet pas son X à ses fils. Une mère hétérozygote transmet l'un ou l'autre de ses X avec probabilité $1/2$.

Pour une mère conductrice et un père non atteint, sous modèle classique, la moitié des fils sont atteints et la moitié des filles sont conductrices. Le risque pour un enfant de sexe non précisé inclut la probabilité du sexe : environ $1/4$ d'enfant atteint si seuls les garçons sont atteints et les sexes équiprobables.

Les manifestations chez des personnes possédant deux X peuvent être modulées par l'inactivation du X et d'autres facteurs. Les arbres simples sont des modèles, pas une description de chaque situation réelle.

4. Probabilités conditionnelles

La formule de Bayes met à jour une probabilité avec une information. Une écriture sûre est $P(A|B) = P(A \text{ et } B)/P(B)$. Il faut nommer les événements avant de multiplier des fractions.

Si une personne non atteinte issue de $Aa \times Aa$ a un risque $2/3$ d'être porteuse, et que son partenaire est Aa certain, le risque d'un enfant aa vaut $(2/3) \times 1 \times (1/4) = 1/6$. Les conditions « même gène », « même modèle » et « pénétrance complète » sont indispensables.

Pour plusieurs enfants, les événements peuvent être indépendants conditionnellement au génotype des parents. Si ce génotype est incertain, les observations des premiers enfants peuvent modifier la probabilité parentale et donc les calculs suivants. L'indépendance doit être formulée au bon niveau.

5. Hardy-Weinberg

Pour deux allèles de fréquences p et q , $p + q = 1$. Sous les hypothèses du modèle, les fréquences génotypiques sont p^2 , $2pq$ et q^2 . Le modèle sert de référence pour une population, pas de loi automatiquement applicable à toute famille.

Si une maladie récessive rare à pénétrance complète a une fréquence $q^2 = 1/10000$, alors $q = 0,01$, $p = 0,99$ et le portage vaut $2pq = 0,0198$, soit 1,98 %. Confondre q avec q^2 sous-estimerait fortement le nombre de porteurs.

Unions non aléatoires, structure de population, sélection ou autres mécanismes peuvent modifier les proportions. Un calcul utilisant une fréquence globale peut être inadéquat pour un sous-groupe sans information supplémentaire.

6. Modes plus complexes

Une transmission mitochondriale est principalement maternelle dans le modèle humain classique ; hétéroplasmie et répartition des copies modifient les manifestations. Les affections dues à des gènes nucléaires codant des protéines mitochondriales ne suivent pas nécessairement ce mode.

L'empreinte parentale fait dépendre l'expression de l'origine parentale de certains allèles. Les traits multifactoriels résultent de plusieurs variants et facteurs environnementaux ; ils ne se résument pas à un carré de Punnett monogénique.

Un arbre vertical suggère parfois une transmission dominante, et une fratrie atteinte avec parents non atteints suggère parfois un modèle récessif. Ces motifs sont des indices : taille de famille, hasard, pénétrance et mosaïcisme peuvent produire des apparences similaires.

Calculer un risque pédagogique

1. Définir le modèle • 2. Décrire les génotypes possibles • 3. Conditionner sur les informations • 4. Multiplier les étapes pertinentes • 5. Vérifier les hypothèses

Le schéma ne remplace pas une consultation ou une interprétation de variants dans une situation personnelle.



Développement : plusieurs enfants, test négatif et hypothèses de population

Probabilité d'au moins un enfant atteint

Pour deux parents Aa dans le modèle récessif simple, le risque par enfant est $1/4$. La probabilité que trois enfants soient tous non atteints est $(3/4)^3 = 27/64$.

La probabilité d'au moins un enfant atteint est donc $1 - 27/64 = 37/64$, environ 57,8 %. Additionner trois fois $1/4$ donnerait 75 % et compterait plusieurs fois les situations où deux ou trois enfants sont atteints.

La probabilité d'exactly un enfant atteint parmi trois vaut $3 \times (1/4) \times (3/4)^2 = 27/64$. Le facteur trois représente les trois positions possibles de l'enfant atteint. « Au moins un » et « exactement un » sont des événements différents.

Conditionner sur un test négatif

Supposons un risque préalable de portage de $2/3$, un test détectant 90 % des porteurs et une spécificité supposée de 100 % dans cet exemple. Un porteur a 10 % de probabilité de test négatif ; un non-porteur a 100 %.

La probabilité d'être porteur et négatif vaut $(2/3) \times 0,10 = 1/15$. La probabilité totale de négatif vaut $1/15 + 1/3 = 2/5$. Après un test négatif, le risque de portage vaut $(1/15)/(2/5) = 1/6$.

Un résultat négatif diminue donc le risque sans nécessairement l'annuler. La valeur dépend de la sensibilité réelle, de la cible du test et du risque préalable. Les nombres ici sont pédagogiques et ne doivent pas être appliqués à un test clinique non décrit.

Dépendance entre enfants lorsque le parent est incertain

Si le génotype des parents est connu, les transmissions successives sont indépendantes dans le modèle. Si l'un des parents a seulement une probabilité d'être porteur, les observations de la famille fournissent de l'information sur ce génotype.

La naissance d'un enfant atteint peut rendre le portage parental certain dans un modèle récessif simple sans autre explication. Les risques suivants se recalculent alors avec cette information.

Il est donc incorrect de dire sans nuance que « tous les événements familiaux sont indépendants ». Ils sont indépendants conditionnellement aux états biologiques pertinents, mais partagent une incertitude commune sur ces états.

Dominance et homozygotie

Un phénotype dominant chez un parent ne dit pas toujours s'il est hétérozygote ou homozygote. Le modèle $Aa \times aa$ donnant $1/2$ exige d'abord l'hypothèse d'hétérozygotie.

Certaines situations rendent un génotype peu compatible avec les observations, mais cette exclusion doit être justifiée. Une dominance complète au niveau du phénotype ne décrit pas forcément la viabilité ou la sévérité de l'homozygote.

La pénétrance peut dépendre de l'âge et de facteurs modificateurs. Un chiffre unique simplifie une réalité qu'il faut préciser lorsqu'on résout un problème.

Hardy-Weinberg et porteurs non atteints

Si $p = 0,99$ et $q = 0,01$, la fréquence des hétérozygotes dans toute la population vaut 0,0198. Parmi les non-atteints d'une affection récessive complètement pénétrante, la fréquence conditionnelle vaut $2pq/(1 - q^2)$, presque identique ici mais conceptuellement distincte.

Pour un allèle moins rare, la différence peut devenir plus visible. L'approximation $2q$ suppose p proche de un ; elle ne doit pas être appliquée si q est élevé.

Les fréquences peuvent varier entre populations. Utiliser une fréquence agrégée pour un groupe particulier sans justification peut biaiser le risque. Le modèle de population doit correspondre à l'information disponible.

Liaison génétique

Pour deux loci liés, multiplier indépendamment les probabilités d'allèles peut être faux. Il faut connaître les haplotypes parentaux et la fréquence de recombinaison.

Un parent AB/ab avec $r = 0,10$ produit, dans le modèle symétrique simple, 45 % AB, 45 % ab, 5 % Ab et 5 % aB. Les quatre gamètes ne sont pas équiprobables. Le calcul de risque dépend donc de l'organisation des allèles sur les chromosomes.

Présenter une réponse

Une réponse complète donne les hypothèses, l'événement recherché, la formule et l'unité ou le pourcentage final. Écrire seulement « $1/6$ » ne montre pas si l'on parle de portage parental, de transmission d'allèle ou de manifestation chez l'enfant.

Cas d'intégration : conditionner sur une information nouvelle

Deux parents hétérozygotes pour une affection autosomique récessive ont un enfant dont on sait qu'il n'est pas atteint. Dans le modèle mendélien à pénétrance complète et sans autre mécanisme, quelle est sa probabilité d'être porteur hétérozygote ? Pourquoi la réponse n'est-elle plus simplement une moitié ?

Résolution et discussion

Avant l'information clinique, les probabilités des génotypes AA, Aa et aa sont respectivement $1/4$, $1/2$ et $1/4$. Le fait de savoir l'enfant non atteint exclut aa dans les hypothèses annoncées. L'espace des possibilités restantes possède une probabilité totale de $3/4$. La probabilité conditionnelle d'être Aa vaut donc $(1/2)/(3/4) = 2/3$.

Le numérateur décrit l'événement compatible recherché ; le dénominateur décrit toutes les situations compatibles avec l'information disponible. Cette normalisation est indispensable. Supprimer simplement le cas atteint sans renormaliser laisserait des probabilités dont la somme serait inférieure à un.

L'observation concernant cet enfant ne modifie pas automatiquement la probabilité mendélienne pour une nouvelle conception, si les génotypes parentaux sont connus et si les événements de transmission sont indépendants. Le prochain enfant conserve ici un risque de $1/4$ d'être aa. Probabilité conditionnelle pour une personne déjà observée et probabilité pour une conception future répondent à deux questions différentes.

Si la pénétrance n'était pas complète ou si l'affection ne s'exprimait qu'après un certain âge, être actuellement non atteint n'exclurait plus nécessairement aa. Il faudrait introduire les probabilités d'observation selon chaque génotype. Le résultat 2/3 est donc une conséquence d'un modèle précis, pas une règle universelle applicable à toute fratrie comportant une maladie récessive.

Application : actualiser un modèle à pénétrance incomplète

Un enfant a une probabilité initiale de 1/2 d'avoir reçu un allèle dominant. Dans un modèle fictif, 80 % des porteurs expriment le caractère avant un âge donné, aucun non-porteur ne l'exprime et tous les autres facteurs sont exclus. L'enfant n'exprime pas le caractère à cet âge.

La probabilité de ce constat chez un porteur vaut 0,20 ; chez un non-porteur, elle vaut 1. Bayes donne $P(\text{porteur} \mid \text{non manifestant}) = (0,5 \times 0,20) / (0,5 \times 0,20 + 0,5 \times 1) = 1/6$. L'absence de manifestation diminue la probabilité, sans la rendre nulle.

Si cette personne a un enfant avec un non-porteur, le risque de transmettre cet allèle dans le modèle vaut $(1/6) \times (1/2) = 1/12$. Le risque de manifestation avant le même âge vaut encore $1/12 \times 0,80 = 1/15$. Ces chiffres illustrent les hypothèses fournies ; ils ne sont pas des estimations pour une affection réelle.

Approfondissement v0.4 : génotype latent et observations successives

Conditionner avant de multiplier

Une personne a une probabilité 2/3 d'être Aa et 1/3 d'être AA ; son partenaire est Aa certain. Dans le modèle récessif à pénétrance complète, un enfant est atteint avec probabilité 1/4 si le premier parent est Aa et zéro s'il est AA. Le risque marginal du premier enfant est donc 1/6.

Deux enfants partagent le même génotype parental inconnu. La probabilité qu'ils soient tous deux atteints vaut $(2/3) \times (1/4)^2 = 1/24$. Elle n'est pas $(1/6)^2 = 1/36$. La multiplication par 2/3 ne doit être faite qu'une fois, car on ne tire pas un nouveau génotype parental à chaque conception.

Actualiser après deux enfants non atteints

La probabilité de deux enfants non atteints vaut $(2/3)(3/4)^2 + (1/3) \times 1 = 17/24$. Parmi ces familles compatibles, la part provenant d'un parent Aa vaut $(2/3)(3/4)^2 = 3/8$. La probabilité actualisée de portage est donc $(3/8)/(17/24) = 9/17$.

Le risque pour un troisième enfant devient $(9/17) \times (1/4) = 9/68$, soit environ 13,2 %. L'observation n'a pas changé la mécanique mendélienne chez un parent Aa : le risque conditionnel reste 1/4. Elle a changé notre information sur le génotype du parent.

Au moins un et exactement un

Dans ce même modèle, la probabilité d'au moins un enfant atteint parmi trois vaut $(2/3)[1 - (3/4)^3] = 37/96$. La probabilité d'exactement un vaut $(2/3) \times 3 \times (1/4)(3/4)^2 = 9/32$. La branche AA ne contribue à aucun de ces événements mais doit rester dans les calculs d'événements tels que « aucun atteint ».

Une fréquence conditionnelle n'est pas une causalité

Ces calculs supposent un seul locus, un partenaire réellement Aa, une pénétrance complète et aucune autre cause du phénotype. Si l'observation est imparfaite ou dépend de l'âge, il faut remplacer « non atteint » par un événement observable avec sa probabilité propre dans chaque génotype. Un arbre de probabilités permet de conserver cette distinction.

La même méthode s'applique à une incertitude partagée dans d'autres domaines : choisir d'abord l'état latent, calculer les observations conditionnelles dans chaque branche, puis sommer et renormaliser. Le raccourci qui multiplie des probabilités marginales n'est valable que lorsque l'indépendance marginale est justifiée. Ici, les enfants sont indépendants conditionnellement aux génotypes, mais pas avant de connaître ceux-ci.

Exercices progressifs

1. Pour Aa × Aa, quelle probabilité d'un enfant Aa avant toute information ?
2. Même croisement : quelle probabilité de portage si l'enfant est non atteint ?
3. Un père transmet-il son chromosome X à son fils dans le modèle XY classique ?
4. Sous Hardy-Weinberg, $q = 0,02$ donne quelle fréquence aa ?
5. Si un allèle dominant est transmis avec probabilité $1/2$ et possède une pénétrance de 80 %, quel risque de manifestation selon ce modèle ?
6. Pourquoi multiplier directement $1/2$ par 0,80 ne répond-il pas à la question après observation de l'absence de manifestation ?
7. Après un seul enfant non atteint dans ce modèle, calculer la probabilité parentale de portage puis le risque du deuxième enfant.

Corrections argumentées

1. $1/2$.
2. $2/3$, après exclusion des aa.
3. Non. Il lui transmet son Y ; le X vient de la mère.
4. $q^2 = 0,0004 = 0,04 \%$, soit $1/2500$.
5. $0,5 \times 0,8 = 0,4$, soit 40 %, sous les hypothèses précisées et en excluant d'autres causes.
6. Le produit 0,40 est la probabilité initiale d'avoir hérité de l'allèle et de manifester le caractère. Après observation, il faut conditionner sur les non-manifestants et renormaliser leurs probabilités. Ce changement d'univers est précisément le rôle du dénominateur de Bayes.
7. $P(\text{non atteint}) = (2/3)(3/4) + 1/3 = 5/6$. $P(\text{Aa et non atteint}) = 1/2$. Le portage actualisé vaut $(1/2)/(5/6) = 3/5$. Le risque du prochain enfant est $(3/5)(1/4) = 3/20 = 15 \%$. Il faut actualiser une fois le génotype latent, puis appliquer la transmission.



Référence et périmètre

NHGRI, glossaire de génétique et génomique. Calculs originaux sous modèles explicites ; aucune estimation personnelle n'est proposée.

Référence complémentaire consultée : NHGRI, Mendelian Inheritance, définition institutionnelle de la transmission mendélienne.

Références du complément v0.4, consultées le 2 octobre 2026 : NHGRI, transmission mendélienne. Exemples et calculs originaux.



Partie 3 • Méiose, recombinaison et anomalies chromosomiques

Prérequis et objectifs

Connaître ploidie, chromatides et allèles. Le cours suit la séparation des homologues puis des chromatides, explique la liaison génétique et distingue anomalies de nombre et de structure. Les comptages doivent toujours préciser le stade.

1. Une réplication, deux divisions

Avant la méiose, la réplication transforme une cellule diploïde $2n, 2C$ en cellule $2n, 4C$. La première division sépare les chromosomes homologues et donne des cellules $n, 2C$. La seconde sépare les chromatides sœurs et donne des cellules $n, 1C$.

Il n'y a pas de nouvelle phase de réplication entre les deux divisions dans le schéma classique. Si l'ADN doublait de nouveau, la réduction finale attendue ne serait pas obtenue.

La méiose produit de la diversité par assortiment des homologues et recombinaison. La fécondation ajoute une autre combinaison de génomes parentaux.

2. Appariement et crossing-over

Pendant la prophase I, les homologues s'apparient et peuvent échanger des segments entre chromatides non sœurs. Ces événements de recombinaison contribuent à créer de nouvelles associations d'allèles et à la bonne ségrégation.

Un crossing-over n'échange pas obligatoirement la totalité d'un chromosome. Il ne se produit pas entre les deux chromatides sœurs dans la description classique qui explique le brassage interhomologue.

Les chiasmas représentent des conséquences visibles de ces échanges et des associations des homologues. Leur distribution n'est pas uniforme le long du génome ; des régions recombinent davantage que d'autres.

3. Assortiment indépendant et liaison

Deux loci sur des chromosomes différents se comportent souvent comme indépendants dans le modèle mendélien. Deux loci proches sur le même chromosome ont tendance à être transmis ensemble : ils sont liés.

Pour un individu AB/ab , les haplotypes parentaux sont AB et ab . Une recombinaison entre les loci peut produire Ab et aB . La fréquence des recombinants r varie de 0 à 0,5 dans le modèle usuel à deux loci.

Une fréquence de 50 % ne prouve pas que les loci soient sur des chromosomes différents : des loci très éloignés sur le même chromosome peuvent sembler indépendants en raison de recombinaisons multiples.

4. Carte génétique

Pour de petites distances, 1 % de recombinaison correspond approximativement à 1 centimorgan. Ce rapport n'est pas une conversion fixe en nombre de paires de bases ; la fréquence locale des échanges varie.

Un test-cross avec un double récessif peut révéler les gamètes du parent étudié. Si l'on observe 420 AB, 430 ab, 70 Ab et 80 aB, le total est 1000 et les recombinants 150, donc $r = 15\%$. La distance simple estimée est d'environ 15 cM, avec sous-estimation possible si des échanges multiples ne sont pas détectés.

La phase doit être identifiée : AB/ab et Ab/aB n'ont pas les mêmes classes parentales. Choisir les plus petites classes comme recombinants est une stratégie compatible dans ce cas simple, mais l'interprétation exige le modèle du croisement.

5. Non-disjonction et anomalies de nombre

Une non-disjonction des homologues en méiose I ou des chromatides en méiose II peut produire des gamètes avec un chromosome en plus ou en moins. Après fécondation, une trisomie ou une monosomie peut apparaître.

Dans une non-disjonction d'un bivalent en méiose I, tous les produits de cette méiose sont déséquilibrés pour ce chromosome dans le schéma simple. En méiose II dans une seule des deux cellules, deux produits peuvent rester normaux, un être $n + 1$ et un $n - 1$.

Une polyploïdie concerne des jeux chromosomiques entiers ; une aneuploïdie concerne un nombre anormal de chromosomes particuliers. Les termes ne sont pas interchangeables.

6. Anomalies de structure

Une délétion retire un segment ; une duplication en ajoute une copie ; une inversion retourne un segment ; une translocation déplace ou échange des régions entre chromosomes.

Un réarrangement équilibré ne modifie pas nécessairement la quantité totale de séquence détectable à une résolution donnée. Il peut néanmoins interrompre un gène, modifier sa régulation ou poser un risque de gamètes déséquilibrés. « Équilibré » ne signifie donc pas automatiquement « sans conséquence ».

Un caryotype observe de grandes structures ; une analyse de nombre de copies détecte d'autres déséquilibres à une résolution dépendante de la méthode. Une analyse de copies peut manquer un réarrangement équilibré. La méthode doit correspondre à la question.

7. Mosaïcisme et prélèvement

Une anomalie survenue après la fécondation peut ne concerner qu'une partie des cellules. La proportion varie selon les tissus et l'histoire du développement. Un résultat sur le sang ne décrit pas automatiquement chaque organe.

La résolution et le nombre de cellules analysées limitent la détection. L'absence d'anomalie dans un prélèvement ne démontre pas son absence absolue dans toute cellule de l'organisme.



Suivre une méiose

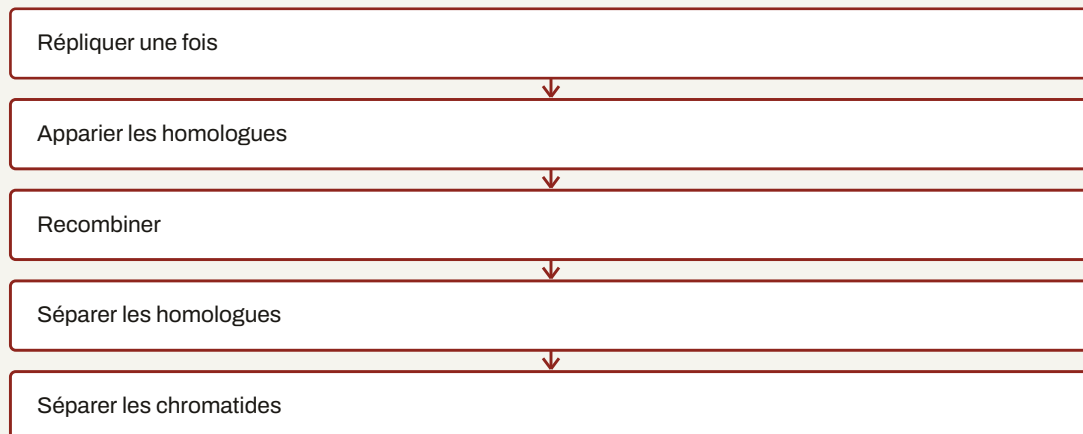


Schéma de principe original • lire les conditions et limites dans le texte

Ce schéma est une vue d'ensemble ; les sous-étapes de prophase et les mécanismes moléculaires de recombinaison demandent des développements spécifiques.

Développement : phase, doubles échanges et gamètes déséquilibrés

Quatre chromatides et un échange

Après réplication, une paire d'homologues comprend quatre chromatides. Un crossing-over simple entre deux chromatides non sœurs peut rendre deux chromatides recombinants et laisser deux chromatides non recombinants pour la région étudiée.

Un échange observé au niveau d'une paire ne signifie donc pas que tous les gamètes deviennent recombinants. La fréquence de recombinaison mesurée dans les descendants est liée aux produits transmis, pas seulement au nombre d'événements dans chaque méiose.

Des échanges multiples peuvent impliquer différentes chromatides et rendre la relation plus complexe. C'est une raison pour laquelle une carte génétique n'est pas un simple comptage visuel de cassures.

Doubles crossing-over

Deux échanges entre deux loci peuvent rétablir une association parentale aux marqueurs externes, tout en ayant recombinaison dans une région intermédiaire. Un test à deux marqueurs peut donc manquer certains événements.

Pour de grandes distances, la fréquence observée sous-estime le nombre total d'échanges. Elle tend vers 50 % et ne dépasse pas cette limite dans le modèle usuel, alors que la distance de carte peut dépasser 50 cM.

Ajouter un troisième marqueur aide à détecter certains doubles échanges et à établir l'ordre des loci. Les classes les plus rares sont souvent informatives dans un croisement à trois points, mais l'interprétation dépend de la viabilité et des autres hypothèses du problème.

Phase de liaison

AB/ab est une phase de couplage ; Ab/aB une phase de répulsion pour les mêmes allèles. Avec $r = 0,20$, la première produit 40 % AB, 40 % ab, 10 % Ab et 10 % aB dans le modèle simple.

Dans la seconde, les deux classes à 40 % deviennent Ab et aB. Oublier la phase inverse donc la prédiction de gamètes majoritaires sans changer la fréquence de recombinaison.

La phase peut être déduite d'un arbre ou de données de descendance. Elle doit être notée explicitement avant de calculer les probabilités de combinaisons.

Non-disjonction selon le stade

Une non-disjonction en première division maintient ensemble les deux homologues au lieu de les répartir. Après la seconde division, les quatre produits sont anormaux pour le chromosome concerné : deux $n + 1$ et deux $n - 1$ dans le schéma simple.

Une non-disjonction en seconde division dans une des deux cellules produit deux gamètes normaux, un $n + 1$ et un $n - 1$. Les proportions diffèrent donc selon le stade.

Après fécondation par un gamète normal, ces gamètes peuvent conduire à des états trisomiques ou monosomiques. La viabilité et le développement ultérieurs ne sont pas identiques pour toutes les anomalies ; les proportions de naissances observées ne reflètent pas directement celles des gamètes produits.

Translocation équilibrée

Chez un porteur d'une translocation réciproque, l'appariement méiotique doit tenir compte de segments répartis sur plusieurs chromosomes. Différents modes de ségrégation peuvent donner des gamètes équilibrés ou déséquilibrés.

Un individu peut donc être peu ou pas affecté tout en ayant une production de gamètes à risque de déséquilibre. Le risque exact dépend des chromosomes, points de cassure et autres facteurs ; il ne se déduit pas d'un pourcentage universel.

Un réarrangement apparemment équilibré peut aussi interrompre un gène ou modifier son environnement régulateur. L'absence de perte visible au caryotype n'exclut pas toutes les conséquences moléculaires.

Inversion et recombinaison

Une inversion change l'orientation d'un segment. Pendant l'appariement, une structure en boucle peut être nécessaire pour aligner les régions homologues.

Des crossing-over dans la région inversée peuvent produire des chromatides déséquilibrés selon le type d'inversion et les événements. La recombinaison observée peut donc être réduite ou sélectionnée dans les descendants.

Le simple fait de posséder les mêmes gènes ne suffit pas à garantir un comportement méiotique identique : leur ordre et leur orientation comptent.

Résolution des méthodes

Le caryotype montre de grandes structures ; une puce de nombre de copies recherche des pertes ou gains ; des méthodes de séquençage et d'analyse de jonctions peuvent préciser des réarrangements.



Un résultat normal doit être formulé avec la résolution et le type d'anomalie recherchés. « Aucun déséquilibre détecté » ne signifie pas « aucune modification génétique possible ».

Cas d'intégration : distinguer non-disjonction en première et seconde division

On suit une seule paire de chromosomes homologues dans une méiose modèle. Comparer les gamètes obtenus si les deux homologues ne se séparent pas en première division, ou si les chromatides d'un seul chromosome ne se séparent pas en seconde division après une première division normale.

Résolution et discussion

Lors d'une non-disjonction en première division, les deux homologues gagnent le même pôle. Après la seconde division, le schéma simple produit deux gamètes contenant deux exemplaires du chromosome suivi et deux gamètes n'en contenant aucun. Aucun des quatre gamètes n'a le nombre normal pour cette paire dans ce modèle.

Lors d'une non-disjonction dans une seule des deux cellules de seconde division, deux gamètes issus de l'autre cellule restent normaux pour le chromosome suivi. Les deux gamètes anormaux contiennent respectivement deux exemplaires et aucun. Le bilan est donc un gamète avec une copie supplémentaire, un avec une copie manquante et deux normaux.

Après fécondation par un gamète normal, les constitutions possibles incluent une trisomie ou une monosomie pour le chromosome considéré. Le vocabulaire $n + 1$ pour un gamète doit être distingué de $2n + 1$ pour le zygote. Les chromatides et les chromosomes ne sont pas comptés selon le même critère : le nombre de centromères et l'état de séparation comptent.

Les recombinaisons peuvent compliquer l'origine segmentaire des chromatides, et la viabilité des produits n'est pas équivalente pour tous les chromosomes. Les proportions du schéma sont celles d'un événement méiotique idéal, pas les proportions nécessairement observées parmi des naissances. La sélection et les étapes du développement peuvent modifier fortement ce qui est finalement mesuré.

Application : séparer fréquence observée et paramètre de recombinaison

Un croisement test donne 150 recombinants parmi 1 000 descendants. L'estimation est r observé = 0,15. Dans un modèle de descendants indépendants, sans sélection ni erreur de classement, l'écart-type d'échantillonnage estimé vaut $\sqrt{r(1-r)/N} = \sqrt{0,15 \times 0,85/1000} \approx 0,0113$.

Une plage approximative à 95 % par la méthode normale est donc $0,15 \pm 1,96 \times 0,0113$, soit environ 12,8 % à 17,2 %. Le résultat ne fixe pas r exactement à 15 % : une répétition peut donner un autre compte par hasard.

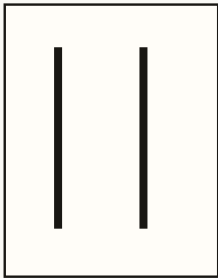
Cet intervalle ne corrige ni une survie différente des classes, ni un mauvais repérage des haplotypes parentaux, ni les recombinaisons multiples masquées. La distance en centimorgans reste également une autre grandeur ; l'approximation distance $\approx 100r$ est surtout adaptée aux petites distances.

Lecture visuelle : Méiose : ADN et ploïdie ne mesurent pas la même chose

Méiose : ADN et ploïdie ne mesurent pas la même chose

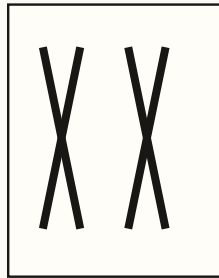
Avant S

$2n ; 2C$



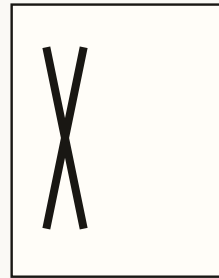
Après S

$2n ; 4C$



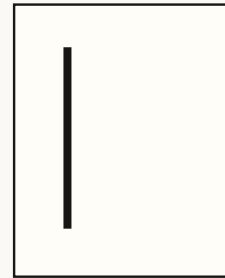
Après méiose I

$n ; 2C$



Après méiose II

$n ; 1C$



Un seul couple homologue dessiné ; C = ADN d'un génome haploïde.

Chaque vignette représente une cellule ; sans crossing-over représenté.

Chaque vignette représente une cellule ; un seul couple homologue est dessiné. Avant S : $2n/2C$; après S : $2n/4C$; après méiose I : $n/2C$; après méiose II : $n/1C$. Les recombinaisons sont omises et aucune réplication n'est placée entre les divisions.

Question de lecture. Une cellule après méiose I est-elle diploïde parce que chaque chromosome possède deux chromatides ?

Correction. Non : elle possède un seul jeu d'homologues, donc n. Le nombre de chromatides et la ploïdie sont distincts.

Approfondissement v0.4 : résoudre un croisement à trois marqueurs

Organiser les huit classes

Un croisement test donne, sur 1000 descendants, $ABC = 360$, $abc = 360$, $Abc = 40$, $aBC = 40$, $ABc = 90$, $abC = 90$, $AbC = 10$ et $aBc = 10$. On suppose classement fiable, viabilité égale et modèle où les classes parentales sont majoritaires et les doubles recombinants minoritaires.

Les parentaux sont ABC et abc. Les classes les plus rares sont AbC et aBc. En comparant ABC à AbC, seul B change ; de même abc et aBc diffèrent en B. Le marqueur central est donc B et l'ordre retenu A-B-C. Cette conclusion dépend du repérage correct des doubles recombinants ; elle n'est pas une règle aveugle sur n'importe quel jeu de données biaisé.



Compter dans chaque intervalle

Pour l'intervalle A-B, les simples recombinants Abc et aBC représentent 80 descendants. Les doubles recombinants AbC et aBc ont aussi recombinaison cet intervalle : on ajoute 20. La fréquence est donc $100/1000 = 0,10$.

Pour B-C, les classes ABc et abC apportent 180 descendants ; les mêmes 20 doubles sont ajoutés. La fréquence est $200/1000 = 0,20$. La carte approchée additionne ces intervalles : $10 + 20 = 30$ cM, sous les limites de l'approximation et de possibles événements supplémentaires non détectés.

Pourquoi les marqueurs externes donnent moins

Les doubles échanges rétablissent l'association parentale A-C. Si l'on observe seulement les deux marqueurs externes, les recombinants sont les $80 + 180 = 260$ simples, soit 26 %. La somme $10 \% + 20 \%$ comptait les doubles deux fois, tandis que l'observation externe ne les compte pas. On retrouve $r_{AC} = r_{AB} + r_{BC} - 2 \times \text{fréquence des doubles} = 0,10 + 0,20 - 0,04 = 0,26$ dans ce modèle.

Comparer doubles observés et attendus

Si les événements des deux intervalles étaient indépendants au niveau du modèle utilisé, la fréquence attendue des doubles serait $0,10 \times 0,20 = 0,02$, soit 20 descendants. Ici, elle coïncide avec l'observation. Le coefficient de coïncidence vaut observés/attendus = 1 et l'interférence définie par 1 moins ce coefficient vaut zéro.

Une observation de dix doubles pour vingt attendus donnerait un coefficient de 0,5 et une interférence de 0,5, mais il faudrait recalculer les fréquences d'intervalles à partir du nouveau tableau complet. Remplacer un seul effectif sans ajuster le total ou les autres classes produirait un jeu incohérent. Ces outils décrivent la répartition des échanges ; ils ne convertissent pas automatiquement des centimorgans en paires de bases.

Exercices progressifs

1. Y a-t-il réplication entre méiose I et II ?
2. Quel couple se sépare en méiose I ?
3. Sur 800 descendants, 120 sont recombinants. Calculer r.
4. Une fréquence de 50 % prouve-t-elle des chromosomes différents ?
5. Une translocation équilibrée est-elle nécessairement sans effet biologique ?
6. Si N est multiplié par quatre à fréquence comparable, comment varie cet écart-type d'échantillonnage ?
7. Dans le tableau fourni, combien de descendants seraient classés recombinants si l'on ne lisait que A et C ? Pourquoi ne faut-il pas répondre 300 ?

Corrections argumentées

1. Non dans le déroulement classique.
2. Les chromosomes homologues ; les chromatides sœurs restent associées.
3. $120/800 = 0,15$, soit 15 %.



4. Non. Des loci éloignés sur le même chromosome peuvent donner cette valeur apparente.
5. Non. Elle peut interrompre un gène, changer une régulation ou produire des gamètes déséquilibrés.
6. Il est divisé par deux, car il varie comme $1/\text{racine}(N)$. Augmenter l'effectif réduit la fluctuation aléatoire mais ne supprime pas un biais de sélection ou une erreur de classement.
7. Il y en a 260, les quatre classes simples totalisant $40 + 40 + 90 + 90$. Les vingt doubles ont des associations A-C parentales. Le total 300 additionne les comptages des deux intervalles et inclut deux fois les vingt doubles : $300 - 2 \times 20 = 260$.

Référence et périmètre

NHGRI, glossaire de génétique, méiose et crossing-over, NCBI Bookshelf, recombinaison. Cartes simplifiées et exemples originaux.

Référence complémentaire consultée : NHGRI, Meiosis, définition institutionnelle de la méiose.

Références du complément v0.4, consultées le 2 octobre 2026 : NHGRI, crossing-over. Exemples et calculs originaux.

Partie 4 · PCR, séquençage et analyse expérimentale

Prérequis et objectifs

Connaître complémentarité, réplication et transcription. Le cours apprend à choisir une méthode selon la question, prévoir ses témoins et interpréter ses limites. Détecter une séquence, mesurer une quantité et démontrer une fonction sont trois résultats différents.

1. Extraction et qualité

Une extraction libère l'acide nucléique et élimine une partie des contaminants. Concentration, pureté et intégrité sont trois dimensions distinctes. Un extrait très concentré peut contenir des inhibiteurs de polymérase ou des fragments trop courts pour l'objectif.

L'ARN est sensible aux ribonucléases et aux conditions de conservation. Un ratio d'absorbance fournit un indice de pureté, mais ne prouve pas à lui seul l'intégrité ni l'absence d'inhibiteurs.

Un témoin d'extraction négatif suit les mêmes manipulations sans échantillon biologique pertinent. Il permet de repérer certaines contaminations introduites avant l'amplification, que le seul témoin sans matrice de PCR ne couvre pas entièrement.

2. PCR

La PCR utilise deux amorces qui encadrent la région cible et sont orientées pour que leurs extrémités 3' conduisent la synthèse vers cette région. Les cycles alternent séparation des brins, hybridation des amorces et élongation.

Dans un modèle idéal, la quantité double à chaque cycle : $N = N_0 \times 2^n$. En pratique, l'efficacité n'est pas toujours parfaite et la réaction atteint un plateau. Une bande finale ne permet donc pas toujours de comparer directement les quantités initiales.

La spécificité dépend des amorces et des conditions. Un produit de la taille attendue peut être non spécifique ; séquençage ou autre contrôle peut être nécessaire. Un résultat négatif peut venir d'une absence de cible, mais aussi d'un inhibiteur, d'une extraction défectueuse ou d'une variation dans la zone d'amorce.

3. RT-PCR et qPCR

La transcription inverse convertit un ARN en ADN complémentaire. Une RT-PCR comprend cette étape avant amplification. Le sigle RT ne signifie pas automatiquement « temps réel » ; qPCR désigne une amplification suivie quantitativement par un signal.

Un témoin sans transcriptase inverse aide à détecter une contamination par ADN génomique dans une analyse d'ARN. Des amorces traversant une jonction d'exons peuvent améliorer la spécificité, sans remplacer tous les contrôles.

En qPCR, le cycle de quantification C_q est atteint plus tôt si la cible initiale est plus abondante, à efficacité et seuil comparables. Une différence de trois cycles correspond à un facteur huit si le doublement est idéal.

La méthode relative $2^{-\Delta\Delta Cq}$ exige notamment des efficacités appropriées et une référence stable. Une référence dont l'expression change avec le traitement peut fausser ou inverser l'interprétation.

4. Séquençage

Le séquençage Sanger utilise des terminaisons de chaîne et permet de lire une séquence sur une région adaptée. Un chromatogramme mélangé peut refléter plusieurs allèles, une contamination ou une amplification non spécifique selon le motif observé.

Les technologies de séquençage massif lisent de nombreux fragments. Les lectures sont alignées ou assemblées, puis analysées pour détecter des variations. Les régions répétées, certains biais de couverture et la longueur des lectures limitent les conclusions.

La profondeur est le nombre de lectures couvrant une position ; la couverture peut désigner la fraction de la cible suffisamment lue. Une forte profondeur moyenne ne garantit pas que chaque base soit bien couverte.

La fraction allélique dépend du génotype, de la composition cellulaire, du nombre de copies et des biais techniques. Dix lectures variantes sur vingt ne démontrent pas à elles seules un hétérozygote constitutionnel sans contexte.

5. Choisir selon la question

Un caryotype détecte de grandes anomalies chromosomiques. Une analyse de nombre de copies peut détecter des pertes ou gains plus fins, avec sa résolution propre. Une analyse de séquence peut détecter substitutions et petites insertions/délétions dans les régions correctement interrogées.

Aucune technique ne détecte automatiquement tous les variants. Une méthode ciblant quelques exons ne renseigne pas sur tout le génome, les régions régulatrices ou tous les réarrangements.

Question	Méthode possible	Contrôle essentiel
Cible ADN présente ?	PCR	Négatif et positif
ARN relatif modifié ?	RT-qPCR	Sans RT et référence stable
Séquence d'un amplicon ?	Sanger	Produit spécifique
Variants nombreux ?	Séquençage massif	Couverture et validation
Fonction d'un élément ?	Test fonctionnel	Perturbation et restauration

6. Édition et causalité

Une édition ciblée peut modifier un locus, mais une population traitée peut rester hétérogène. Détecter une modification globale ne signifie pas que chaque cellule possède le même génotype ou que les deux allèles sont modifiés.

Il faut vérifier le site visé et les conséquences pertinentes : ARN, protéine ou fonction selon la question. Des effets inattendus, sélection de clones ou dommages associés peuvent contribuer au

résultat.

Une expérience de restauration avec une version fonctionnelle peut renforcer un lien causal. Elle doit rester comparable : expression excessive, localisation différente ou présence d'une étiquette peuvent créer une situation artificielle.

Les répétitions biologiques indépendantes sont essentielles. Dix PCR du même ADN évaluent une répétabilité technique, pas la variabilité entre dix organismes ou cultures indépendants.

Construire une preuve moléculaire

1. Définir la question • 2. Choisir la cible et la méthode • 3. Prévoir les témoins • 4. Vérifier le signal • 5. Relier au mécanisme sans dépasser la preuve

Ce schéma insiste sur les limites : amplification ne signifie pas fonction, et modification ne signifie pas correction complète.

Développement : dessiner des amorces et interpréter une quantification

Orientation des amorces

Sur un brin de référence écrit 5' vers 3', l'amorce avant possède une séquence correspondant à la région de départ et s'hybride au brin complémentaire. L'amorce arrière correspond au complément inverse d'une région située à l'autre extrémité de la cible.

Les extrémités 3' des deux amorces pointent vers la région à amplifier. Si elles pointent vers l'extérieur, les extensions ne produisent pas le fragment borné attendu dans une PCR classique.

Une séquence complémentaire écrite dans le mauvais sens peut sembler correcte lettre par lettre mais être inutilisable. L'orientation doit donc être inscrite sur chaque brin avant de choisir les amorces.

Amplification non idéale

Si l'efficacité par cycle est E, la quantité suit approximativement $N = N_0(1 + E)^n$ pendant la phase exponentielle. E = 1 correspond au doublement ; E = 0,90 à un facteur 1,90.

Une différence de trois cycles correspond alors à $1,90^3 \approx 6,86$ plutôt qu'à huit si l'on utilise ce modèle. Les efficacités doivent être comparables pour interpréter directement les écarts de Cq.

Au plateau, substrats limitants, accumulation de produits et pertes d'activité modifient la progression. La quantité finale après de nombreux cycles n'est donc pas proportionnelle de manière simple à la quantité initiale.

Exemple $\Delta\Delta Cq$

Dans un témoin, Cq cible = 25 et Cq référence = 20, donc $\Delta Cq = 5$. Dans un échantillon traité, Cq cible = 23 et Cq référence = 20, donc $\Delta Cq = 3$. $\Delta\Delta Cq = 3 - 5 = -2$.

Avec efficacité idéale et référence stable, le rapport vaut $2^2 = 4$. Le traitement est compatible avec une quantité relative de cible quatre fois plus élevée.

Si la référence était elle-même modifiée, ce rapport serait biaisé. Plusieurs références validées ou une stratégie de normalisation adaptée peuvent améliorer l'interprétation. Le signe de $\Delta\Delta C_q$ doit être vérifié avant d'annoncer hausse ou baisse.

Témoins distincts

Un témoin sans matrice vérifie certaines contaminations de la préparation PCR. Un témoin d'extraction suit les étapes en amont. Un témoin sans RT vérifie la contribution d'ADN dans une étude d'ARN. Un témoin positif montre que le système peut amplifier une cible appropriée.

Un contrôle interne peut révéler une inhibition dans l'échantillon. Si cible et contrôle échouent, on ne peut pas conclure simplement à l'absence de cible.

Ces témoins ne sont pas redondants : chacun teste une hypothèse de défaillance différente. Leur choix dépend du chemin complet allant du prélèvement au résultat.

Profondeur et incertitude

Si une position est couverte par vingt lectures et que deux portent un variant, la fraction observée est 10 %, mais son incertitude est importante. Une petite variation de nombre de lectures change fortement le pourcentage.

Des erreurs de séquençage peuvent produire quelques lectures alternatives. La qualité des bases, l'orientation des lectures, les duplicats et le contexte répétitif doivent être examinés avant d'interpréter un faible signal comme mosaïcisme.

Une profondeur élevée ne corrige pas automatiquement un biais systématique d'amplification. Multiplier les lectures d'un même artefact peut renforcer une impression trompeuse de précision.

Sanger et mélange d'allèles

Une substitution hétérozygote peut donner deux pics à une position dans un chromatogramme propre. Une insertion ou délétion hétérozygote peut décaler les superpositions en aval et produire un motif beaucoup plus complexe.

La lecture doit tenir compte de la qualité et d'éventuels produits non spécifiques. Une amplification d'un seul allèle peut masquer l'autre, notamment si une variation gêne une amorce.

Édition et validation

Après édition, séquencer le site ciblé permet de décrire les allèles présents. Un mélange de petites insertions/délétions n'est pas équivalent à une correction précise attendue.

La fonction doit être testée au niveau pertinent : ARN, protéine, localisation ou phénotype cellulaire. Une modification du génome sans changement fonctionnel peut être insuffisante pour l'objectif.

Des clones indépendants et une restauration adaptée aident à distinguer effet du locus et particularité d'un clone. Une population unique sélectionnée peut accumuler des différences étrangères à la cible.

Cas d'intégration : interpréter une PCR quantitative relative

Un gène cible donne un Cq de 24 dans un témoin et de 21 dans un échantillon. Un gène de référence donne respectivement 20 et 19. Calculer la variation relative selon la méthode $2^{-\Delta\Delta Cq}$ en supposant des efficacités compatibles et une référence stable. Pourquoi le seul écart de trois cycles de la cible serait-il insuffisant ?

Résolution et discussion

Pour le témoin, $\Delta Cq = 24 - 20 = 4$. Pour l'échantillon, $\Delta Cq = 21 - 19 = 2$. On obtient $\Delta\Delta Cq = 2 - 4 = -2$, puis $2^{-\Delta\Delta Cq} = 4$. La quantité relative normalisée est donc estimée quatre fois plus élevée. Utiliser directement les trois cycles de différence de la cible aurait donné huit, en oubliant la différence portée aussi par la référence.

Cette normalisation vise à limiter certains effets de quantité initiale et de préparation. Elle ne corrige pas automatiquement une référence dont l'expression varie biologiquement entre conditions. La stabilité des gènes de référence doit être évaluée pour l'expérience considérée ; plusieurs références appropriées peuvent renforcer l'analyse.

La base deux suppose une amplification proche du doublement par cycle et des efficacités suffisamment comparables. Des inhibiteurs, des amorces inefficaces ou des produits non spécifiques peuvent invalider cette simplification. Une courbe d'étalonnage et des contrôles de spécificité renseignent sur ces points.

Un témoin sans matrice recherche une contamination ou certains produits parasites. Un témoin sans transcription inverse aide à repérer une contribution d'ADN lors de l'étude d'ARN. Aucun de ces témoins ne remplace l'autre, car leurs omissions diffèrent. Enfin, un changement d'ARN ne démontre pas nécessairement une même variation de protéine active : traduction, dégradation et modifications post-traductionnelles constituent d'autres niveaux de régulation.

Application : quantifier le risque de manquer un allèle

Dans un modèle idéal, un allèle présent à la fraction $f = 0,10$ est échantillonné par N lectures indépendantes sans biais. La probabilité de ne voir aucune lecture variée est $(1 - f)^N$. Avec dix lectures, elle vaut $0,9^{10} \approx 34,9\%$. Zéro lecture variée ne serait donc pas une preuve forte d'absence.

Avec trente lectures indépendantes, cette probabilité tombe à $0,9^{30} \approx 4,24\%$. On améliore la sensibilité du simple échantillonnage, mais on n'a pas pour autant validé un appel de variant : erreurs de lecture, seuils de qualité et duplicats restent à examiner.

Les répétitions issues d'une même molécule amplifiée ne sont pas toujours des observations indépendantes. Un biais d'amplification contre l'allèle peut persister à forte profondeur apparente. Le calcul mesure seulement la limite binomiale du modèle, et non la performance clinique complète d'une méthode.

Approfondissement v0.4 : étalonnage de qPCR et contrôle d'inhibition

Déduire l'efficacité de la pente

Pendant la phase exponentielle idéale, le seuil fixe N_{seuil} est atteint lorsque $N_{\text{seuil}} = N_0(1+E)^{C_q}$. En prenant le logarithme décimal, on obtient $C_q = \text{constante} - \log_{10}(N_0)/\log_{10}(1+E)$. La pente m de C_q en fonction de $\log_{10}(N_0)$ vaut donc $-1/\log_{10}(1+E)$, d'où $E = 10^{(-1/m)} - 1$.

Une pente d'environ $-3,322$ correspond à $E = 1$: une dilution au dixième retarde le seuil d'environ 3,322 cycles. Une pente de $-3,60$ donne un facteur par cycle d'environ 1,896 et une efficacité d'environ 89,6 %. La pente est négative parce que davantage de matrice initiale fait diminuer C_q .

Utiliser une dilution sans oublier son coût en cible

Diluer un extrait diminue à la fois la cible et les inhibiteurs éventuels. Sans inhibition et avec doublement idéal, une dilution au dixième doit augmenter C_q de 3,322. Si l'extrait non dilué atteint le seuil à 29 et le dilué au dixième à 30, le retard n'est que d'un cycle. Ce résultat est compatible avec une levée partielle d'inhibition qui compense une partie de la perte de cible.

Il ne démontre pas seul quel composé inhibe la réaction. Des erreurs de pipetage, des produits non spécifiques ou une variation d'efficacité peuvent aussi perturber l'écart. Un contrôle interne ajouté de manière appropriée et une gamme de dilutions aident à distinguer les hypothèses.

Comparer des cibles d'efficacités différentes

Si les facteurs par cycle sont connus et constants, un rapport relatif normalisé peut s'écrire $F = \text{acible}^{(C_q \text{ témoin} - C_q \text{ échantillon})} / \text{aréférence}^{(C_q \text{ témoin} - C_q \text{ échantillon})}$. Chaque différence utilise les C_q de son propre essai. Lorsque les deux facteurs valent deux, on retrouve la forme $2^{-\Delta\Delta C_q}$.

Avec une avance de deux cycles pour la cible à facteur 1,9 et une avance d'un cycle pour la référence à facteur 2, le rapport est $1,9^2/2 = 1,805$. Utiliser automatiquement deux pour les deux essais donnerait deux. La correction reste tributaire d'efficacités réellement mesurées dans des conditions pertinentes, et d'une référence biologiquement stable.

Distinguer répliquats et étalons

Des répétitions techniques évaluent la dispersion d'une préparation ; une gamme d'étalonnage teste une relation entre quantité et signal ; des répétitions biologiques évaluent la variabilité entre échantillons indépendants. Ces trois éléments répondent à des questions différentes. Une belle droite d'étalonnage ne garantit pas une extraction comparable ni une absence de biais biologique dans la normalisation.

Exercices progressifs

1. Deux amorces de PCR doivent-elles s'allonger vers l'extérieur de la cible ?
2. Quel facteur idéal correspond à quatre cycles de différence ?
3. À quoi sert le témoin sans RT ?
4. Une profondeur moyenne de 100 lectures garantit-elle 100 lectures sur chaque base ?
5. Une bande de bonne taille prouve-t-elle la séquence exacte ?

6. Quelle probabilité de ne lire aucun allèle variant à fraction réelle $1/2$ avec dix lectures idéales indépendantes ?
7. Une gamme donne une pente $-3,322$. Prévoir l'écart de C_q après dilution au quart, puis calculer un rapport normalisé avec avance de trois cycles cible et d'un cycle référence, toutes deux à doublement idéal.

Corrections argumentées

1. Non. Leurs extensions doivent converger vers la région à amplifier.
2. $2^4 = 16$, sous efficacité idéale comparable.
3. À repérer une amplification possible d'ADN contaminant sans conversion préalable de l'ARN.
4. Non. La couverture est hétérogène ; certaines régions peuvent être insuffisamment lues.
5. Non. Un produit non spécifique de taille proche reste possible.
6. $(1/2)^{10} = 1/1024 \approx 0,0977\%$. Le chiffre suppose une amplification équilibrée et des lectures indépendantes ; il ne tient pas compte d'une perte d'allèle pendant l'amplification.
7. La pente correspond à un facteur proche de deux par cycle. Une dilution au quart demande $\log_2(4) = 2$ cycles supplémentaires. Pour le rapport normalisé, $2^3/2^1 = 4$. La dilution est une modification technique connue ; la normalisation compare ici deux essais et ne doit pas être confondue avec elle.

Références et périmètre

NHGRI, *Polymerase Chain Reaction*, NCBI Bookshelf, transcription. Principes et expériences pédagogiques, sans protocole clinique validé.

Références du complément v0.4, consultées le 2 octobre 2026 : NHGRI, PCR. Exemples et calculs originaux.

Références mécanistiques vérifiées le 2 octobre 2026 : Bustin et al., 2009, MIQE : notice et résumé des recommandations expérimentales.