



Mineure Epidémiologie & Recherche Clinique

Les apports de la bio-informatique en épidémiologie et
recherche clinique

Etude de cas : La crise COVID-19



Jérôme Vétillard

VP R&D Qualees

Industry Advisor Healthcare & Life Sciences Microsoft



En temps normal, le développement d'un médicament prend :

- jusqu'à 10 ans,
- coûte entre 1.5 et 2.5 milliards de dollars USD

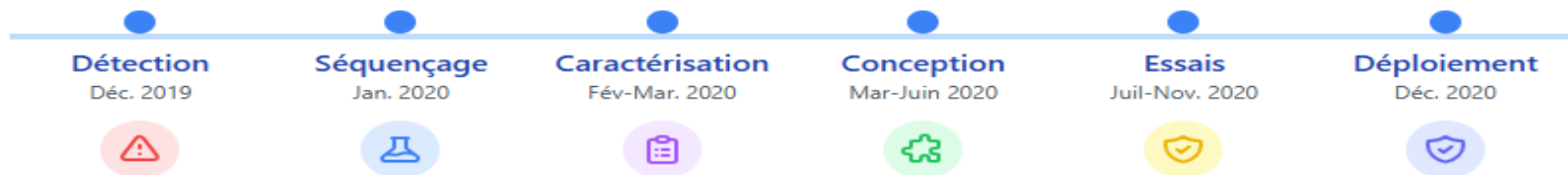
Cette fois-ci, il faudra que ça aille beaucoup plus **VITE**

(10 ans confinés c'est un peu trop long)...

365 jours d'innovation vaccinale

De la découverte du virus au déploiement vaccinal

Chronologie des grandes étapes



"Le développement d'un vaccin en moins d'un an : un exploit inédit qui a transformé notre compréhension de l'innovation biomédicale"

Jalons majeurs

31 décembre 2019

Premiers cas signalés à Wuhan

11 janvier 2020

Publication du premier séquençage complet

8 décembre 2020

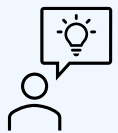
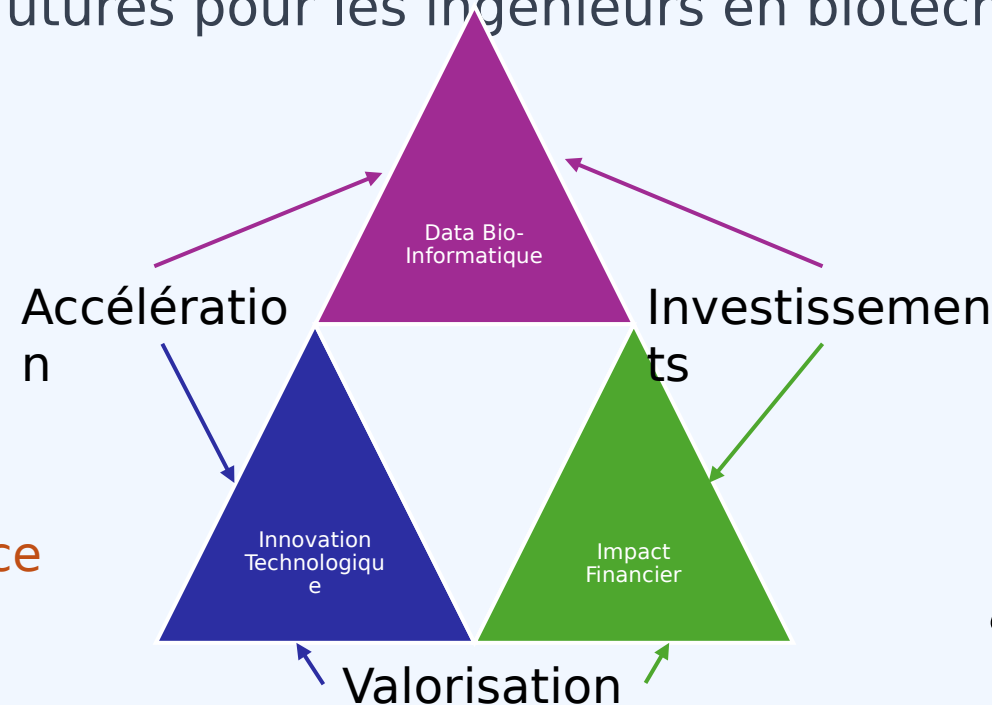
Premières vaccinations au Royaume-Uni

Objectifs et Enjeux

"Comment la bioinformatique et la data ont révolutionné notre capacité d'innovation ?«

Objectifs de la présentation

- Comprendre l'intégration des outils bio-informatiques dans la recherche clinique
- Analyser la corrélation entre innovation technologique et impact financier
- Identifier les technologies transférables à d'autres domaines de la biologie industrielle
- Explorer les applications futures pour les ingénieurs en biotechnologie



Point clé EBI : Importance pour les ingénieurs en biologie industrielle de maîtriser cette triple



Anecdote : Le business model de la médecine de précision

Méthodologie Pédagogique



Concept narratif

Format "365 jours - style 24h chrono" avec découpage en segments temporels

365
JOURS



Outils et supports

- ✓ **Timeline interactive**
Enrichie au fil du cours
- ✓ **Visualisations 3D**
Et analyses comparatives
- ✓ **Exercices pratiques courts**
5 minutes par segment



Technologies abordées

NGS Alignement de séquences Modélisation moléculaire Machine learning ERP Systèmes EDC
Analyse bayésienne Dynamique moléculaire Pipelines bioinformatiques

Segment 1

Jours 0-30



Segment 2

Jours 31-60



Segment 3

Jours 61-120



Segment 4

Jours 121-200



Segment 5

Jours 201-300



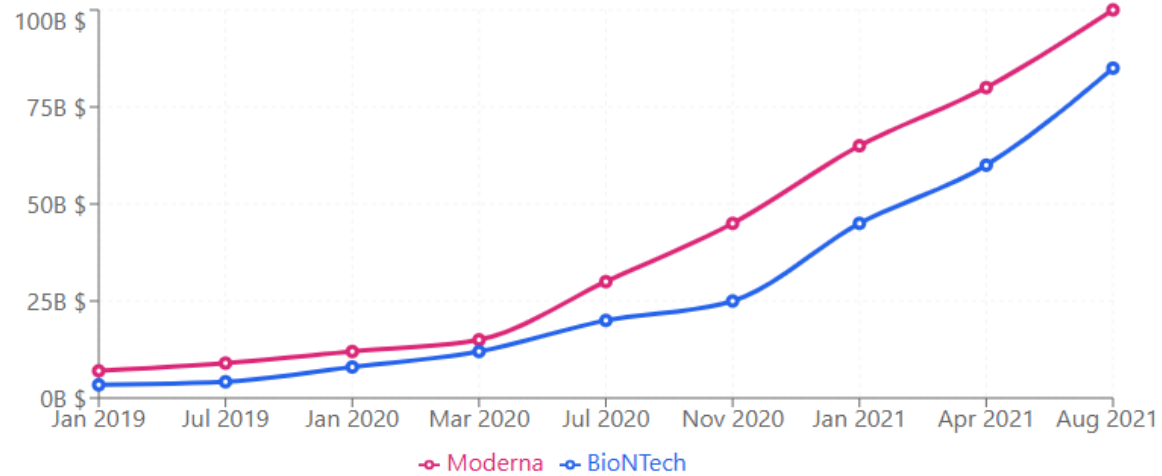
Segment 6

Jours 301-365



Introduction au Versant Financier

Évolution des capitalisations boursières (2019-2021)



Moderna

+1,330%

De 7 milliards \$ (2019) à plus de 100 milliards \$ (août 2021)



BioNTech

+2,400%

De 3,4 milliards \$ (2019) à plus de 85 milliards \$ (août 2021)



Enjeux connexes

⚠ Dimensions éthiques

- Accès équitable aux vaccins
- Partage des bénéfices
- Transparence des données

🛡 Aspects réglementaires

- Procédures d'urgence (EUA)
- Cadres réglementaires adaptés
- Surveillance post-commercialisation

📄 Propriété Intellectuelle

- Brevets sur technologies ARNm
- Licences et collaborations
- Débat sur la levée des brevets

J0 à J30 : THE OUTBREAK

Le 1^{er} décembre, le premier cas de Covid-19 aurait été détecté à Wuhan.

Il s'agirait d'un patient de 55 ans tombé malade le 17 novembre 2019 à Wuhan.

le 15 décembre, le nombre de cas s'élève à 27.

Le 11 janvier 2020, la Chine annonçait LE PREMIER Décès d'une « mystérieuse pneumonie » (mais excluant le « SRAS »)

Le 13 janvier, l'OMS annonce qu'un premier cas est identifié en Thaïlande, il s'agit d'une Chinoise de 61 ans, provenant de Wuhan

Trois cas du nouveau coronavirus chinois sont confirmés en France, sur des patients hospitalisés à Bordeaux et Paris, le 24 janvier. Il s'agit des premiers cas européens.

Le 30 janvier, le nombre de personnes infectées par le



Chiffres clés au 30 janvier 2020

Cas confirmés

7,818

Officiellement rapportés

Cas estimés réels

~75,800

Selon modèles épidémiologiques

Pays touchés

20

Sur 5 continents

Décès

170

Principalement en Chine

Discours du Directeur Général de l'Organisation Mondiale de la Santé en janvier 2020 : [LIEN](#)

Urgence de **S**anté **P**ublique de **P**ortée **I**nternationale déclarée 30 jours après les premiers cas :

« Un événement extraordinaire dont il est déterminé qu'il constitue un risque pour la santé publique dans d'autres états en raison du risque de propagation internationale de maladies et qu'il peut requérir une action internationale coordonnée »

Structure de la Veille Épidémiologique

Comment détecte-t-on une pandémie émergente ?



Niveaux de surveillance

Local

Hôpitaux, cliniques, médecins généralistes, laboratoires locaux

National

Centres nationaux de contrôle des maladies (CDC, Santé Publique France)

International

OMS, réseaux régionaux (ECDC), collaborations entre pays

Technologique

Surveillance numérique, IA, big data, réseaux sociaux



Indicateurs d'alerte

Signaux épidémiologiques

- Clusters inhabituels de maladies
- Taux d'attaque élevés
- Propagation rapide géographique
- Mortalité anormale

Seuils d'alerte

- Écart statistique vs moyenne historique
- Transmission interhumaine établie
- Propagation dans ≥ 3 pays
- $R_0 > 1$ (reproduction épidémique)

Critères OMS (USPPI)

- Gravité de l'impact sur la santé publique
- Caractère inhabituel ou inattendu
- Risque de propagation internationale
- Risque d'entraves aux voyages/commerce



Méthodes de détection

Surveillance syndromique

Suivi des symptômes et syndromes sans diagnostic précis

- Détection précoce des anomalies

Surveillance de laboratoire

Tests, séquençage génomique, identification de nouveaux agents

- Confirmation et caractérisation

Surveillance numérique

Analyse des recherches web, médias sociaux, mobilité

- Détection ultra-précoce

Surveillance environnementale

Échantillonnage des eaux usées, surveillance animale

- Détection des réservoirs et vecteurs

Veille Épidémiologique & Outils Informatiques

Révolution numérique dans la détection des pandémies

↗ Évolution des systèmes de surveillance

Pré-1990s : Surveillance traditionnelle

- Rapports manuscrits
- Notification par téléphone
- Délai de semaines à mois
- Analyse manuelle

Réactif et lent

1990-2010 : Digitalisation

- Premiers systèmes EDS électroniques
- Bases de données centralisées
- Partage par email et web
- Analyse statistique automatisée

Plus rapide, plus structuré

2010-2020 : Surveillance numérique

- Surveillance basée sur l'internet
- Intégration des données mobiles
- Analyse des médias sociaux
- Systèmes d'alerte précoce

Proactif, temps réel

Post-2020 : IA et Big Data

- Apprentissage profond
- Analyse prédictive
- Modélisation en temps réel
- Intégration de sources multiples

Prédictif, automatisé



Collecte de données

Réseaux de capteurs IoT

Surveillance environnementale, mesure de la mobilité, monitoring des paramètres vitaux

Web scraping

Extraction automatisée de données depuis actualités, sites gouvernementaux, forums

API et plateformes d'échange

GISAID, HealthMap, ProMED, systèmes EDC interconnectés

Applications mobiles

Déclaration de symptômes par les citoyens, suivi des contacts



Analyse et traitement

Traitement du langage naturel (NLP)

Analyse de textes non structurés, détection de mentions de maladies, analyse de sentiment

Machine Learning

Détection d'anomalies, classification de symptômes, prédiction de propagation

Bioinformatique

Analyse de séquences génomiques, phylogénie, identification de variants

Modélisation épidémiologique

Modèles SIR/SEIR, simulations stochastiques, matrices de contact



Visualisation et diffusion

Dashboards interactifs

Tableaux de bord temps réel (ex: Johns Hopkins COVID-19), filtres dynamiques

Cartographie SIG

Systèmes d'information géographique, heat maps, cartographie dynamique

Systèmes d'alerte

Notifications automatisées, alertes SMS/email, gradation des niveaux d'alerte

APIs publiques

Partage de données structurées, open data, intégration avec systèmes tiers

Veille Épidémiologique & Outils Informatiques

Révolution numérique dans la détection des pandémies

Outils informatiques qui ont fait la différence dans la détection du COVID-19

BlueDot (Canada)

Système d'intelligence artificielle qui a détecté le COVID-19 dès le 30 décembre 2019, avant l'annonce officielle de l'OMS.

Technologie : Analyse automatisée de 100 000+ articles en 65 langues quotidiennement, données de vols aériens, démographie

HealthMap (Harvard)

A identifié le cluster de pneumonie atypique à Wuhan le 30 décembre 2019 en détectant les mentions dans les médias locaux.

Technologie : Web scraping, traitement du langage naturel, visualisation géospatiale automatisée

Nextstrain

Plateforme open-source qui a permis le suivi en temps réel de l'évolution génomique du SARS-CoV-2.

Technologie : Analyse phylogénétique, visualisation interactive, intégration avec GISAID

Google Mobility Reports

Analyse anonymisée des déplacements pour évaluer l'efficacité des mesures de confinement et prédire la propagation.

Technologie : Big data de mobilité, agrégation anonymisée, visualisation comparative



Avantages des outils informatiques

- ✓ Détection plus précoce (jours/semaines avant les méthodes classiques)
- ✓ Intégration de sources de données hétérogènes et massives
- ✓ Réduction des biais humains dans l'interprétation
- ✓ Scalabilité face aux volumes croissants de données
- ✓ Modélisation prédictive pour anticiper les évolutions



Limites et défis actuels

- ⚠ Faux positifs et "infodémie" (bruit informationnel)
- ⚠ Fracture numérique (couverture inégale selon les régions)
- ⚠ Enjeux de vie privée et questions éthiques
- ⚠ Dépendance aux infrastructures techniques
- ⚠ Nécessité de validation humaine des alertes algorithmiques

Course contre la montre génomique

Séquençage en Urgence - Évolution technologique 2003-2023

Illumina NovaSeq 6000

DÉBIT MAX

6 Tb

par run

PRÉCISION

99.9%

Q30, paired-end

- Séquençage par synthèse (SBS)
- Reads courts (150-300 pb)
- Prix moyen run: 4 000-12 000€

DURÉE RUN

13-48h

selon flow cell

POIDS

147 kg

Haute précision



Oxford Nanopore MinION Mk1C

Portable

DÉBIT MAX

50-100 Gb

par flow cell

PRÉCISION

95-99%

TEMPS RÉEL

1min - 72h

analyse en direct

POIDS

450g

ultraportable



Score Phred :

- Le **score Phred** est une mesure logarithmique de la qualité d'une lecture de séquençage en biologie moléculaire, notamment dans les technologies de séquençage haut débit (**Next Generation Sequencing, NGS**).
- Formellement, le score Phred Q est défini par :

$$Q = -10 \times \log_{10}(P)$$

où P est la probabilité que la base soit appelée de manière incorrecte (c'est-à-dire une erreur de lecture).

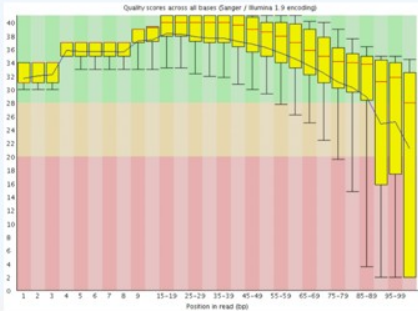
- Par exemple :
 - **Phred 30** ($Q = 30$) signifie que la probabilité d'erreur est 1/1000 (soit 0,1%).
 - **Phred 20** équivaut à 1/100 (1% d'erreur).

Comparaison des techniques

Critère			
Temps de préparation			
Temps d'exécution			
Précision			
Coût par Gb			
Débit par run	3-6 Tb (NovaSeq)	50-100 Gb (MinION)	90-160 Gb (Revio)
Longueur des reads	150-300 bp	10-100 kb (moyenne 20 kb)	15-25 kb (HiFi)
Investissement initial	800K-1M€ (NovaSeq)	900-1500€ (MinION)	750-850K€ (Revio)

Contrôle qualité (FastQC) - Métriques critiques

Scores de qualité par base (Phred) Illumina NovaSeq 6000



Score Phred ≥ 30 : 93.2% des bases
Erreur estimée: $< 1/1,000$ bases
✓ PASS - Haute qualité de lecture

Course contre la montre génomique

Séquençage en Urgence - Évolution technologique 2003-2023

Paramètres d'analyse FastQC - Séquençage viral

Paramètres Généraux

Génome de référence: NC_045512.2
Taille: 29,903 bp
GC%: 38%

Seuils de Qualité

Phred minimum: Q20
Couverture min: 100x
Contamination max: 5%

Résultats Analyse

Variants identifiés: 7
Contamination humaine: 2.3%
ORFs complets: 11/11

Rubrique	Champ	Signification	Pourquoi c'est important
Paramètres généraux	Génome de référence : NC_045512.2	Accession GenBank de la souche « Wuhan-Hu-1 » du SARS-CoV-2, utilisée ici comme séquence de référence pour l'alignement et l'appel de variants. PMC	C'est le cadre de comparaison : toutes les mutations (variants) sont repérées comme des écarts par rapport à cette séquence.
	Taille : 29 903 bp	Longueur totale du génome ARN du SARS-CoV-2 (≈30 kb). PMC	Confirme qu'on traite bien un virus complet ; toute couverture manquante est facile à repérer.
	GC % : 38 %	Pourcentage de guanines + cytosines dans le génome. PMC	Sert à détecter des contaminations (un GC% très différent peut révéler une autre espèce) et à

Course contre la montre génomique

Séquençage en Urgence - Évolution technologique 2003-2023

Paramètres d'analyse FastQC - Séquençage viral

Paramètres Généraux

Génome de référence: NC_045512.2

Taille: 29,903 bp

GC%: 38%

Seuils de Qualité

Phred minimum: Q20

Couverture min: 100×

Contamination max: 5%

Résultats Analyse

Variants identifiés: 7

Contamination humaine: 2.3%

ORFs complets: 11/11

Rubrique	Champ	Signification	Impact qualité
Seuils de qualité	Phred minimum : Q20	Score Phred $\geq 20 \Rightarrow$ probabilité d'erreur $\leq 1\%$ ($Q = -10 \log_{10} P_{err}$) pour chaque base.	Garantit que la quasi-totalité des lectures utilisées pour l'analyse variantielle sont fiables.
	Couverture min : 100×	Chaque nucléotide doit être lu ≥ 100 fois.	Réduit fortement les faux positifs/négatifs lors de l'appel de variants.
	Contamination max : 5 %	Seuil toléré pour la proportion de séquences non	Au-delà, l'échantillon pourrait être jugé trop « sale » et nécessiter une

Course contre la montre génomique

Séquençage en Urgence - Évolution technologique 2003-2023

Paramètres d'analyse FastQC - Séquençage viral

Paramètres Généraux

Génome de référence: NC_045512.2
Taille: 29,903 bp
GC%: 38%

Seuils de Qualité

Phred minimum: Q20
Couverture min: 100x
Contamination max: 5%

Résultats Analyse

Variants identifiés: 7
Contamination humaine: 2.3%
ORFs complets: 11/11

Rubrique	Champ	Signification	Lecture biologique
Résultats analyse	Variants identifiés : 7	Sept positions nucléotidiques différent par rapport à NC_045512.2.	Correspond à la dérive accumulée de votre souche ; la combinaison de mutations sert au typage (PANGO lineage, Nextstrain clade, etc.).
	Contamination humaine : 2,3 %	2,3 % des lectures alignent sur le génome humain (ou autres contaminants).	Sous le seuil (5 %) ⇒ échantillon considéré « propre ».
	ORFs complets : 11/11	Les 11 cadres de lecture ouverts canoniques (ORF1ab, S, 3a, E, M, G, 7a, 7b, 8, N).	Confirme l'intégrité génomique ; indispensable pour des études fonctionnelles

Course contre la montre génomique

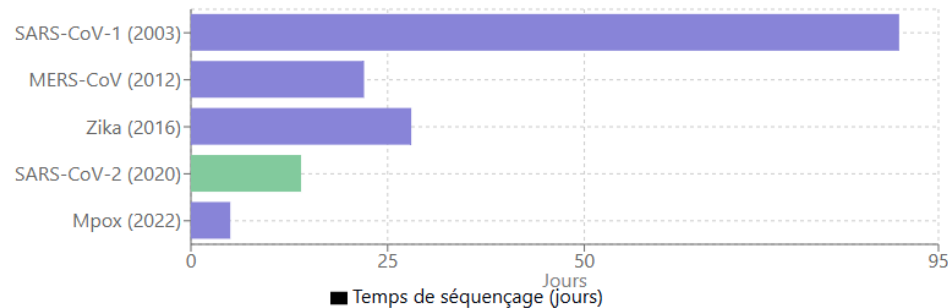
Séquençage en Urgence - Évolution technologique 2003-2023

Évolution du temps de séquençage

14 jours

pour le SARS-CoV-2 (2020)
vs 90 jours pour SARS-CoV-1 (2003)

Amélioration de 84%

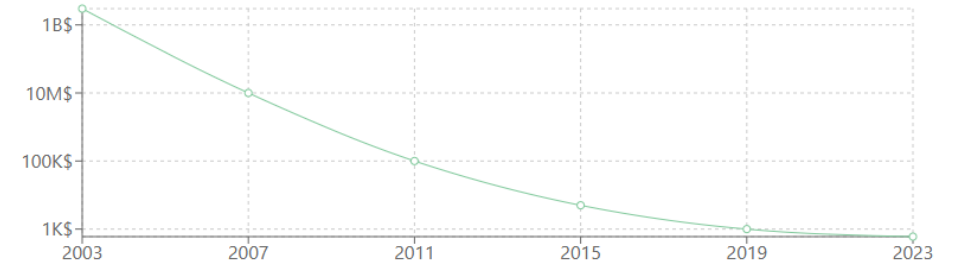


Évolution du coût de séquençage

3B\$ → 600\$

Coût du séquençage d'un génome humain

Baisse de 99.99998%



Coût du séquençage génome complet

Facteurs de réduction: amélioration des chimies, miniaturisation, automatisation
La loi de Moore a été dépassée pour le séquençage (réduction plus rapide)

Technologiques

- Parallélisation massive (millions de reads)
- Séquençage temps réel (Nanopore)
- Chimies haute-fidélité
- Miniaturisation des appareils

Organisationnels

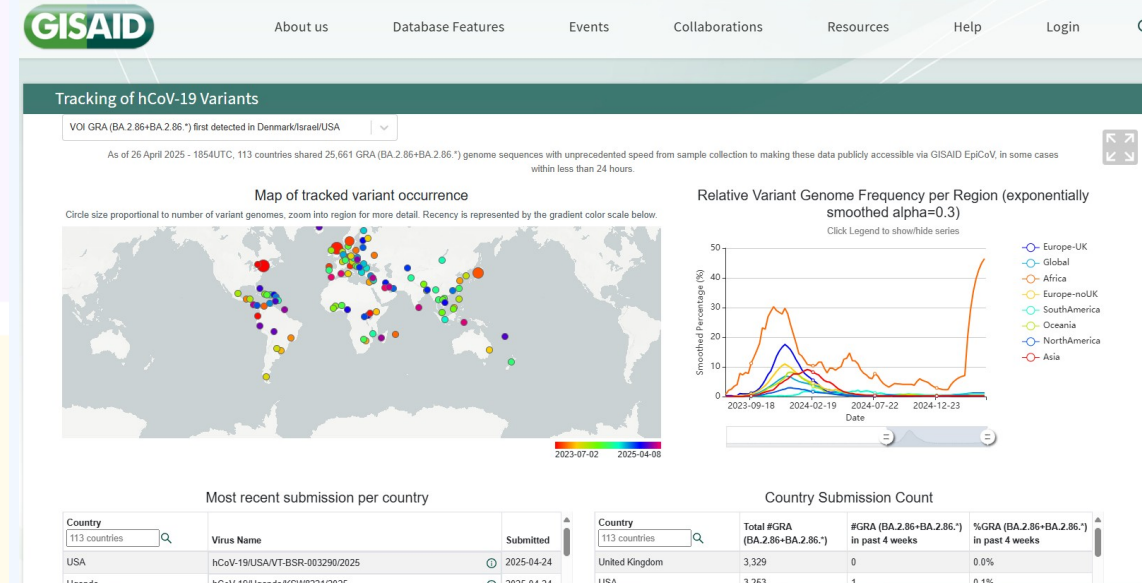
- Partage international (GISAID)
- Financement d'urgence ciblé
- Protocoles standardisés
- Formation équipes dédiées

Informatiques

- Algorithmes d'assemblage optimisés
- Analyses cloud distribuées
- Pipeline automatisés (Nextflow)
- Intelligence artificielle (basecalling)

Analyses comparées

- SARS-CoV-1: analyse manuelle
- SARS-CoV-2: 200K+ génomes/an
- Mpox: 95% réduction délai
- Surveillance continue (eaux usées)

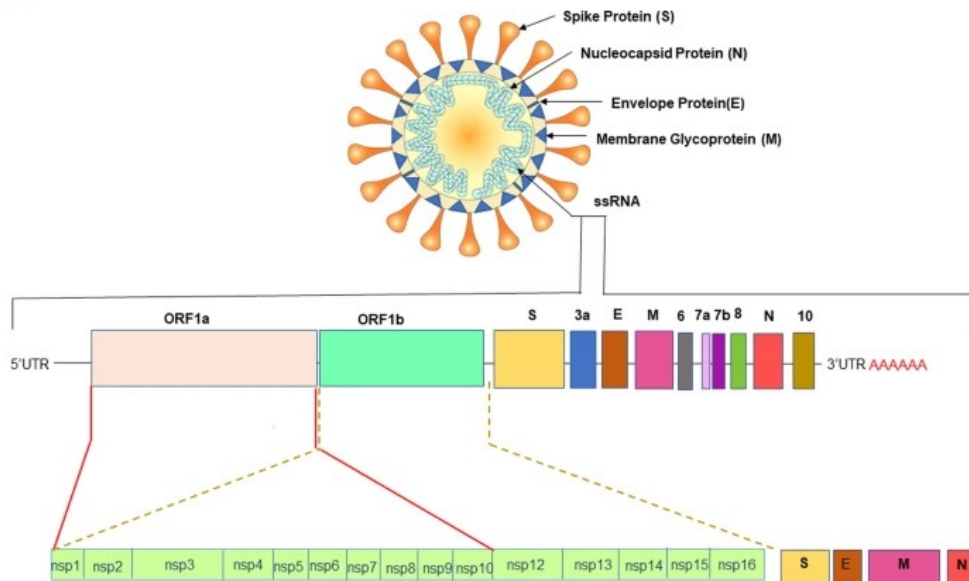


Course contre la montre génomique

Séquençage en Urgence - Évolution technologique 2003-2023

Le 24 janvier 2020, le ministère français chargé de la santé confirmait trois premiers cas de patients touchés par le coronavirus SARS-CoV-2. Le 29 janvier 2020, l'Institut Pasteur, en charge de la surveillance des virus respiratoires en France, a séquencé intégralement le génome du coronavirus dit « 2019-nCoV ». C'est une première en Europe depuis le début de l'épidémie, permise grâce à la Plateforme de microbiologie mutualisée (P2M) de l'Institut Pasteur.

Fig. 2



Schematic presentation of the SARS-CoV-2 genome Structure. SARS-CoV-2 has a spherical structure. The virus has an outer lipid envelope, covered with spike glycoprotein. The SARS-CoV-2 represents a typical Betacoronavirus genome organisation. The full-length RNA genome comprises of approximately 29,903 nucleotides and has a replicase complex (comprised of ORF1a and ORF1b) at the 5'UTR. The ORF1a encodes for nsp1–nsp10, while ORF1b encodes for nsp1–nsp16. Four genes that encode for the Structural proteins: Spike gene, Envelope gene, Membrane gene, Nucleocapsid gene and a poly (A) tail at the 3'UTR. The accessory genes are distributed in between the structural genes

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

[FASTA](#) [Graphics](#)

[Go to:](#) ☐

LOCUS	NC_045512	29903 bp ss-RNA	linear	VRL 18-JUL-2020
DEFINITION	Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome.			
ACCESSION	NC_045512			
VERSION	NC_045512.2			
DBLINK	BioProject: PRJNA485481			
KEYWORDS	RefSeq.			
SOURCE	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)			
ORGANISM	Severe acute respiratory syndrome coronavirus 2 Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Coronidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus; Severe acute respiratory syndrome-related coronavirus.			
REFERENCE	1 (bases 1 to 29903)			
AUTHORS	Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C. and Zhang, Y. Z.			
TITLE	A new coronavirus associated with human respiratory disease in China			
JOURNAL	Nature 579 (7798), 265-269 (2020)			
PUBMED	32015508			

Algorithmes d'Alignement et Analyse

Exploration des outils bioinformatiques pour l'analyse phylogénétique

VISUALISATION

Alignement multiple de séquences (MAFFT)



☐ Région hautement conservée

☐ Site de liaison au récepteur

☐ Point chaud de mutation

☐ Marqueur spécifique de clade

PARAMÈTRES Paramètres critiques des algorithmes

BLAST

Matrix:	BLOSUM62
Matrice de substitution d'acides aminés	
Gap Open:	11
Pénalité d'ouverture de gap	
Gap Extend:	1
Pénalité d'extension de gap	
E-value:	10^{-5}
Seuil de significativité statistique	
Word Size:	6
Taille minimum des correspondances exactes	

BLAST (Basic Local Alignment Search Tool) est un algorithme de bioinformatique emblématique, conçu pour rechercher des similarités locales entre séquences biologiques (ADN, ARN, protéines). Contrairement à MAFFT (alignement multiple global), BLAST est orienté vers la détection rapide de correspondances locales significatives entre une séquence requête et une base de données.

- **BLOSUM** (*Blocks Substitution Matrix*) : matrices développées à partir de blocs d'alignement de séquences protéiques conservées.

- **BLOSUM62** est la matrice par défaut pour BLASTP car elle offre un compromis robuste pour des séquences **modérément similaires**.

- **Utilisation** : si vous cherchiez des séquences très divergentes (évolution lointaine), on utiliserait BLOSUM45 ; pour des séquences proches BLOSUM80 ou BLOSUM90

OPTIMISATIONS Accélération des analyses

- Parallélisation des requêtes BLAST sur infrastructure HPC (×18 plus rapide)
- Utilisation de DIAMOND pour alignements protéiques préliminaires (×500 plus rapide que BLASTP)
- Indexation avec HMMER pour recherche de motifs conservés (sensibilité accrue de 35%)
- Pré-filtrage avec MinHash pour éliminer les séquences non pertinentes (réduction de 80% du temps)
- Implémentation GPU-MAFFT pour alignements multiples (×12 plus rapide)

MAFFT

Algorithm:	G-INS-i
Alignement global itératif pour haute précision	
Scoring Matrix:	200PAM/k=2
Matrice pour séquences éloignées	
Gap Open Penalty:	1.53
Pénalité d'ouverture de gap	
Offset:	0.123
Valeur de décalage pour les gaps terminaux	
Max Iterations:	1000
Nombre maximum d'itérations pour convergence	

G-INS-i : Global Iterative Refinement. Il s'agit du mode le plus précis de MAFFT, mais aussi l'un des plus coûteux en temps de calcul. Il est basé sur un alignement global (alignement sur toute la longueur des séquences), contrairement à L-INS-i ou E-INS-i qui tolèrent mieux des insertions massives ou réarrangements.

Algorithmes d'Alignement et Analyse

Exploration des outils bioinformatiques pour l'analyse phylogénétique

VISUALISATION Alignement multiple de séquences (MAFFT)

1 11 21 31 41 51 61 71 81 91

SARS-CoV-2 (Wuhan- ATGTTTGTCTTTCTTGTCTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTACAACCAAGAACTCAATTACCCCTGCATAC

Indices pour l'analyse:

- La position 60 montre une substitution C→G spécifique à la lignée humaine
- Le site de liaison au récepteur (positions 50-60) présente 92% de similarité avec RaTG13
- La région 70-85 contient la plus grande diversité de séquence entre les coronavirus
- L'identité globale entre SARS-CoV-2 et RaTG13 est de 96.2%
- Pangolin-CoV partage des caractéristiques avec SARS-CoV-2 aux positions 75-80

MERS-CoV ATGCATGTTTCTTTATTCTTACGATAACAACTCACTAGTGCCTCTTCTTATCAGTACTGTATTAGTACTTATCATACCTC
ACAATCTCCGAGGGTTTGTTCACCTGAGGGCGTTTA

☒ Région hautement conservée

☐ Site de liaison au récepteur

☐ Point chaud de mutation

☐ Marqueur spécifique de clade

MINI-EXERCICE Interprétation d'un résultat d'alignement

Instructions pour l'analyse:

1. Identifiez les positions montrant une variation entre SARS-CoV-2 et RaTG13
2. Comparez ces positions avec les séquences de Pangolin-CoV
3. Déterminez les mutations spécifiques à la lignée humaine
4. Évaluez le degré de conservation des sites de liaison au récepteur
5. Calculez le pourcentage d'identité entre les trois premiers virus

Question directrice

"Que nous révèle cet alignement sur l'origine du virus ?"

1. **Proximité phylogénétique avec RaTG13** : L'identité globale de 96,2% entre SARS-CoV-2 et le coronavirus de chauve-souris RaTG13 indique clairement que ce dernier est le plus proche parent connu du virus responsable de la COVID-19. Cette forte similarité confirme l'origine zoonotique du virus dans les populations de chauves-souris.
2. **Différences dans le site de liaison au récepteur** : Bien que globalement similaire (92% de similarité avec RaTG13), le site de liaison au récepteur (positions 50-60) présente des différences significatives, notamment la substitution C→G à la position 60, spécifique à la lignée humaine. Ces modifications sont cruciales pour comprendre l'adaptation du virus à l'hôte humain.
3. **Mosaïque génétique avec des éléments du Pangolin-CoV** : L'alignement révèle que le Pangolin-CoV partage certaines caractéristiques avec SARS-CoV-2 aux positions 75-80, précisément dans des régions où SARS-CoV-2 diffère de RaTG13. Cela suggère soit un événement de recombinaison, soit une évolution convergente impliquant potentiellement un hôte intermédiaire.
4. **Distance significative avec SARS-CoV-1 et MERS-CoV** : L'alignement montre clairement que SARS-CoV-2 est génétiquement distant des précédents coronavirus humains épidémiques, ce qui explique ses propriétés uniques de transmission et pathogénicité.
5. **Zones de grande diversité** : La région 70-85, identifiée comme point chaud de mutation, présente la plus grande diversité entre ces coronavirus. Cette variabilité génétique illustre le processus évolutif continu et la capacité d'adaptation de ces virus.

En conclusion, cet alignement soutient fortement l'hypothèse d'une origine naturelle du SARS-CoV-2 à partir de coronavirus de chauve-souris, avec une possible implication d'hôtes intermédiaires (potentiellement le pangolin) ou d'événements de recombinaison qui ont contribué à son adaptation à l'hôte humain. La présence de mutations spécifiques dans le site de liaison au récepteur explique également sa capacité à infecter efficacement les cellules humaines.

Partage des Données

GISAID et GenBank : démocratisation des données

Interface de soumissionGISAID/GenBank

Soumettre

Annuler

Interface de soumission GISAID/GenBank

Éthique : Dilemmes du partage de données sensibles

- **Propriété intellectuelle vs bien commun** : Tension entre les droits des chercheurs et les besoins de santé publique mondiale
- **Implications géopolitiques** : Inégalités d'accès aux ressources et partage équitable des bénéfices

Témoignage

"Le partage rapide des séquences génomiques durant la pandémie a permis de développer des tests diagnostiques et des vaccins en un temps record. C'est la preuve que la science ouverte sauve des vies."

— Dr. Sophie Dubois, Institut Pasteur

L'évolution du partage des données et ses bénéfices pour le bien commun

Évolution historique

- D'un modèle fermé à une science ouverte et collaborative
- Des données confinées aux laboratoires vers des plateformes accessibles mondialement
- Accélération du mouvement suite aux crises sanitaires mondiales

Bénéfices scientifiques

- Accélération considérable du rythme des découvertes
- Exemple COVID-19 : du séquençage aux vaccins en moins d'un an
- Validation croisée des résultats à l'échelle internationale

Impacts économiques

- Démocratisation des capacités d'innovation
- Émergence de nouveaux acteurs dans des pays en développement
- Coexistence possible entre partage et valorisation commerciale

Statistiques

742K Séquences déposées	VS	154K Partagées précédemment
-----------------------------------	----	---------------------------------------

Dimension éthique

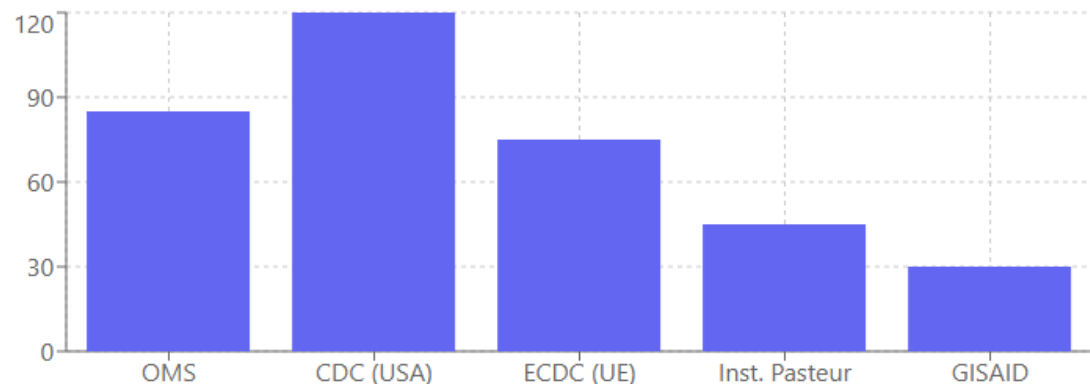
- Réduction des inégalités d'accès au savoir scientifique
- Défis persistants : reconnaissance des contributeurs
- Question du partage équitable des bénéfices issus des données

Perspectives d'avenir

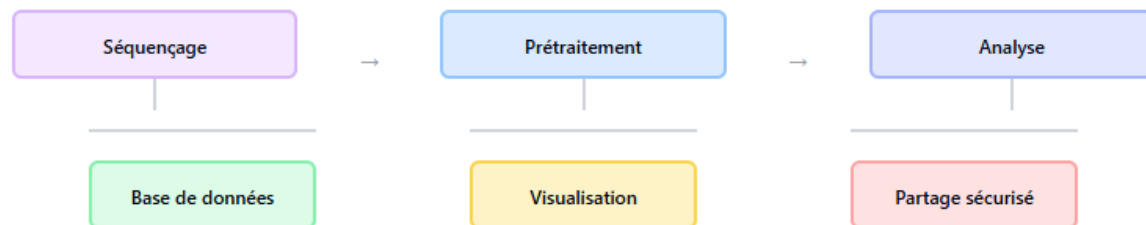
- Vers des standards internationaux de partage de données
- Nécessité de cadres juridiques adaptés
- Potentiel transformateur pour la santé mondiale

Impact de l'Infrastructure IT

Investissements IT d'urgence (M€)



Pipeline d'analyse de données génomiques



Points clés

- **Capacité de traitement parallèle :**
Analyse simultanée de milliers de séquences génomiques grâce aux clusters HPC et technologies cloud
- **Stockage et distribution des données :**
Systèmes distribués permettant l'accès et le partage rapide des données entre chercheurs internationaux
- **Sécurisation des informations sensibles :**
Protocoles de chiffrement et contrôles d'accès adaptés aux données de santé publique

Applications industrielles

Parallèle avec le bioprocessing de souches industrielles



Optimisation de souches

Analyses génomiques pour sélectionner et améliorer les souches microbiennes utilisées dans la production industrielle



Détection de contaminations

Surveillance en temps réel des bioréacteurs pour identifier les contaminations microbiennes avant qu'elles n'affectent la production

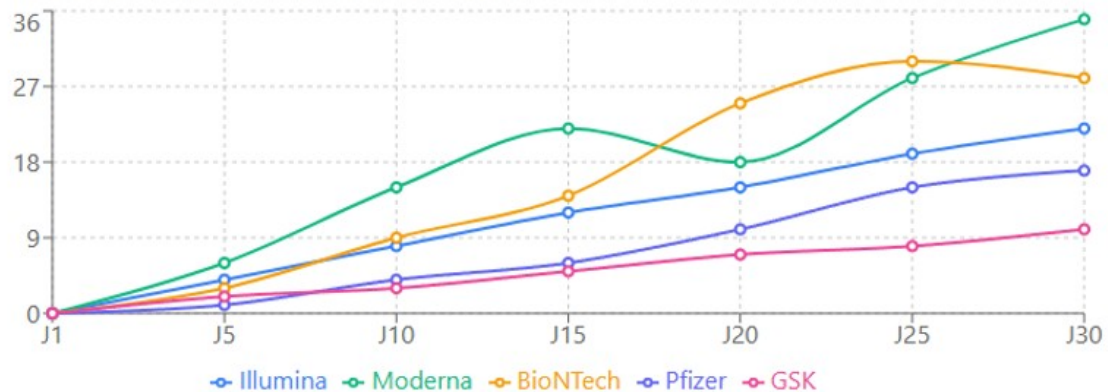
Microsoft Genomics

Powering genomic data analysis on Azure

Point Financier et Transition

Premiers signaux financiers

Mouvements boursiers initiaux (% variation)



Points clés

- Valorisation des entreprises disposant de plateformes de séquençage
Hausse significative des actions d'Illumina (+22%) et autres acteurs du séquençage génomique
- Premiers investissements spéculatifs dans les biotechs
Afflux de capitaux vers les sociétés possédant des technologies d'ARNm et de vectorisation virale

Perspective d'analyste

"Nous assistons à un phénomène sans précédent où le marché anticipe déjà la course au vaccin. Les investisseurs ne parient plus seulement sur les résultats, mais sur la capacité d'innovation et la rapidité de déploiement des technologies de rupture. Cette crise pourrait redessiner durablement le paysage pharmaceutique mondial."

Analyste senior chez Goldman Sachs

Transition

Ces mouvements financiers précoces reflètent la mobilisation sans précédent du secteur privé face à la menace émergente.

Phase actuelle:

Détection

Analyse génomique

Caractérisation du virus

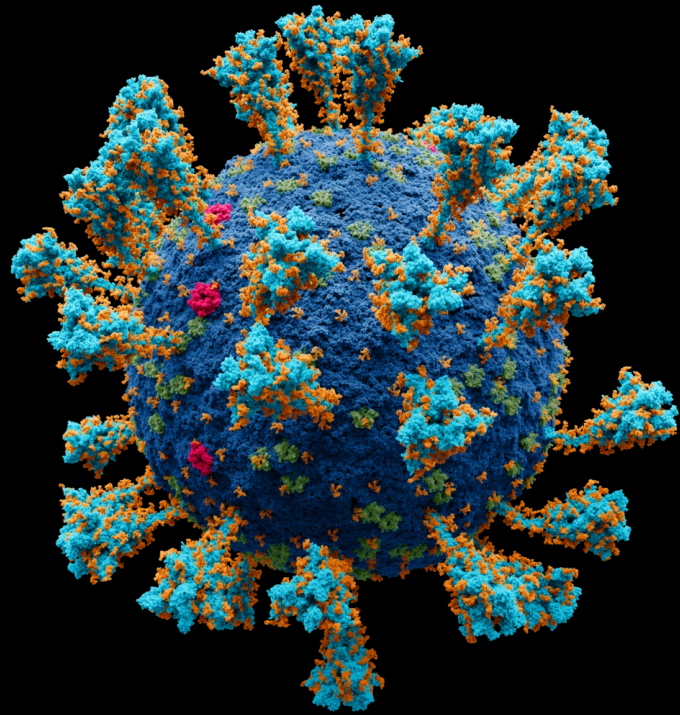
Développement vaccinal

Phase suivante:

Notre prochaine section se concentrera sur la caractérisation approfondie du virus

Qui est le mieux « outillé » pour tirer parti de la formidable vague d'investissements public/privé pour faire face à la crise sanitaire ?

Plateformes vaccinales à ARNm (Moderna / BioNTech – Pfizer), Tests PCR, perte de vitesse des « pharmas » traditionnelles (sauf Astra Zeneca et son vaccin à adénovirus de chimpanzé)...



J31 à J60 : De la séquence à la structure

J1 : Alerte sanitaire initiale

J7 : Isolement du virus

J12 : Premier séquençage

J15 : Partage GISAID

J18 : Pipeline d'analyse

J23 : Premiers modèles 3D

J27 : Identification protéine S

J30 : Séquence consensus

Figure 1

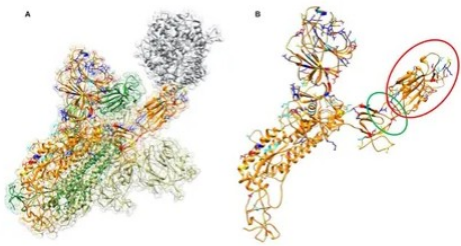


Figure 1. Structure of SARS-CoV-2 S proteins and ACE2 binding domain. **(A)** The structure of S protein trimer binding with ACE2 binding domain. ACE2 is shown in gray color. The three S protein monomers are represented in yellow, orange, and green colors, respectively. The mutations from SARS-CoV to SARS-CoV-2 in this study are labeled in four colors on a single chain of S protein: Red represents residues which are mutated to be more negative; Blue represents residues which are mutated to be more positive; yellow represents residues which are mutated from polar to hydrophobic; cyan represents residues which are mutated from hydrophobic to polar. **(B)** Structure of a single S protein monomer. The RBD shown in red circle is flipping out to reach ACE2. The green circle region highlights the hinge between RBD and the rest of S protein.

(On connaissait déjà SARS SARS COV1

Objectifs de cette phase

- **Caractérisation structurale complète**
Identification des domaines fonctionnels et motifs structuraux clés
- **Identification des cibles potentielles**
Recherche de sites actifs et régions accessibles pour les médicaments
- **Modélisation des interactions hôte-pathogène**
Simulation des mécanismes d'entrée cellulaire et réplication

Question interactive

Quelles méthodes d'analyse structurale connaissez-vous?

Cristallographie aux rayons X

Cryo-microscopie électronique

Résonance magnétique nucléaire (RMN)

Diffusion des neutrons

Modélisation in silico

Statistique clé

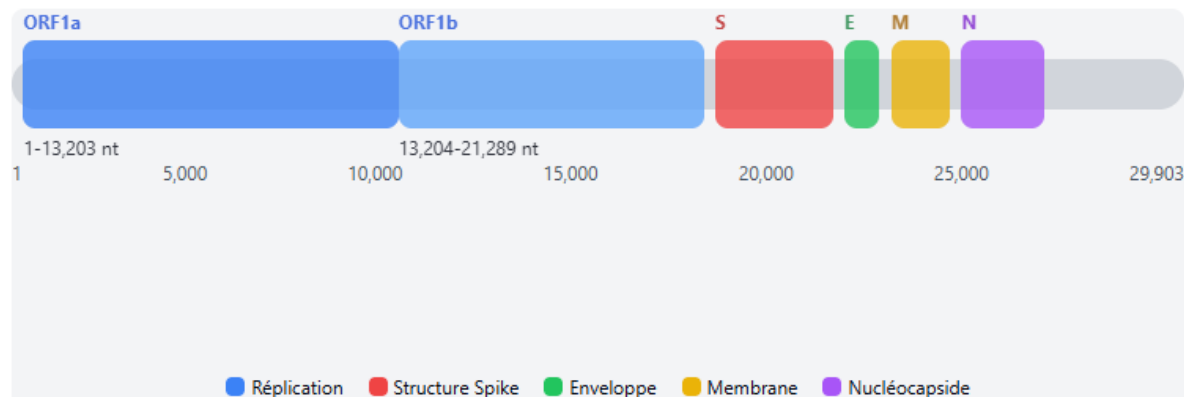
287

Laboratoires mobilisés sur l'analyse structurale



Au-delà de la séquence brute

Génome annoté du SARS-CoV-2



Interface Clustal Omega/MUSCLE

CLUSTAL O(1.2.4) multiple sequence alignment

```
SARS-CoV-2 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFPFFS
SARS-CoV MFIFLLFLTLSGSDLDRCTTFDDVQAPNYTQHTSSMRGVVYPDEIFRSDTLTLTQDLFL
MERS-CoV MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIYPQ
*:::.* . . . . : * :.* : . :
SARS-CoV-2 NVTWFHAIHVS GTNGTKRFDNPVLPFNDGVYFAST EKSNIIRGWIFGTTLD SKTQSL LIV
SARS-CoV KCNTWFHAIHVS GTNGIKRFDNPVLPFNDGVYFAST EKSNIIRGWIFGTTLD TKTQSL LI
MERS-CoV GRTYSNITITYQGLF-PYNDKVFSSVLH LTMESDNKFKNGILLVNPYSNMWFHDILYY
.* : *:: . . . . : * . : :.* : * :::.* : * :.*::
```

Région conservée de la protéine Spike (acides aminés 1-120)

Points techniques

• Annotation automatique vs manuelle

Les outils bioinformatiques (Prokka, RAST) offrent une première passe d'annotation automatique, mais l'expertise humaine reste essentielle pour affiner les résultats

• Détection des ORFs (cadres de lecture ouverts)

ORF (*Open Reading Frame*) : Séquence d'ADN/ARN sans codon stop prématuré, débutant par un codon d'initiation (START) et se terminant par un codon STOP, permettant la traduction en protéine. Identification par algorithmes de prédiction et validation par conservation évolutive.

• Identification des gènes clés (S, E, M, N)

Utilisation d'homologie de séquence avec d'autres coronavirus et validation fonctionnelle par expression protéique

Apports de l'alignement multiple vs simple

Alignement multiple

- ✓ Identification des domaines conservés évolutivement
- ✓ Détection des sites fonctionnels cruciaux
- ✓ Prédiction de la structure tertiaire
- ✓ Analyse phylogénétique robuste

Alignement simple

- ✓ Identification de similarités basiques
- ✓ Calcul rapide de pourcentages d'identité
- ✓ Détection d'insertions/délétions simples
- ✗ Incapacité à capturer les relations complexes

Clustal Omega : Outil d'alignement multiple de séquences de dernière génération utilisant des arbres guide HMM, permettant l'alignement rapide de milliers de séquences avec une précision supérieure.

MUSCLE (Multiple Sequence Comparison by Log-Expectation) : Algorithme itératif d'alignement optimisant le score selon deux mesures de similarité différentes.

Fonctionnalités principales :

- Alignement de séquences de nucléotides ou protéines issues de différentes espèces
- Identification des régions conservées et des mutations spécifiques
- Génération de matrices de distance pour analyses phylogénétiques
- Visualisation des motifs de conservation par symboles de consensus (* : .)
- Exportation vers d'autres outils de prédiction structurale ou fonctionnelle

L'arbre généalogique viral

Points méthodologiques

• Infrastructure d'exécution

Calcul haute performance : Exécution sur clusters HPC institutionnels et services cloud (AWS, Google Cloud, Azure) plutôt que sur les séquenceurs eux-mêmes. Frameworks comme Galaxy et Nextflow permettent l'automatisation et la reproductibilité des pipelines d'analyse.

• Algorithmes IQ-TREE et MEGA

IQ-TREE : Inférence phylogénétique par maximum de vraisemblance avec sélection automatique de modèle et tests ultrarapides de bootstrap.

MEGA (Molecular Evolutionary Genetics Analysis) : Suite intégrée pour l'alignement de séquences, l'inférence d'arbres évolutifs et les tests d'hypothèses évolutives.

• Modèles d'évolution moléculaire

Formules mathématiques (GTR+G+I, HKY85, JTT) qui décrivent la probabilité des différents types de substitutions nucléotidiques ou d'acides aminés au cours de l'évolution, tenant compte des variations de taux entre sites.

• Apport de l'IA et algorithmes avancés

Apprentissage profond : RAXML-NG et PhyloGFN utilisent des réseaux neuronaux pour accélérer l'exploration de l'espace des arbres possible. **Réseaux génératifs** : Inférence d'arbres par GAN et VAE pour des ensembles de données massifs de séquences (> 100k). **Algorithmes de graphes** : BayesPhylogenies et BEAST utilisent l'inférence bayésienne pour quantifier l'incertitude.

Applications en biotech industrielle



Analyse de souches de production

Les mêmes méthodes phylogénétiques permettent d'optimiser des souches industrielles (ex: *Pichia pastoris*) en:

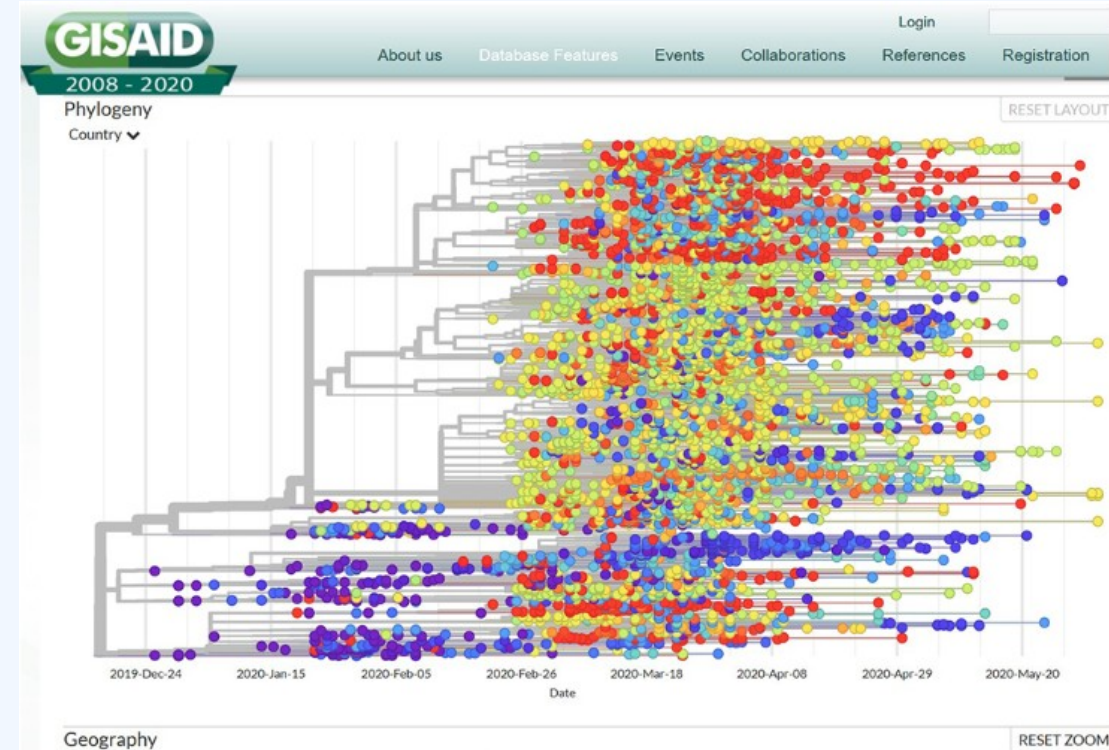
- Identifiant les lignées à haut rendement
- Corrélant génotype et performances de production
- Sélectionnant des souches résistantes aux stress industriels



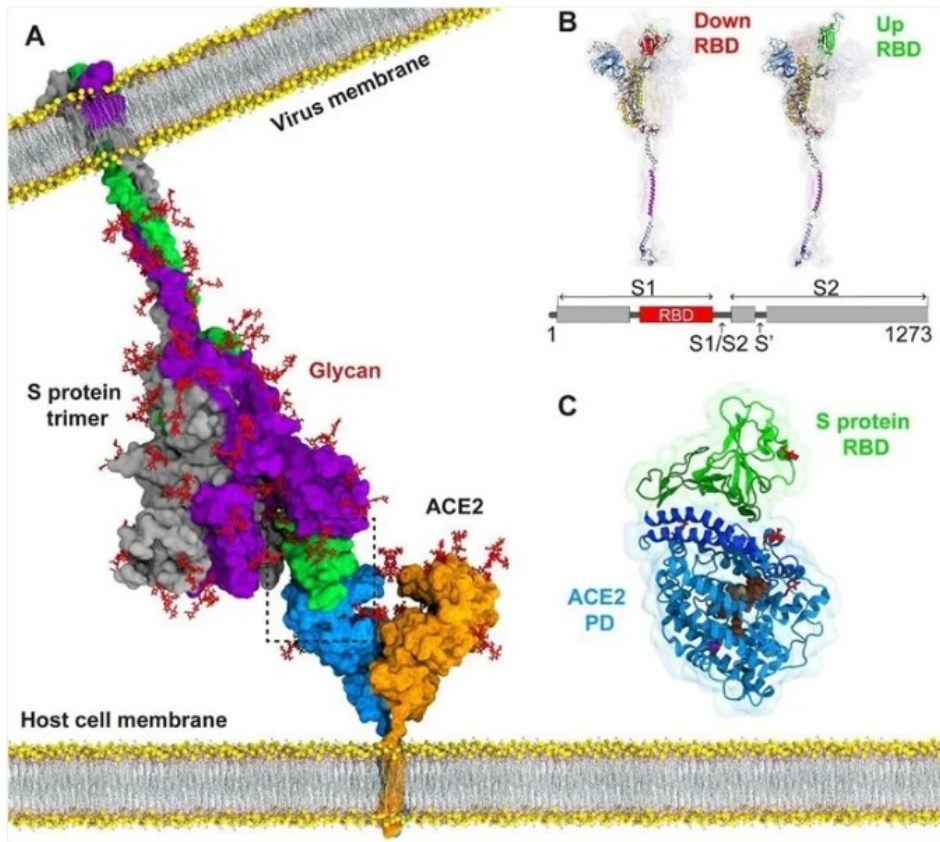
Traçabilité des modifications génétiques

Utilisation critique pour:

- Suivre l'histoire des modifications génétiques successives
- Documenter les changements pour conformité réglementaire
- Identifier la source de dérives de production
- Assurer la propriété intellectuelle des souches développées



La clé d'entrée virale



Atomic model for binding of the SARS-CoV-2 S protein to the ACE2 receptor on the host cell membrane. (A) The structure of the full-length S protein in complex with ACE2. The S protein is a homotrimer (green, purple, and grey) and embedded into the viral membrane. ACE2 is a homodimer (blue and orange) and embedded into the host cell membrane. The full length structure of the S protein in complex with ACE2 was modeled using the full length S protein model (48) and the crystal structure of the S protein RBD in complex with ACE2 (PDB ID: 6M17). Both proteins were manually inserted into the membrane by their transmembrane domains. (B) The structure of an S protomer in the down and up position of its RBD. S1/S2 and S2' are the cleavage sites of the S protomer upon ACE2 binding. (C) MD simulations were performed for RBD of the S protein in complex with the PD of ACE2. Catalytic residues of ACE2, glycans, and Zn²⁺ and Cl⁻ ions are shown in brown, red, yellow and purple, respectively.

Annotations des domaines clés

■ Domaine RBD (Receptor Binding Domain)

Région d'environ 200 acides aminés (aa 319-541) responsable de la liaison directe au récepteur ACE2 humain. Cible principale des anticorps neutralisants et site majeur de mutations adaptatives.

■ Sites de glycosylation

22 sites N-glycosylés qui forment un "bouclier" protecteur contre la reconnaissance immunitaire. Ces glycanes contribuent également à la stabilité structurale et peuvent influencer la conformation.

■ Régions conservées

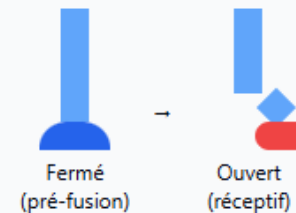
Domaines S2 et sites de clivage protéolytique hautement conservés parmi les coronavirus. Ces régions sont essentielles au mécanisme de fusion membranaire et représentent des cibles potentielles pour antiviraux à large spectre.

■ Régions variables

Boucles exposées au solvant, particulièrement dans le sous-domaine NTD (N-Terminal Domain), sujettes à une forte pression de sélection immunitaire et site de nombreuses mutations d'échappement dans les variants préoccupants.

Flexibilité conformationnelle et implications

États conformationnels



Implications fonctionnelles

- Équilibre dynamique entre états qui influence l'infectiosité
- Cibles d'anticorps spécifiques à certaines conformations
- Mutations stabilisant l'état "ouvert" augmentent l'affinité pour ACE2
- Base du mécanisme d'entrée cellulaire déclenché par liaison au récepteur

La clé d'entrée virale

Structure of the SARS-CoV-2 spike protein trimer.

(A) Left: side view of the trimeric spike ectodomain with 3 RBDs in the down-conformation; right: top view of the trimeric spike protein showing RBDs in gray, forest green, and orchid (PDB: 6VXX). (B) Left: side view of the trimeric spike ectodomain with 1 RBD in the up-conformation; right: top view of the trimeric spike protein showing 1 up RBD in gray (PDB: 7BNN). (C) A schematic layout of the spike protein is shown at the top. Right: structure of a monomer displaying the RBD in the open conformation. Spike protein structure shows the receptor-binding subunit S1 and the membrane-fusion subunit S2 separated by the furin-like protease site (S1/S2). Different subdomains of the spike protein are the NTD in green, RBD in gray containing RBM in cyan at their top, the fusion peptide in pink, second cleavage site S2' in red, and HR1 and HR2 in olive (PBB: 7BNN). The scissors represent the S1/S2 boundary at amino acid position 685. Left: The open conformation RBD highlights the 3 different regions: receptor-binding ridge, flat surface, and 443–450 loop of the RBM that form the ACE2-binding region. ACE2, angiotensin-converting enzyme 2; FP, fusion peptide; NTD, N-terminal domain; RBD, receptor-binding domain; RBM, receptor-binding motif; SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2.

More »

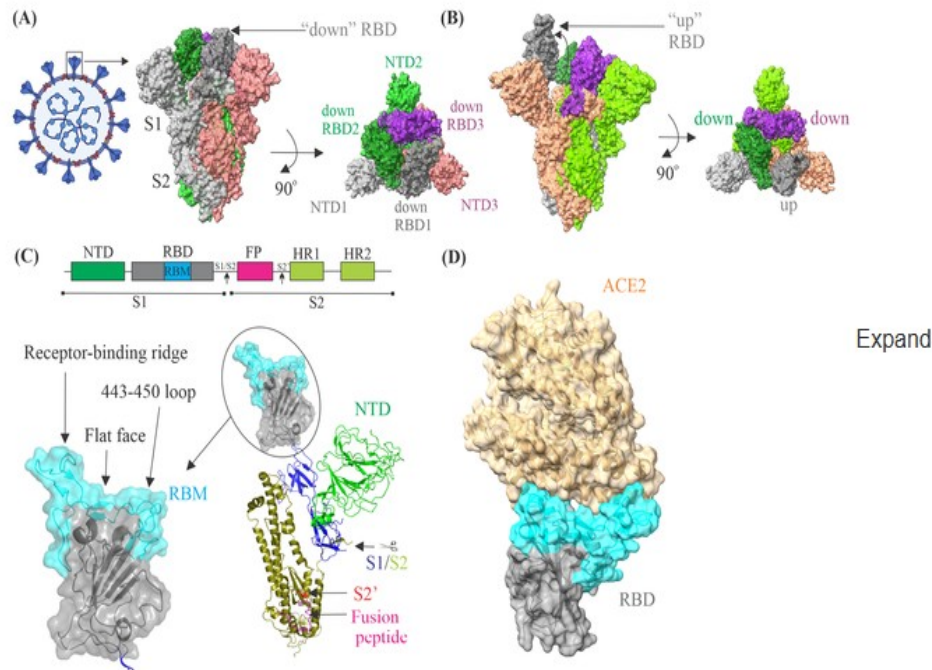
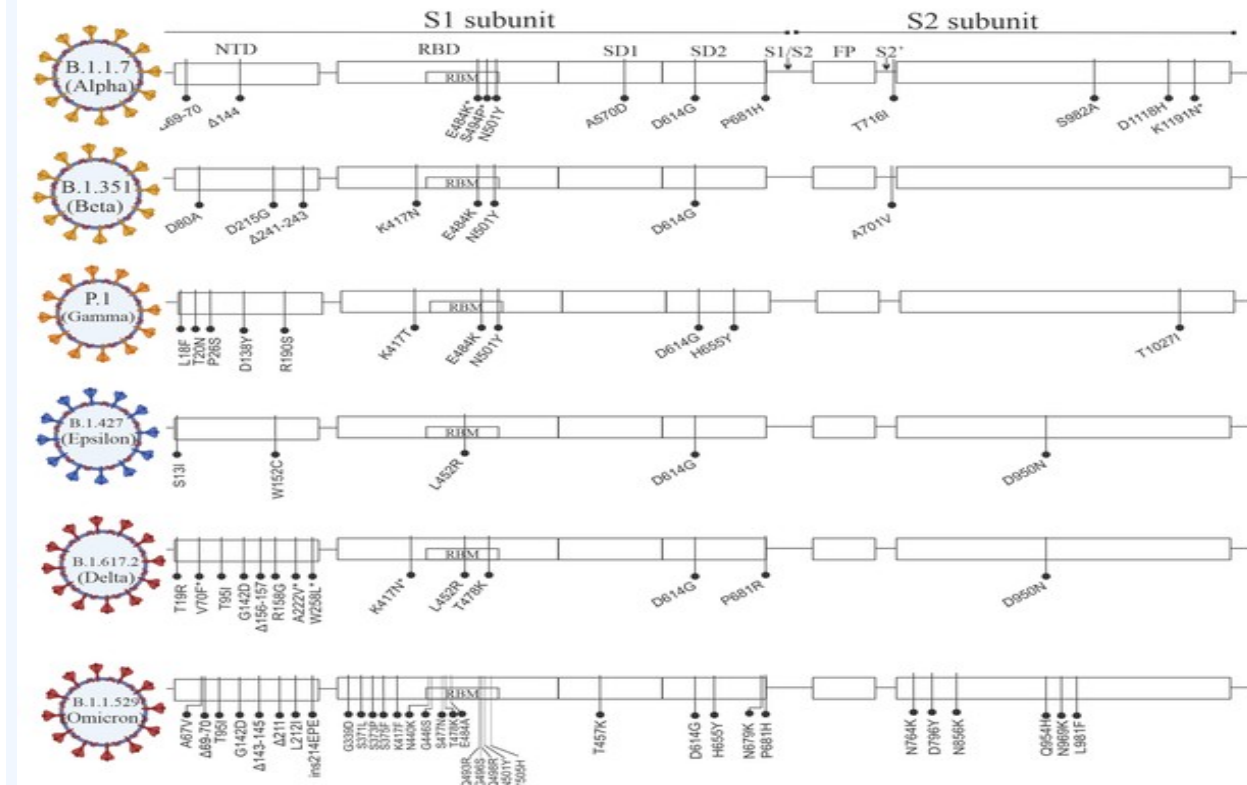
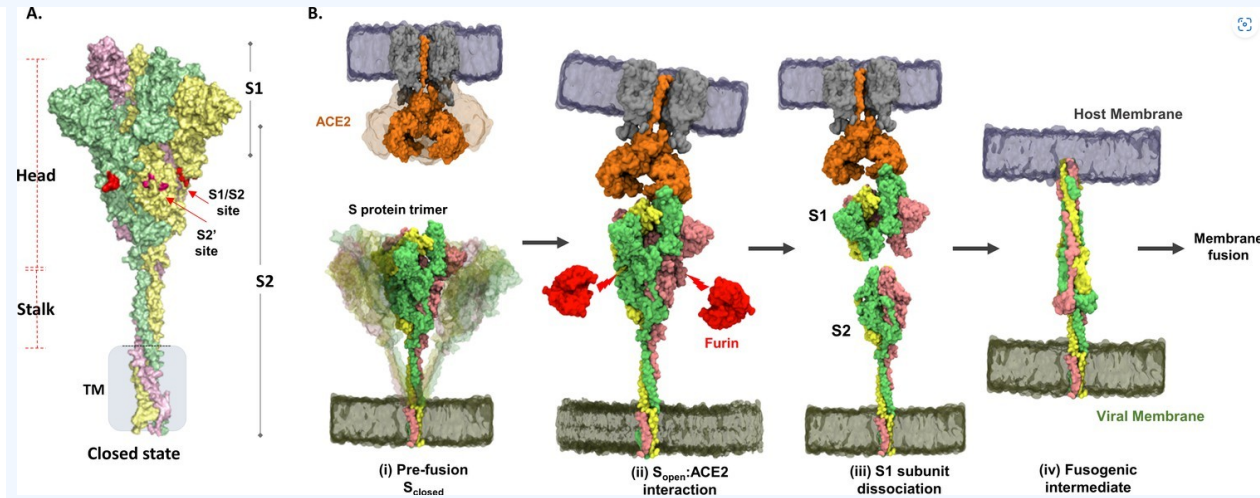


Fig 2.

Schematic overview of the SARS-CoV-2 variants.

The variant being monitored B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), and B.1.427 (Epsilon) and the variant of concern B.1.617.2 (Delta) and B.1.1.529 (Omicron) showing amino acid modifications in comparison to the ancestral Wuhan-Hu-1 sequence (NC_045512.2). FP, fusion peptide; NTD, N-terminal domain; RBD, receptor-binding domain; RBM, receptor-binding motif; SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2.



Apports de l'Intelligence Artificielle

Compréhension structurale et prédiction des mutations virales



Compréhension de la structure virale

- **AlphaFold/RoseTTAFold :**
Prédiction de la structure des protéines virales avec une précision atomique en heures au lieu de mois par cristallographie
- **Modèles d'apprentissage profond :**
Identification rapide des sites de liaison et des épitopes potentiels sur la protéine Spike
- **Simulations moléculaires accélérées :**
Dynamique moléculaire assistée par IA pour comprendre les changements conformationnels lors de l'infection
- **Cartographie des interactions :**
Modélisation complexe des interactions virus-récepteur ACE2 pour identifier les résidus critiques



Prédiction des mutations problématiques

- **Modèles évolutifs :**
Prédiction des trajectoires évolutives du virus avec identification précoce des variants préoccupants
- **Analyse d'échappement immunitaire :**
Anticipation des mutations permettant au virus d'échapper à l'immunité induite par les vaccins
- **Fitness viral :**
Évaluation in silico de l'impact des mutations sur la transmissibilité et la virulence
- **Surveillance génomique en temps réel :**
Algorithmes d'IA analysant des millions de séquences pour détecter l'émergence de nouveaux variants

Accélération critique



Réduction du temps d'analyse structurale de **plusieurs mois à quelques jours**

Précision prédictive



Identification précoce des variants Alpha, Delta et Omicron **avant leur domination épidémique**

Adaptation vaccinale



Reformulation des vaccins en **moins de 100 jours** pour répondre aux nouveaux variants

Algorithmes d'Intelligence Artificielle

Applications spécifiques en recherche vaccinale

Algorithme	Type	Applications	Impact
AlphaFold 2	Réseau neuronal profond + attention	Prédiction de la structure 3D des protéines virales	Structure de la protéine Spike avec précision <1Å
ESM-2 / ESM-MSA	Transformer bidirectionnel	Prédiction des effets des mutations sur la stabilité protéique	Identification des mutations stabilisantes pour les vaccins
NetMHC / DeepLigand	CNN + architecture RNN	Prédiction de la présentation des épitopes du SARS-CoV-2	Sélection d'antigènes optimaux pour vaccins ARNm
MuSyC / EVEscape	Modèle génératif + apprentissage par renforcement	Prédiction des mutations d'échappement immunitaire	Anticipation des variants problématiques (Delta, Omicron)
MoleculeKit / RFdiffusion	GNN + modèles de diffusion	Conception de systèmes de délivrance lipidiques	LNP optimisés pour la délivrance d'ARNm
MDsim / Folding@Home	Simulation accélérée par IA + computing distribué	Dynamique moléculaire virus-récepteur	Identification des résidus clés pour l'infection
CoVSurver / Nextstrain	Analyse phylogénétique + prédiction bayésienne	Surveillance génomique mondiale et évolution virale	Détection précoce des lignées émergentes

1

Séquençage et analyse génomique

Analyse des séquences virales par des modèles basés sur les transformers

2

Prédiction structurale

Application d'AlphaFold et ESM-2 pour prédire structures et interactions

3

Modélisation immunologique

Prédiction des épitopes et de la réponse immunitaire par CNNs spécialisés

4

Design vaccinal

Optimisation des séquences d'ARNm et des systèmes de délivrance

Performances comparatives : Algorithmes traditionnels vs IA

Prédiction structurale



Prédiction d'épitopes



Détection de variants problématiques

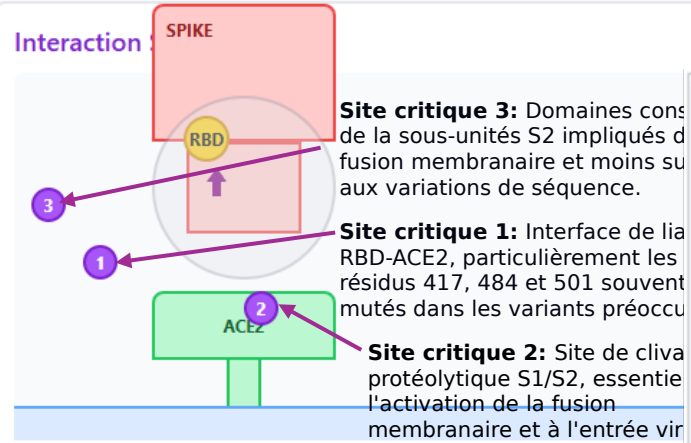


Optimisation ARNm



Cibler la spike : pourquoi et comment

Interaction



Points stratégiques du ciblage

• Accessibilité aux anticorps

La protéine Spike est abondamment exposée à la surface du virion, la rendant facilement accessible aux anticorps. Cette exposition confère un avantage pour la reconnaissance et le ciblage des virus circulants qui nécessiteraient une réponse cellulaire.

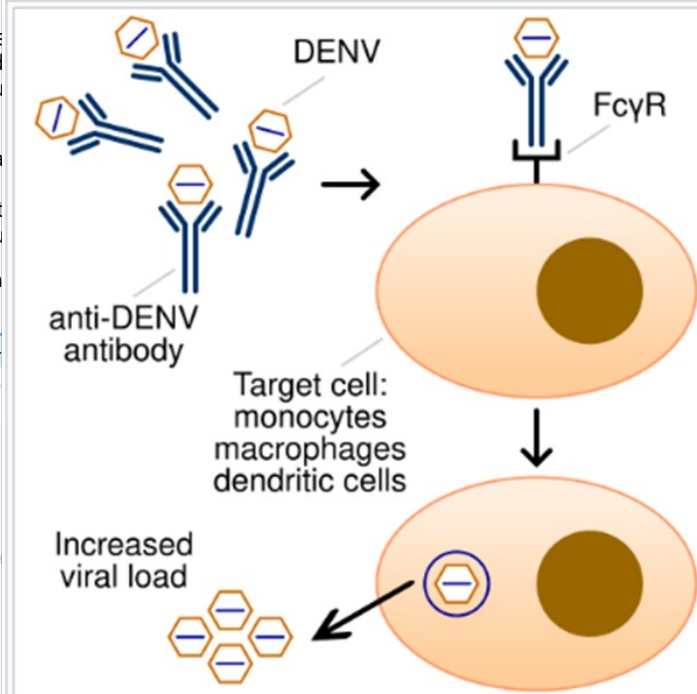
• Conservation relative de la séquence

Certaines régions de la Spike, particulièrement la sous-unité S2, sont relativement conservées en raison de contraintes fonctionnelles. Cibler ces régions augmente les chances de succès thérapeutiques.

• Rôle fonctionnel critique

La Spike est absolument essentielle pour l'entrée virale, sans alternative. Bloquer son activité neutralise efficacement l'infection, rendant les virus difficiles à éliminer sans compromettre la viabilité virale.

Pourquoi cibler la spike et pas une autre protéine ?



Un **anticorps** (ici représentés par des structures bleues en forme de Y) lié aux particules virales (DENV) puis aux récepteurs gamma Fc (FcγR) des **cellules immunitaires** (ici représentées par les ovales orangés) permettant au virus d'infecter cette cellule.

Antibody Dependent Enhancement

Avantages

- La surface virale essentielle à l'infection
- Facilement accessible
- Facilement neutralisants
- Facilement immunogènes

Inconvénients potentiels

- × Variabilité de séquence (surtout dans RBD)
- × Glycosylation complexe (masquage d'épitopes)
- × Potentiel de facilitation de l'infection (ADE)
- × Instabilité conformationnelle

Autres cibles potentielles: N (nucléocapside), M (membrane), et E (enveloppe)

Design d'inhibiteurs enzymatiques

Applications en biotechnologie

	Équivalent en design d'inhibiteurs
• Inhibiteurs compétitifs directs	
• Inhibiteurs non-compétitifs	
• Inhibition par encombrement stérique	
• Résistance aux médicaments	

La Spike partage de nombreuses similitudes avec le développement d'inhibiteurs enzymatiques: les deux nécessitent de connaître la structure, la fonction, et les contraintes évolutives de la cible.

— Prof. Elena Cortez, Structural Biology Department

Technologies de modélisation SARS-CoV-2

Techniques structurales

- **Cryo-EM:** Résolution sub-3Å de la Spike complète
- **Cristallographie RX:** Interface RBD-ACE2 à 2.45Å
- **Tomographie cryo-électronique:** Visualisation in situ

Méthodes in silico

- **MD:** Dynamique moléculaire (échelle µs)
- **AlphaFold2/RoseTTAFold:** Prédiction structures
- **Docking moléculaire:** HADDOCK, Autodock

Comparaison avec SARS-CoV (2003)

Affinité pour ACE2:

- Modifications structurales de la boucle de contact dans le RBD
- Résidus clés: L455, F486, Q493, S494, N501, Y505
- Site de clivage furine additionnel à l'interface S1/S2

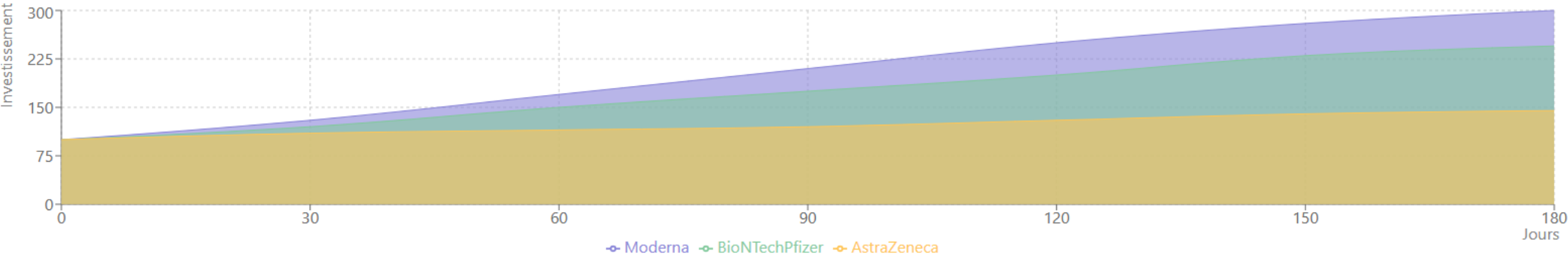
La plus forte affinité pour ACE2 et la présence d'un site de clivage furine contribuent à la transmissibilité accrue du SARS-CoV-2 comparé au SARS-CoV original.

10-20x supérieure

Intensification des Investissements IT

L'accélération des ressources computationnelles

Évolution des investissements IT/R&D depuis jour 0



Entreprises impliquées et leur stratégie

Entreprise	Stratégie
Moderna	+200% de capacité de calcul
BioNTech/Pfizer	Partenariat avec géants IT
AstraZeneca	Approche plus traditionnelle

Points clés

- Capacités HPC (High Performance Computing):** Augmentation significative des capacités de calcul haute performance
- Cloud computing d'urgence:** Déploiement rapide de ressources cloud pour faire face aux besoins immédiats
- Ressources dédiées à la modélisation:** Allocation prioritaire des ressources informatiques pour la simulation moléculaire

Synthèse et Transition

Bilan à J+60 : structure comprise, objectif vaccin

Récapitulatif

Acquis majeurs de cette phase:

- Structure virale complètement élucidée
- Cibles d'intervention identifiées
- Stratégie d'attaque définie
- Modèles d'interaction validés

Implications pour la conception vaccinale

Points faibles identifiés:

- Domaines protéiques exposés
- Régions stables et conservées

Stratégies potentielles d'immunisation:

- Ciblage multivalent
- Présentation optimisée des antigènes
- Adjuvants spécifiques

Lien avec l'ingénierie des protéines

Stabilisation de conformations:

Verrouillage des épitopes dans leur conformation immunogène optimale

Optimisation d'épitopes:

Amélioration de la reconnaissance immunitaire et de la présentation

Applicabilité en biocatalyse industrielle:

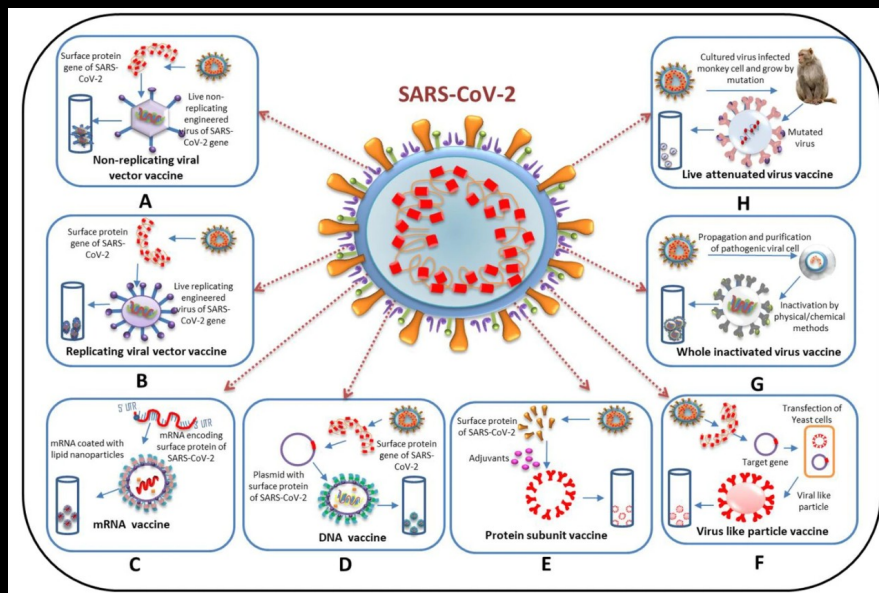
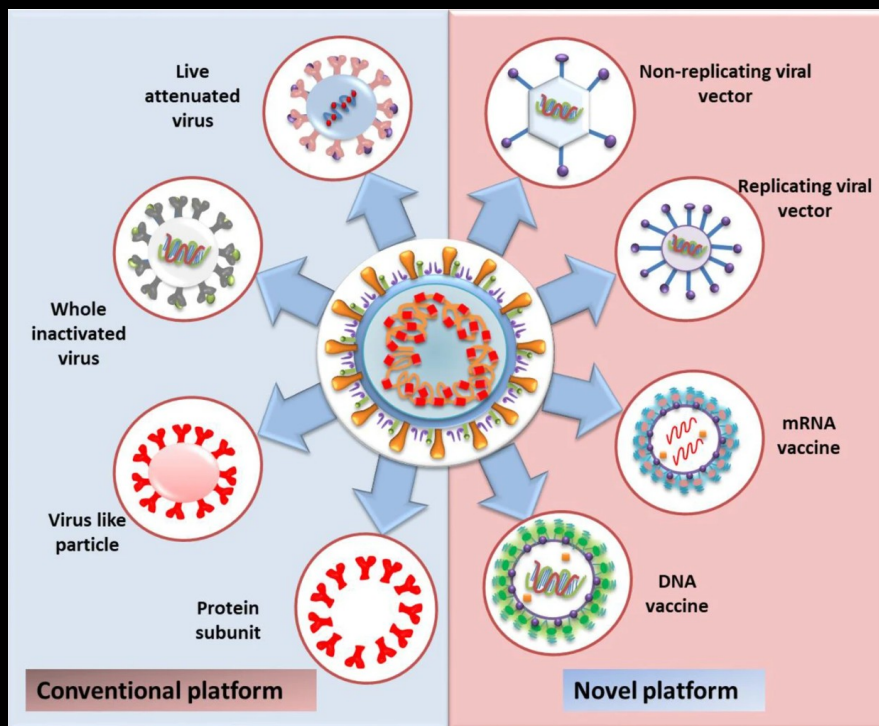
Transfert de technologies vers d'autres domaines biotechnologiques

Transition

Introduction à la conception de vaccins à ARNm:

Les connaissances structurales acquises nous permettent maintenant d'aborder la phase critique de conception des vaccins à ARNm avec une compréhension approfondie des cibles.





J61 à J120 : conception DES vaccins (strategies vaccinales)

1. Vaccins à ARNm

Acteurs principaux: Moderna, BioNTech/Pfizer, CureVac

Chronologie: Janvier 2020: Séquençage du SARS-CoV-2 / Février 2020: Conception des séquences d'ARNm

Mars-Avril 2020: Tests précliniques / Mai-Juillet 2020: Essais cliniques de phase 1/2

2. Vaccins à vecteurs viraux

Acteurs principaux: Oxford/AstraZeneca, Johnson & Johnson, Gamaleya (Sputnik V)

Chronologie: Février-Mars 2020: Adaptation des vecteurs / Avril-Mai 2020: Tests précliniques / Juin-Août 2020: Essais cliniques de phase 1/2

3. Vaccins à protéines recombinantes

Acteurs principaux: Novavax, Sanofi/GSK

Chronologie: Février-Avril 2020: Conception des protéines / Mai-Juillet 2020: Tests précliniques

4. Vaccins à virus inactives

Acteurs principaux: Sinopharm, Sinovac, Bharat Biotech

Chronologie: Janvier-Mars 2020: Isolation et culture virale / Avril-Mai 2020: Tests précliniques / Juin-Août 2020: Essais cliniques de phase 1/2

5. Vaccins à sous-unités virales

Acteurs principaux: Medicago, Clover Biopharmaceuticals

Chronologie: Mars-Mai 2020: Conception des antigènes / Juin-Août 2020: Tests précliniques / Septembre-Novembre 2020: Essais de phase 1/2

6. Vaccins à ADN

Acteurs principaux: Inovio, Zydus Cadila

Chronologie:

Février-Avril 2020: Conception des constructions ADN / Mai-Juillet 2020: Tests précliniques

Comparaison des délais de développement

Vaccins à ARNm: Les plus rapides (~11 mois de la conception à l'autorisation)

Vaccins à vecteurs viraux: 12-14 mois

Vaccins à virus inactivés: 12-14 mois (selon les régions)

Vaccins à protéines recombinantes: 15-18 mois

Lancement du Projet Vaccinal

Jours 61-120 : Concevoir un vaccin de nouvelle génération

Positionnement dans la chronologie du projet

JOURS 0-60

JOURS 61-120

JOURS 121-180

JOURS 181+

Structure

Conception vaccin

Essais et production

Distribution

Objectifs précis de cette phase

- Concevoir des séquences d'ARNm optimales
- Développer des systèmes de délivrance efficaces
- Établir des protocoles de production évolutifs
- Réaliser des simulations in silico prédictives

Acteurs principaux et approches alternatives

Acteurs principaux:

- Moderna
- BioNTech
- CureVac

Approches alternatives en parallèle:

- Vaccins à vecteurs viraux (Oxford/AstraZeneca)
- Vaccins à protéines recombinantes (Novavax)
- Vaccins à virus inactivés (Sinopharm)

Citation

"Nous sommes en train de réécrire les règles du développement vaccinal. Ce que nous accomplissons en quelques mois prenait auparavant des années. C'est une révolution biotechnologique en action."

— Dr. Stéphane Bancel, PDG de Moderna

Statistique marquante

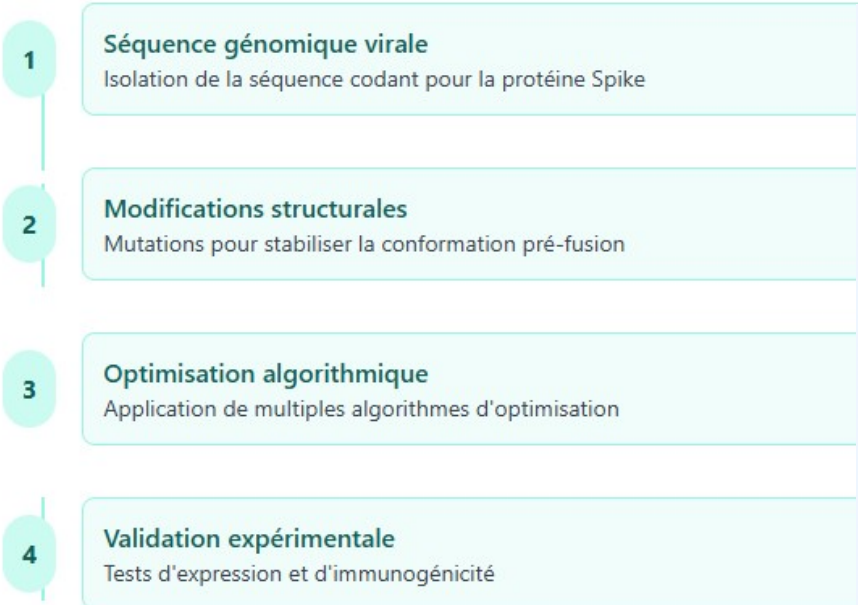
10-15 ans → 1 an

Réduction du temps de développement vaccinal
Vaccins traditionnels vs ARNm

Stratégies de Ciblage et Séquence ARNm

Optimiser la séquence d'ARNm

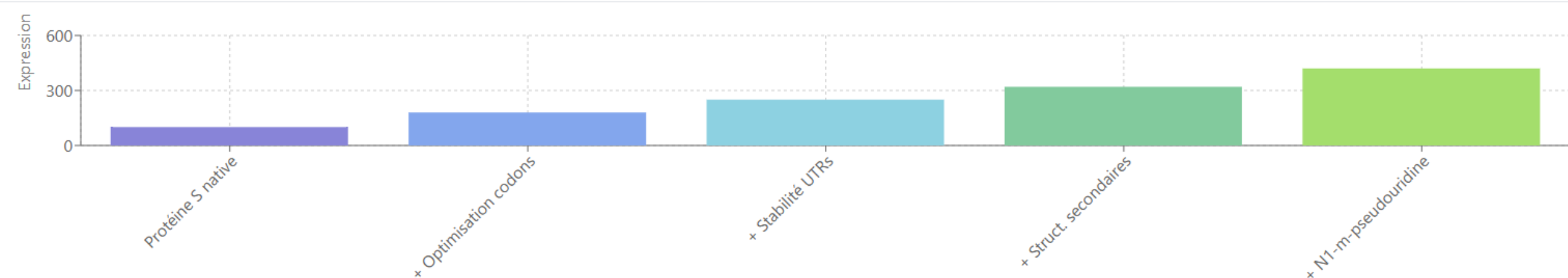
Pipeline d'optimisation de Moderna/BioNTech



Points méthodologiques clés

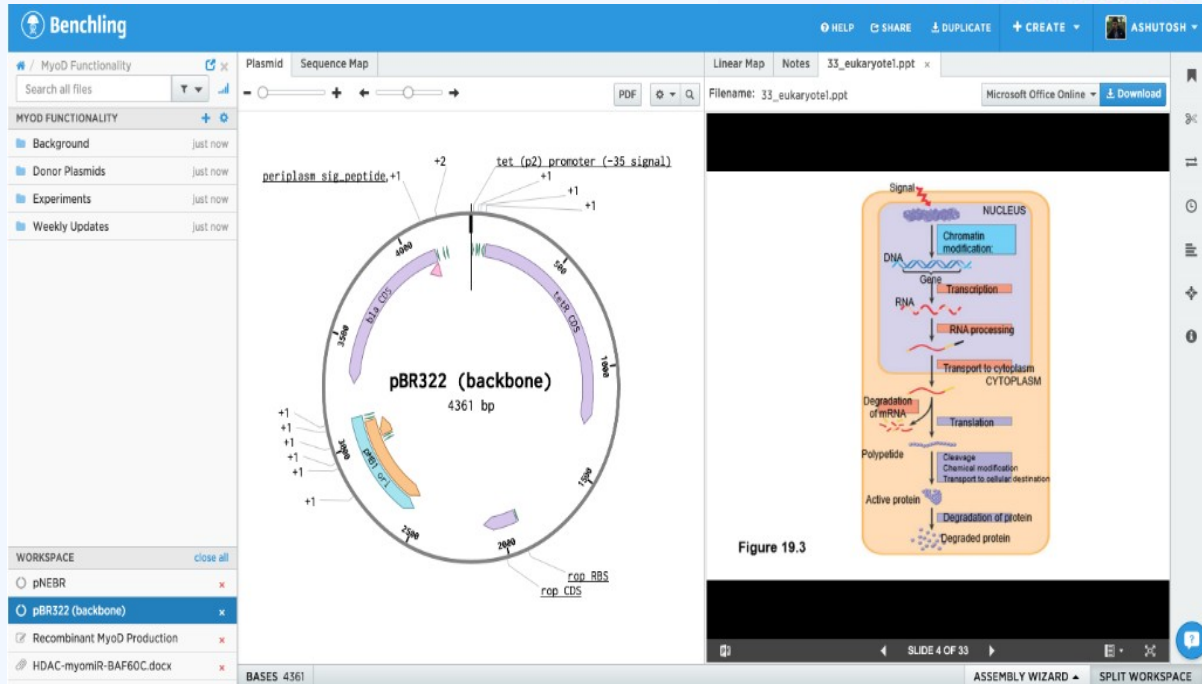
- ✓ **Optimisation des codons:**
Sélection des codons préférentiels pour les cellules humaines, augmentant l'efficacité de traduction de 1.8×
- ✓ **Stabilité de l'ARNm:**
Optimisation des régions 5' et 3' UTR et élimination des séquences déstabilisatrices, prolongeant la demi-vie de 3.2×
- ✓ **Efficacité de traduction:**
Minimisation des structures secondaires limitantes et optimisation du contexte de démarrage Kozak
- ✓ **Modificateurs nucléosidiques:**
Substitution de l'uridine par N1-méthyl-pseudouridine pour échapper à la détection immunitaire et améliorer la stabilité

Gain d'expression prédit : Séquence optimisée vs native



Plateformes de Design de Séquence

Outils de conception in silico



Fonctions principales

1.Édition de séquences avancée

1. Éditeur WYSIWYG de séquences d'ADN, ARN et protéines
2. Annotation automatique et manuelle des caractéristiques
3. Gestion des modifications post-traductionnelles pour les protéines
4. Édition à grande échelle avec recherche/remplacement avancés

2.Conception d'oligonucléotides et de constructions

1. Design automatisé de primers avec vérification des paramètres thermodynamiques
2. Simulation d'assemblage de Gibson, Golden Gate et SLIC
3. Optimisation des codons avec paramètres personnalisables
4. Design CRISPR avec évaluation des off-targets

3.Gestion de données biologiques

1. Organisation hiérarchique des projets et données
2. Système de registre pour plasmides, souches, lignées cellulaires
3. Traçabilité complète avec historique des versions
4. Gestion des échantillons et inventaire de laboratoire

4. Collaboration et partage

1. Collaboration en temps réel sur les mêmes fichiers
2. Contrôle d'accès et permissions granulaires
3. Commentaires et discussions intégrés
4. Intégration avec des outils de gestion de projets

5. Analyse structurale et intégrations

5. Prédiction et visualisation des structures secondaires d'ARN/ADN
6. Visualisation 3D des protéines (via intégration)
7. Alignement de séquences multiples
8. Analyse de sites de restriction et cartes de restriction

6. Automatisation et API

9. API REST complète pour l'automatisation des flux de travail
10. Intégration avec des scripts Python/R pour l'analyse
11. Workflows personnalisables avec conditions et approbations
12. Connexion avec des instruments de laboratoire

Usages spécifiques pour la conception de vaccins ARNm

- Conception et optimisation des séquences d'ARNm avec analyse des structures secondaires
- Modélisation des modificateurs nucléosidiques et leurs effets
- Suivi des versions des constructions pendant le développement
- Collaboration entre équipes multidisciplinaires (bioinformatique, immunologie, production)
- Gestion du pipeline de développement de vaccins de la conception à la production

Points forts : Collaboration cloud, intégration laboratoire, API robuste, Usage privé

Benchling Bioresearch Capabilities



Experiment data capture

Document rich experimental details and assay data. Work together in real time with shared entries, and standardize data collected with templates.



Biomolecule design

Build plasmids and optimize primers with built-in guides. Edit, align, and query sequences for DNA, RNA, AA, and modified oligonucleotides with easy-to-use molecular tools.



Registration

Capture DNA, RNA, AA, and oligos with modifications with experimental metadata in a biologically-aware registry.



Sample management and inventory

Establish an inventory of samples, consumables, and equipment using a hierarchical framework that's linked with the registry and notebook.



Operational insights

Build dashboards and share reports that track metrics and answer questions from all data created within or pulled into Benchling.

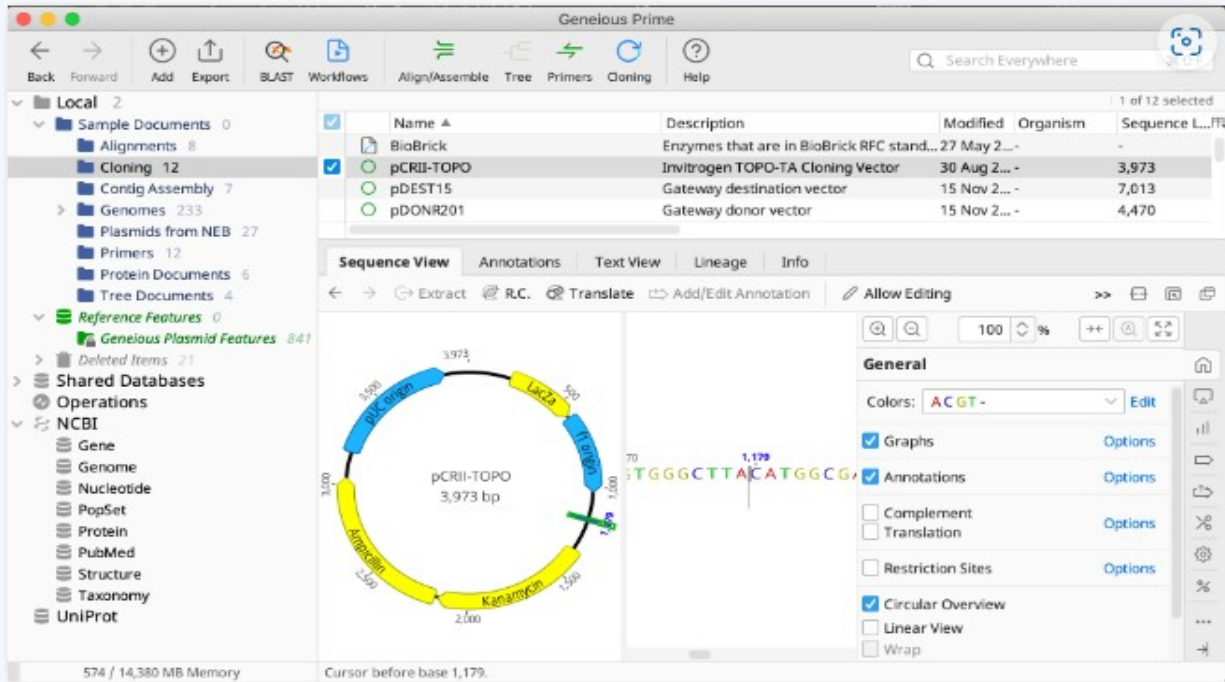


Workflow management

Drag-and-drop tasks in an easy-to-use flowchart to map activities from simple tasks like making a request, through complex process dependencies.

Plateformes de Design de Séquence

Outils de conception in silico



Fonctions principales

1. Analyse et alignement de séquences

1. Alignement de séquences multiples (ClustalW, MUSCLE, MAFFT)
2. Visualisation avancée des alignements avec codage couleur
3. Recherche de motifs et d'homologies
4. Création et édition d'alignements manuels

2. Analyse phylogénétique

1. Construction d'arbres phylogénétiques (NJ, ML, Bayésien)
2. Visualisation et annotation des arbres
3. Analyse de conservation des séquences
4. Calcul de distances génétiques

3. Analyse de séquençage

1. Assemblage de séquences brutes (de novo et par référence)
2. Analyse de chromatogrammes
3. Détection et correction d'erreurs de séquençage
4. Assemblage de génomes

4. Design moléculaire

1. Design de primers et sondes pour PCR/qPCR
2. Planification de clonage et digestive virtuelle
3. Clonage in silico avec vérification des cadres de lecture
4. Design de constructions avec visualisation des jonctions

5. Analyse structurale

5. Prédiction de structures secondaires d'ARN/ADN
6. Analyse des repliements et stabilité (ΔG)
7. Visualisation des épingles à cheveux et autres motifs
8. Détection des problèmes structuraux pour l'expression

6. Gestion de bases de données biologiques

9. Importation directe depuis GenBank, NCBI, ENA
10. Organisation et classification des séquences
11. Recherche avancée dans les collections de séquences
12. Annotation automatique basée sur des références

Usages spécifiques pour la conception de vaccins ARNm

- Analyse comparative des séquences virales pour identifier les cibles vaccinales
- Alignement de variants pour identifier les régions conservées
- Analyse phylogénétique pour suivre l'évolution virale
- Prédiction des structures secondaires d'ARNm critiques
- Conception d'oligonucléotides pour la vérification et l'amplification des constructions

Comparaison et complémentarité constructions

Forces de Benchling

Interface cloud moderne avec collaboration en temps réel

Gestion de données et de workflow intégrée

API robuste et extensibilitéTraçabilité complète et conformité réglementaire

Idéal pour les équipes multidisciplinaires et distribuées

Forces de Geneious

Outils puissants d'alignement et d'analyse phylogénétique

Interface riche pour l'analyse visuelle des séquences

Gestion efficace des données de séquençage

Fonctionnalités plus avancées pour l'analyse comparative

Plus adapté aux analyses approfondies par des bioinformaticiens

Dans le contexte du développement de vaccins ARNm, ces deux plateformes sont souvent utilisées conjointement : Geneious pour l'analyse comparative approfondie et l'exploration des données de séquences, et Benchling pour la conception, l'optimisation et la gestion du pipeline de développement

Prédiction de la Structure d'ARN

RNAfold : prédire pour optimiser

Visualisation : Structures 2D d'ARNm

Structure native



$$\Delta G = -32.5 \text{ kcal/mol}$$

Structures complexes multiples

Structure optimisée



$$\Delta G = -18.2 \text{ kcal/mol}$$

Structure simplifiée, plus acces

Comparaison des repliements:

- Réduction de la complexité structurale
- Minimisation des épingles à cheveux près du site d'initiation
- Élimination des interactions longue distance défavorables
- Amélioration de l'accessibilité pour la machinerie traductionnelle

Points techniques

Principes de prédiction énergétique

- Modèle thermodynamique du plus proche voisin
- Minimisation de l'énergie libre (ΔG)
- Algorithmes de programmation dynamique (Zuker)
- Paramètres énergétiques expérimentaux (Turner)
- Inclusion des pseudo-nœuds et interactions tertiaires

Impact des structures secondaires sur la stabilité

- Protection contre la dégradation par les ribonucléases
- Influence sur la cinétique de traduction
- Masquage/exposition de motifs de reconnaissance
- Effet sur la durée de demi-vie de l'ARNm
- Modulation de l'accessibilité des ribosomes

Élimination des structures indésirables

- Substitutions synonymes ciblées
- Modification du contenu GC local
- Introduction de séquences déstabilisatrices
- Conception de séquences compensatoires
- Optimisation des régions UTR

Applications en biocatalyse

🧠 Optimisation de ribozymes

Conception de ribozymes catalytiques avec activité et spécificité améliorées pour les réactions industrielles comme l'isomérisation, l'hydrolyse et la ligation.

Exemple: Ribozymes hammerhead optimisés pour la dégradation spécifique d'ARN contaminants dans les bioréacteurs

✏️ Engineering d'ARN guides CRISPR

Optimisation des sgRNA pour améliorer la spécificité, l'efficacité et la stabilité dans des applications industrielles comme la modification des voies métaboliques.

Exemple: sgRNA à haute efficacité pour l'édition précise des souches de production d'enzymes industrielles

🛡️ Biosenseurs ARN

Développement de riboswitch et aptamères capables de détecter spécifiquement des métabolites, contaminants ou produits dans les procédés de fermentation.

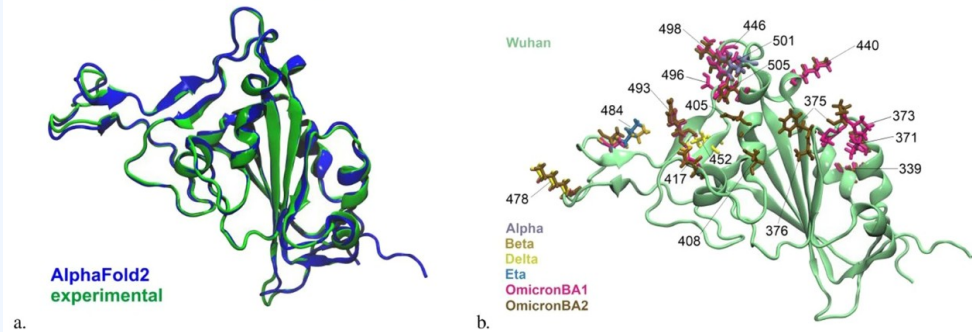
Exemple: Biosenseurs ARN pour la détection en temps réel de produits secondaires indésirables dans la production d'antibiotiques

La maîtrise des structures d'ARN développée pour les vaccins trouve des applications directes dans l'optimisation des biocatalyseurs et systèmes de contrôle moléculaire industriel.

Modélisation 3D de la Spike Protein

AlphaFold et Rosetta : révolution

From: SARS-CoV-2 receptor-binding domain deep mutational AlphaFold2 structures



(a) AF2 aligned Wuhan WT RBD superimposed onto the experimentally determined 6MOJ^{21.22} (RBD-ACE2 complex) clearly shows excellent agreement with respect to local and global structure. The RMSD value is 0.67 Å, which is due to the slight deviation between the structures in "loop" areas such as positions 371 and 478. (b) Variant defining mutations on SARS Cov-19 spike protein RBD. Wuhan RBD is in the cartoon illustration while the variant-defining mutations are illustrated with the licorice drawing method. The residue positions and the color codes are indicated.

Points techniques

Approches par Deep Learning (AlphaFold)

- Architecture à attention croisée traitant les séquences MSA
- Modèles formés sur des structures connues de la PDB
- Prédiction directe de coordonnées 3D avec degrés de confiance
- Temps de calcul: ~heures vs mois par méthodes expérimentales
- Scores pLDDT >90 sur des régions clés de la Spike

Précision vs structures expérimentales

Régions centrales (coeur)

Domaine RBD (liaison au récepteur)

Régions flexibles (boucles)

Points clés:

- Prédiction
- Identification
- Visualisation
- Cartographie

Le RMSD est une mesure mathématique qui calcule la distance moyenne entre les positions des atomes correspondants dans deux structures protéiques après superposition optimale.

Formule mathématique

La formule du RMSD est :

$$\text{RMSD} = \sqrt{(\sum(d^2)/N)}$$

Où :

- d = distance entre chaque paire d'atomes correspondants
- N = nombre total d'atomes comparés
- Σ = somme des distances au carré

Interprétation des valeurs

- < 1Å : Structures pratiquement identiques, différences minimales (précision expérimentale)
- 1-2Å : Très forte similarité structurale, même repliement
- 2-3Å : Similarité importante, même architecture globale
- 3-5Å : Similarité modérée, architecture globale conservée mais différences locales
- > 5Å : Différences structurales significatives

Utilisation dans la modélisation protéique

Dans le contexte de la prédiction structurale comme avec AlphaFold et Rosetta :

- Le RMSD permet de comparer les structures prédites in silico avec les structures déterminées expérimentalement (par cristallographie aux rayons X ou cryo-EM)
- Il peut être calculé sur l'ensemble de la protéine ou sur des sous-régions spécifiques
- Généralement, on utilise les positions des carbones alpha (Cα) pour ce calcul
- Plus le RMSD est faible, plus la prédiction est précise

Importance dans l'analyse de la Spike du SARS-CoV-2

Dans la slide présentée, les valeurs de RMSD indiquent que :

- Les régions centrales (cœur) de la Spike sont prédites avec une très grande précision (RMSD < 1.5Å)
- Le domaine RBD (qui se lie au récepteur ACE2) est également bien prédit (RMSD < 2.0Å)
- Les régions flexibles comme les boucles sont moins précisément prédites (RMSD 3-5Å), ce qui est normal car ces régions sont naturellement mobiles



The structure of the spike protein closely matches the structure determined by Cryo-EM (tan).

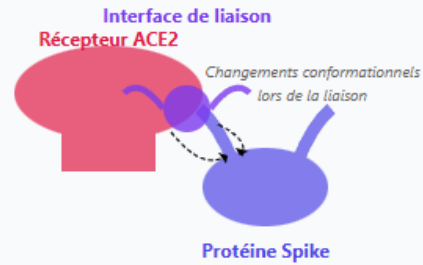
stiques

RMSD < 1.5 Å

RMSD < 2.0 Å

RMSD 3-5 Å

Interaction dynamique spike-ACE2



Observations clés des simulations dynamiques:

- La liaison induit un changement conformationnel dans le RBD
- Exposition sélective des épitopes après liaison
- Coordination des mouvements entre les sous-unités
- Rôle des glycanes dans la stabilisation des interactions

Impact scientifique global d'AlphaFold

AlphaFold a transformé la biologie structurale au-delà du COVID, en rendant accessibles les structures 3D de millions de protéines et en accélérant la recherche dans de nombreux domaines.

Portée

- 214 millions de structures prédites
- Couverture de plus de 1 million d'espèces
- Base de données publique universelle

Applications

- Design de médicaments accéléré
- Ingénierie de protéines industrielles
- Compréhension des maladies génétiques

Avancées méthodologiques

- Intégration IA + biologie structurale
- Nouveaux algorithmes d'apprentissage
- Exploitation des données évolutives

Impact pour COVID-19

- Prédiction précise des variants émergents
- Conception d'anticorps thérapeutiques
- Accélération du développement vaccinal

"AlphaFold représente un progrès extraordinaire... qui change notre compréhension de la biologie"

- Prof. Venki Ramakrishnan, Prix Nobel de Chimie

Simulation de Dynamique Moléculaire

GROMACS et AMBER : simuler pour comprendre

Points techniques

Champs de force utilisés

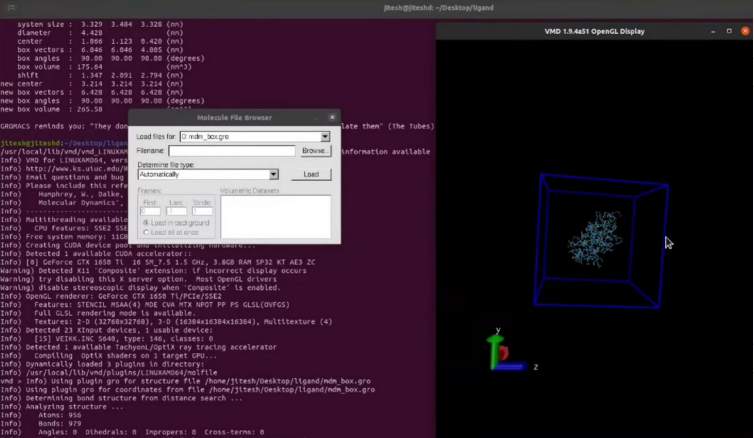
AMBER ff14SB

- Paramétré pour les protéines
- Optimisé pour les régions de boucles
- Modèle d'eau TIP3P

Études comparatives menées pour évaluer la cohérence des résultats avec différents champs de force

CHARMM36m

- Amélioré pour les glycoprotéines
- Paramètres pour modifications PTM
- Modèle d'eau CHARMM-TIP3P



Échelles de temps simulées

Équilibration



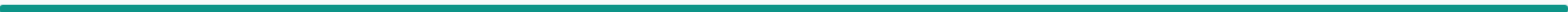
50 ns × 5 réplicas

Production



500 ns × 3 réplicas

Simulations spéciales



1-2 µs ciblés

Pas d'intégration: 2 fs avec contraintes LINCS sur liaisons H

Ressources computationnelles requises

Cœurs CPU/GPU
512-1024 cœurs
128 GPU's NVIDIA A100

Stockage
~15 TB
Trajectoires + analyses

Temps de calcul
~200,000 heures CPU
~5,000 heures GPU

Performance
~50-150 ns/jour
Système de ~1M atomes

Conclusions des simulations:

Les simulations de dynamique moléculaire ont révélé des mécanismes transitoires inaccessibles aux méthodes expérimentales statiques, identifiant des cibles thérapeutiques potentielles et guidant la conception rationnelle des vaccins ARNm en se focalisant sur les conformations les plus immunogènes.

Simulation de Dynamique Moléculaire

GROMACS et AMBER : simuler pour comprendre

Résultats clés



Régions flexibles identifiées

- **RBD (domaine de liaison):**
Fluctuations de 3-5Å, flexibilité accrue en l'absence d'ACE2
- **Boucles S1/S2:**
Haute mobilité (5-8Å RMSF), site de clivage par protéases cellulaires
- **Jonction S2':**
Mouvements corrélés avec l'activation du mécanisme de fusion
- **Domaine NTD:**
Dynamique modulée par les glycosylations, régions d'échappement immunitaire



Points faibles structuraux ciblables

- **Hinges mécaniques:**
Régions charnières avec mouvements amplifiés, sensibles aux anticorps bloquants
- **Site de préactivation:**
Conformation métastable sensible aux perturbations, stabilisable par modifications
- **Poches cryptiques:**
Cavités transitoires apparaissant dans 5-8% des frames, cibles potentielles
- **Interfaces dynamiques:**
Zones de déstabilisation préférentielle lors des transitions conformationnelles



Sites d'interaction caractérisés

- **Interface RBD-ACE2:**
Réseau de 10-15 interactions électrostatiques et H-bonds clés
- **Poche hydrophobe:**
Cavité transitoire (15-25% du temps) pour fixation potentielle d'inhibiteurs
- **Interfaces entre protomères:**
Dynamiques collectives stabilisant la structure trimérique
- **Sites de reconnaissance d'anticorps:**
Cartographie des épitopes accessibles à différentes conformations

Approches de Stabilisation

Ingénierie de la spike pour la stabilisation

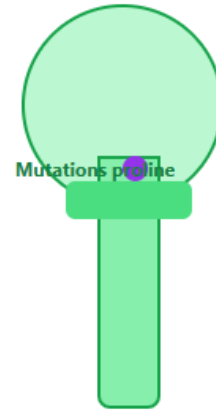
Structure avant/après mutations proline

Structure Native



Conformations multiples possibles

Structure avec mutations K986P et V987P

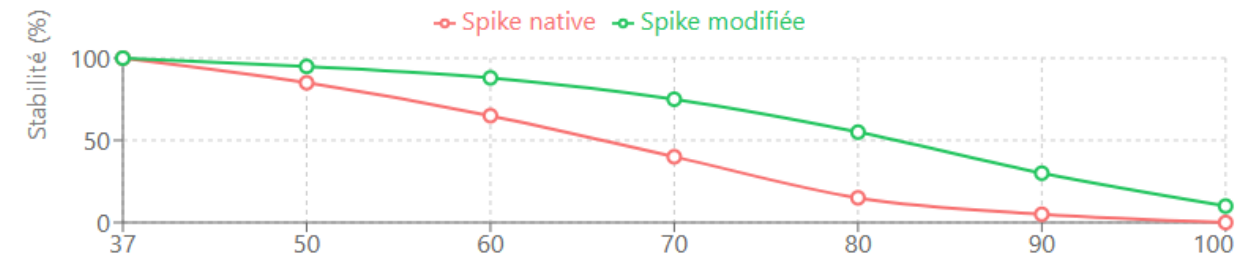


Conformation préfusion verrouillée

Détails techniques

- **Mutations K986P et V987P** : Substitution de la lysine (K) et de la valine (V) par des prolines (P)
- **Verrouillage en conformation préfusion** : La rigidité de la proline empêche le changement conformationnel
- **Amélioration de l'immunogénicité** : Exposition prolongée des épitopes neutralisants

Stabilité thermique comparée



Principes d'ingénierie rationnelle des protéines

Analyse structurale

Modélisation moléculaire

Design itératif

Caractérisation biophysique

Validation fonctionnelle

ARNm vs Vecteur viral : deux philosophies

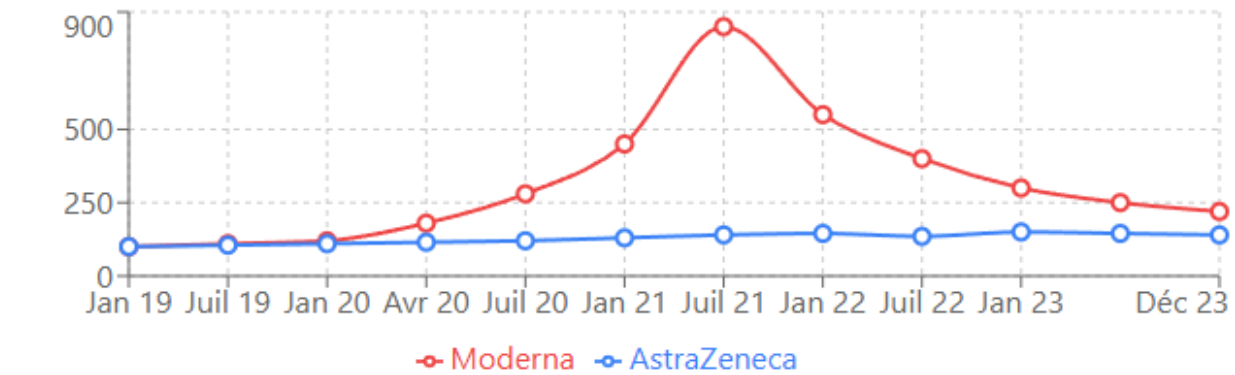
Étude Comparative Moderna vs AstraZeneca

Tableau Comparatif Détaillé

Critère	Moderna (ARNm)	AstraZeneca (Vecteur viral)
Technologie	ARN messenger encapsulé dans des nanoparticules lipidiques	Adénovirus de chimpanzé (ChAdOx1) modifié génétiquement
Approche bioinformatique	Intensive (modélisation 3D, design des séquences optimisé)	Modérée (sélection du vecteur, intégration de la séquence)
Stabilité et conservation	-70°C à -20°C (chaîne du froid stricte requise)	2-8°C (réfrigérateur standard)
Capacité de production	Scaling rapide mais infrastructure spécialisée	Production sur lignées cellulaires plus conventionnelle
Coûts de développement	Élevés (>\$1 milliard, R&D intensive)	Modérés (technologie plus mature)
Adaptabilité aux variants	Très élevée (modification rapide de la séquence ARN)	Moyenne (reprogrammation plus complexe du vecteur)

Impact sur Valorisations Boursières (2019-2023)

Indice de performance (base 100 - janvier 2019)



Analyse SWOT Complète

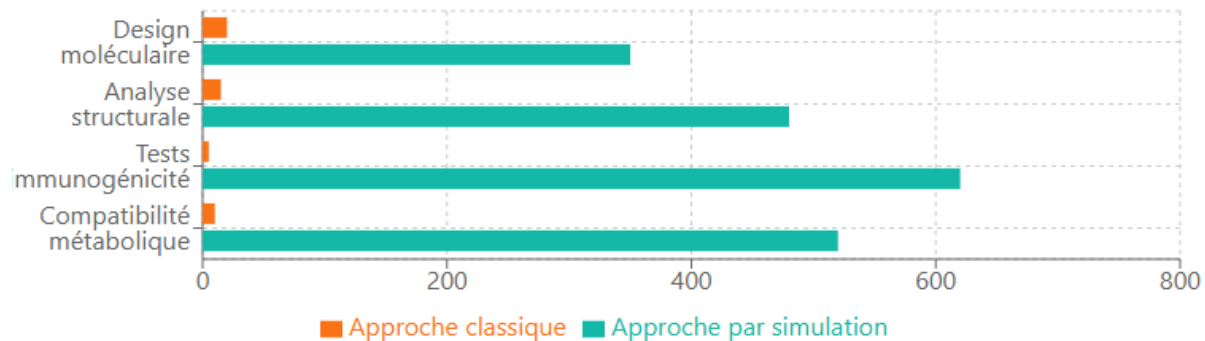
Moderna (ARNm)	AstraZeneca (Vecteur)
Forces	Forces
Faiblesses	Faiblesses
Opportunités	Opportunités
Menaces	Menaces

Réduction drastique des délais

Impact des Simulations sur le Développement des Vaccins

Ressources informatiques mobilisées

Milliers d'heures CPU/GPU par phase de développement



L'approche par simulation nécessite un investissement massif en puissance de calcul, mais permet des économies substantielles en temps et en ressources humaines

Statistiques d'impact

>90%

Réduction du temps de design moléculaire

De plusieurs années à quelques semaines

\$600M+

Économies estimées par vaccin

Réduction des coûts d'échec et d'expérimentation

75%

Réduction des tests animaux

Prédiction in silico des résultats d'innocuité

8x

Augmentation des candidats viables

Plus grande diversité d'approches explorées

Évolution des aspects réglementaires

1

Acceptation progressive des données in silico

La FDA et l'EMA ont développé des cadres d'acceptation pour les données issues de simulations, notamment pour les phases préliminaires

2

Complémentarité avec les tests in vivo

Approche hybride recommandée: les simulations permettent de cibler et réduire les tests biologiques requis

3

Nouveaux standards en développement

Guidelines FDA/EMA 2023 définissant les critères de validation pour les modèles prédictifs et les simulations moléculaires

"Les simulations in silico constituent désormais un élément fondamental de l'arsenal d'évaluation, permettant d'accélérer significativement le développement tout en maintenant les standards de sécurité les plus élevés."

— Rapport 2023 du groupe de travail FDA sur l'innovation biopharmaceutique

Développement classique vs accéléré

A

Développement Classique

Approbation : 6-12 Mois

Essais cliniques : 5 à 8 ans

Développement global 6 à 10 ans

Coûts : 2 à 2.5 Milliards USD

B

Développement Accéléré

Approbation : 2 Mois

Essais cliniques : 10 mois

Développement global 1+ an

Coûts : 1.5 Milliards USD

↓ **Réduction de 70-80% du temps de développement**

J+120 : Prêts pour les essais

Transition vers les phases précliniques et cliniques

Récapitulatif des acquis majeurs



Design optimisé par simulation

Structure 3D stabilisée et efficacité immunogénique prédite à >85%



Formulation nanoparticulaire validée

Stabilité confirmée, biodistribution modélisée et échantillons GMP produits



Protocole de production établi

Rendement optimisé et contrôle qualité automatisé en place



Tests in vitro préliminaires réussis

Expression confirmée et immunogénicité démontrée sur modèles cellulaires

Portefeuille de propriété intellectuelle

Brevets déposés (12)

Séquence ARNm

4 brevets, couverture mondiale

Formulation lipidique

3 brevets, composition propriétaire

Procédés de production

3 brevets, optimisation d'échelle

Méthodes analytiques

2 brevets, contrôle qualité

Stratégies de protection ARNm

- > Protection en couches avec brevets principaux et divisions stratégiques
- > Secrets commerciaux pour certains aspects du procédé de fabrication
- > Accélération des examens obtenue pour 8 brevets (voie prioritaire)

Implications futures

Portefeuille structuré pour permettre des applications thérapeutiques élargies au-delà des vaccins, notamment en oncologie et maladies génétiques, avec possibilité de licences croisées stratégiques.

Transition vers les essais

Précliniques (J+130)

- Toxicologie réglementaire
- Modèles animaux humanisés
- Études de biodistribution
- Données ADME complètes



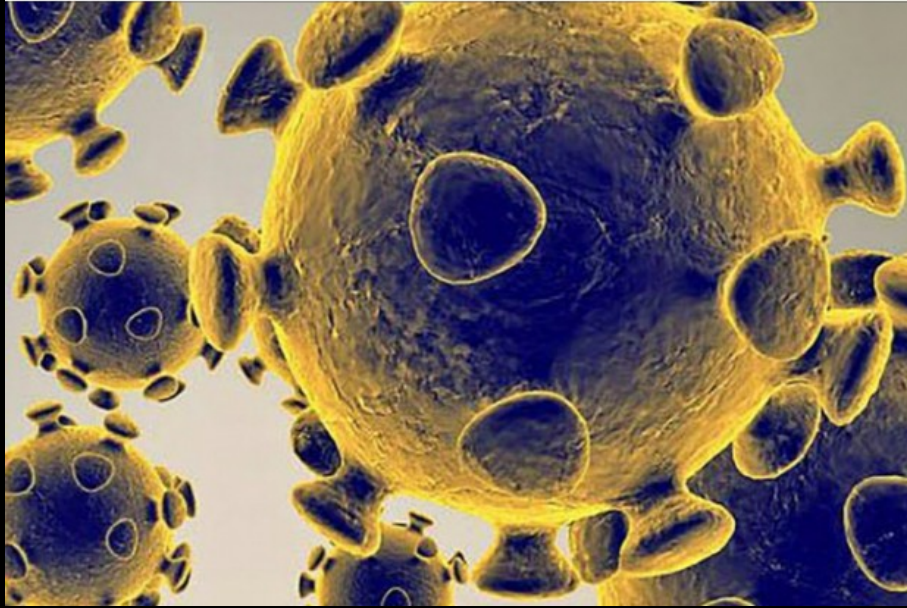
Cliniques (J+210)

- Phase I : protocole accéléré
- Cohortes enrichies par biomarqueurs
- Monitoring en temps réel
- Design adaptatif validé par les autorités

Réduction estimée de 40% du temps de transition grâce aux données in silico acceptées par les autorités

Petit « bémol » sur l'utilisation de données « In Silico » ou encore données synthétiques, or « urgences » type pandémie mondiale, les régulateurs sont beaucoup plus circonspects vis-à-vis de ce type de données et exigent une preuve en « double aveugle », données synthétiques VS données réelles.

La course mondiale pour établir un vaccin contre le nouveau coronavirus bat son plein. Un essai clinique a débuté lundi 16 mars 2020 aux Etats-Unis.



J121 à J200 : TESTS PRECLINIQUES et LANCEMENT DES ESSAIS CLINIQUES

16 DECEMBRE 2019 : PREMIER CAS OFFICIEL DE COVID-19

16 JANVIER 2020 : SEQUENCAGE DE L'ARN DU VIRUS SARS-CoV2

16 Mars 2020 : DEBUT DES ESSAIS CLINIQUES DU VACCIN A ARNm DE MODERNA



Résumé des Phases d'Essais Cliniques

Phase	Raison / Fonction	Typologie de Patients	Nombre de Participants	Durée Typique	Coûts Estimés (USD)	Particularités multicentriques
Phase 0 (exploratoire)	Microdosage pour valider biodisponibilité / pharmacocinétique rapide sans visée thérapeutique	Volontaires sains	10-15	Quelques jours à semaines	1-5 M\$	Rarement multicentrique
Phase I	Évaluer la sécurité, tolérance, profil pharmacocinétique et pharmacodynamique	Principalement volontaires sains (sauf molécules toxiques comme anticancéreux)	20-100	6 mois - 1 an	1-30 M\$	Parfois multicentrique pour accélérer recrutement
Phase II	Évaluer l'efficacité préliminaire, continuer l'évaluation de la sécurité, déterminer la dose optimale	Patients malades ciblés	100-300	1 à 2 ans	10-100 M\$	Multicentrique fréquent pour augmenter la diversité de patients
Phase III	Confirmer efficacité, surveiller effets secondaires à grande échelle, comparer au standard of care, dossier réglementaire	Patients malades (critères stricts)	300-3000+	2 à 5 ans	100-500 M\$ (voire >1 Md\$ pour pathologies complexes)	Souvent multicentrique et multinationale pour robustesse et représentativité
Phase IV (post-AMM, pharmacovigilance)	Surveillance continue du médicament sur le marché, détection d'effets secondaires rares, nouvelles indications	Patients traités en vie réelle	Plusieurs milliers à millions	Illimité dans le temps	Variable (souvent intégré au budget marketing/R&D)	Fortement multicentrique, données issues de registres, vie réelle

Jours 121-200 : La phase critique des essais

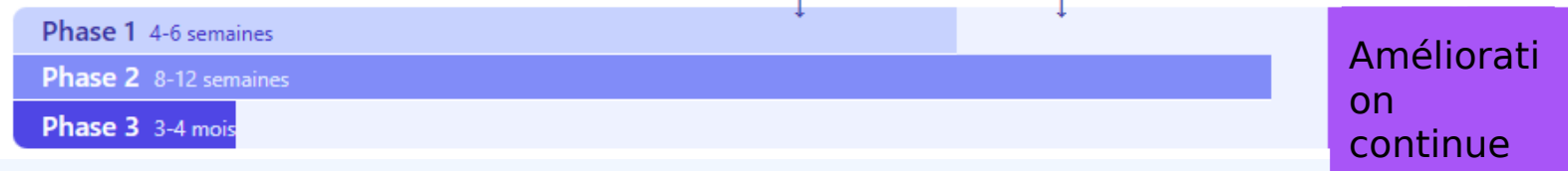
Segment 4 – Tests précliniques et lancement des essais cliniques

Processus accéléré vs traditionnel

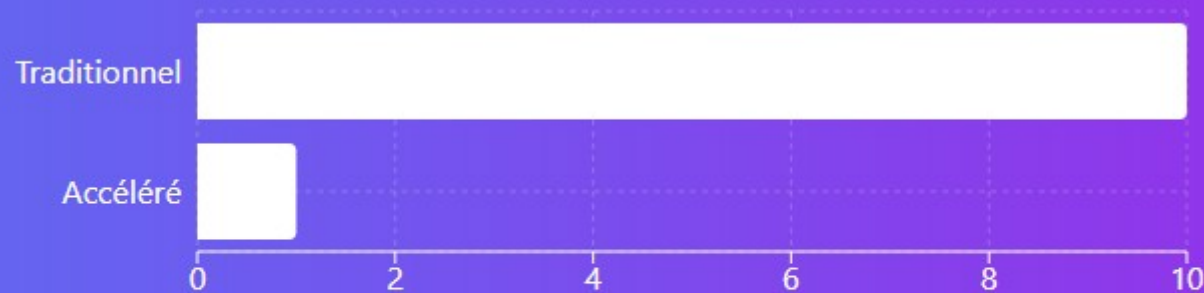
● Processus Traditionnel (séquentiel)



● Processus Accéléré (parallélisé)



Réduction drastique des délais



10+ ans → <1 an

Réduction du temps de développement grâce aux procédures d'urgence et à la parallélisation

Cadre réglementaire spécial

🛡️ Procédures d'urgence FDA (EUA)

- Autorisation d'utilisation d'urgence
- Évaluation accélérée des données intermédiaires
- Réunions fréquentes (toutes les 2 semaines)
- Révision en temps réel des données

🛡️ Framework EMA accéléré

- Procédure d'évaluation accélérée (150 jours)
- Conditional Marketing Authorization (CMA)
- Scientific Advice accéléré
- PRIME scheme pour les thérapies innovantes

⚡ Parallélisation innovante

- Chevauchement des phases 1 et 2
- Production à risque en parallèle
- Évaluation continue des données
- Partage global des données en temps réel

Chronologie des Essais Cliniques des Vaccins ARNm

Accélération et superposition des phases pour les vaccins COVID-19

Chronologie du développement clinique accéléré des vaccins ARNm

Janv-Fév
2020



Études précliniques

- Séquençage du SRAS-CoV-2 (janvier 2020)
- Conception de la séquence ARNm de la protéine Spike
- Tests sur modèles cellulaires et animaux
- Optimisation des nanoparticules lipidiques (LNP)

Mars-Avril
2020

1

Phase 1

45-120 participants

- Première administration à l'humain (16 mars 2020 pour Moderna)
- Tests de dosage (10µg, 30µg, 100µg)
- Évaluation de la sécurité primaire
- Dosage des anticorps (immunogénicité)
- Durée accélérée: 4-6 semaines (vs. 1-2 ans traditionnellement)

Mai-Juin
2020

1/2

Phase 1/2 Combinée

200-600 participants

- Extension des cohortes d'âge (adultes âgés)
- Optimisation du dosage
- Études d'immunogénicité approfondie
- Évaluation continue des données de sécurité
- Préparation simultanée des sites de Phase 3

Juil-Oct
2020

2/3

Phase 2/3 Combinée

30 000-44 000 participants

- Lancement massif des essais (27 juillet 2020)
- Design à grande échelle randomisé, en double aveugle
- Critères d'efficacité primaire: prévention de la COVID-19
- Critères secondaires: prévention des cas sévères
- Stratification par âge et facteurs de risque
- Analyse intermédiaire planifiée (après 150-170 cas)

Nov 2020



Analyse intermédiaire

- Première analyse d'efficacité (9 novembre 2020 pour Pfizer)
- Efficacité de ~95% annoncée pour Pfizer et Moderna
- Préparation des dossiers réglementaires en parallèle
- Communication immédiate des résultats aux autorités

Déc 2020



Autorisation d'urgence

- Réunions des comités consultatifs (FDA: 10 et 17 décembre 2020)
- Autorisation d'urgence FDA (11 déc. pour Pfizer, 18 déc. pour Moderna)
- Autorisation conditionnelle EMA (21 déc. pour Pfizer, 6 janv. pour Moderna)
- Début de la vaccination massive
- Durée totale: ~11 mois depuis le séquençage (vs 10+ ans traditionnellement)

2021-
présent

4

Phase 4 (Surveillance post-commercialisation)

- Surveillance active de la sécurité
- Études d'efficacité en vie réelle
- Suivi des effets secondaires rares
- Études sur des populations spécifiques (enfants, femmes enceintes)
- Adaptation aux variants
- Études de durabilité de l'immunité et doses de rappel

Innovations clés dans le processus

- > Chevauchement des phases au lieu de séquençement strict
- > Production à risque en parallèle des essais cliniques
- > Rolling submissions et revue en temps réel par les autorités
- > Analyses intermédiaires planifiées pour accélérer les conclusions

Impact sur le futur développement vaccinal

La chronologie accélérée des vaccins ARNm contre la COVID-19 représente une transformation radicale du paradigme de développement pharmaceutique - non seulement par sa rapidité exceptionnelle, mais par la démonstration que sécurité et efficacité peuvent être établies dans ce cadre.

"Les vaccins ARNm ont redéfini ce qui est possible en termes de chronologie de développement, créant un nouveau modèle qui influencera l'industrie pharmaceutique pour les décennies à venir."

— Dr. Anthony Fauci

EDC : la révolution de la donnée clinique

Transformation numérique des essais cliniques

Interfaces des principales plateformes EDC

Medidata Rave

● ● ● Medidata Rave EDC

Patient ID

Subject_123456

Visit

Screening

Vital Signs

Demographics

Oracle Clinical

Oracle Clinical EDC

Patient Information

Clinical Data

Query Management

REDCap

REDCap Data Collection

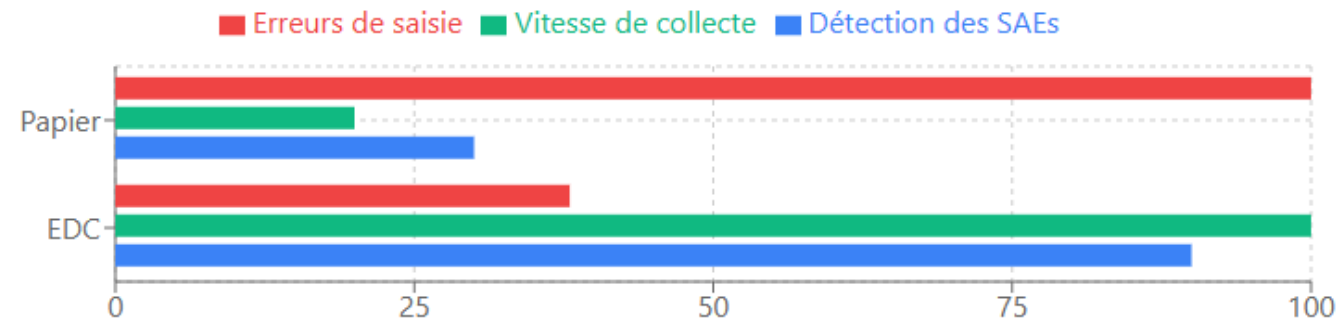
Temperature (°C)

37.2

Blood Pressure

120/80

Avantages quantifiés



-62%
Réduction des erreurs de saisie

5x
Accélération de la collecte

24h
Détection précoce des SAEs

(*) Serious Adverse Events (SAEs)

Architecture technique



- Collecte électronique décentralisée
=> Saisie directe sur sites ou via appareils mobiles/tablettes

- Validation en temps réel
=> Détection immédiate des erreurs logiques et valeurs aberrantes

- Sécurisation et conformité HIPAA/HDS - GDPR



Rapports
en temps réel

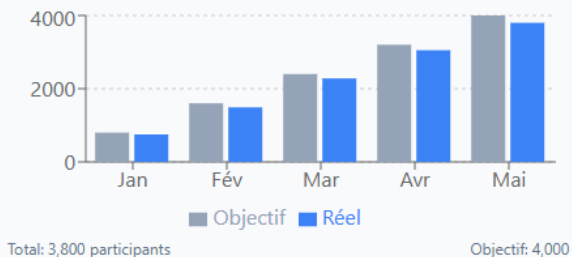
Dashboards : piloter l'essai clinique

Collecte et Analyse en Temps Réel

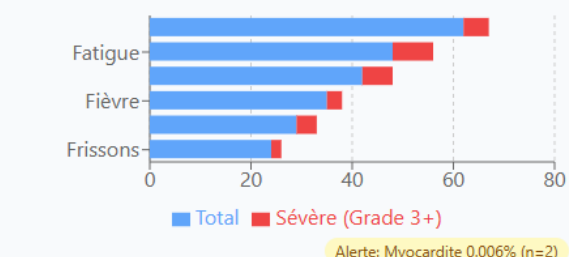
Dashboard Monitoring - Essai Clinique VAX-2023

Mise à jour: Temps Réel

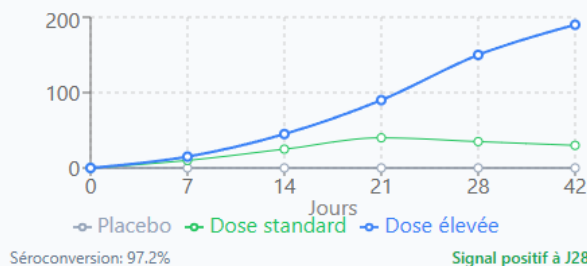
Recrutement en temps réel



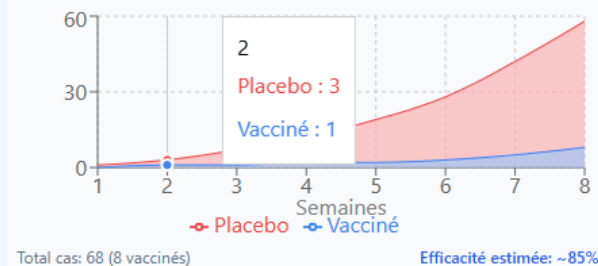
Effets secondaires (%)



Réponses immunitaires (GMC)



Tendances d'efficacité (cas cumulés)



Métriques critiques à surveiller

Recrutement

- Taux de recrutement hebdomadaire
- Diversité démographique
- Taux de rétention (dropout rate)
- Power statistique actualisé

Sécurité

- AEs vs SAEs par groupe
- Signaux d'alerte précoce
- Distribution temporelle post-dose
- Seuil d'arrêt prédéfini (stopping rules)

Immunogénicité

- Titre d'anticorps neutralisants
- Réponse cellulaire T
- Cinétique (pic, déclin)
- Comparaison aux convalescents

Efficacité

- Cas incidents dans chaque bras
- Divergence des courbes de Kaplan-Meier
- Efficacité par sous-groupe
- Efficacité contre variants émergents

Note: Le tableau de bord doit fournir des alertes automatisées lorsque les métriques dépassent les seuils prédéfinis ou montrent des tendances préoccupantes.

Le diagramme "Tendances d'efficacité" présente l'accumulation des cas de maladie au fil du temps (sur 8 semaines) dans les deux groupes de l'essai clinique. C'est un élément crucial du dashboard car il permet de visualiser la divergence progressive entre les courbes du groupe placebo et du groupe vacciné.

Voici ce qu'on peut y observer :

- 1. Courbe du groupe placebo (rouge) :** On voit une augmentation exponentielle des cas, commençant lentement (1 cas en semaine 1) puis s'accéléralant rapidement jusqu'à atteindre 58 cas cumulés en semaine 8.
- 2. Courbe du groupe vacciné (bleue) :** Cette courbe montre une progression beaucoup plus lente et plus linéaire, avec seulement 8 cas cumulés à la semaine 8.
- 3. Point de divergence :** Les courbes commencent à se séparer nettement à partir de la semaine 3-4, ce qui correspond généralement au délai nécessaire pour développer une réponse immunitaire après la vaccination.
- 4. Interprétation :** L'écart grandissant entre les deux courbes est le signal visuel de l'efficacité du vaccin. Plus les courbes s'éloignent l'une de l'autre, plus le vaccin montre un effet protecteur.
- 5. Projection :** Si on prolongeait ces courbes, on s'attendrait à ce que l'écart continue de s'agrandir, renforçant les preuves d'efficacité.

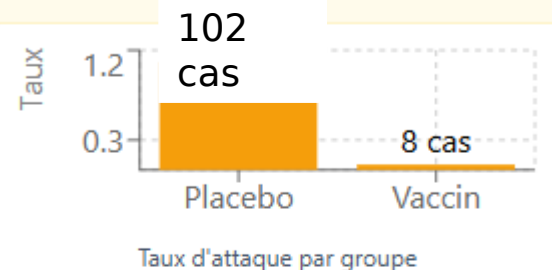
Ce type de visualisation en temps réel est particulièrement précieux pour les comités de surveillance qui peuvent ainsi détecter rapidement si un vaccin montre des signes d'efficacité ou non, parfois bien avant que l'essai n'atteigne son point d'analyse finale prédéfini.

Dashboards : piloter l'essai clinique

Collecte et Analyse en Temps Réel

Mini-atelier : Interprétation de Dashboards

1. Identification d'un signal d'efficacité

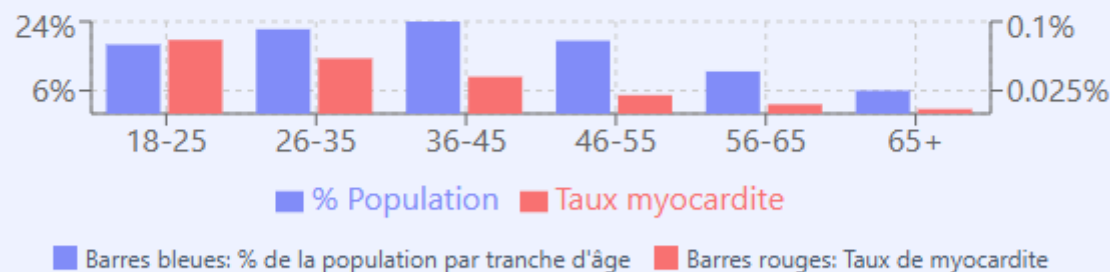


Questions à discuter:

- Quel est le taux d'attaque dans chaque groupe?
- Calculer l'efficacité vaccinale préliminaire
- Est-ce statistiquement significatif?

Note: $\text{Efficacité} = (1 - \text{RR}) \times 100\% = 95\%$

2. Détection de patterns d'effets secondaires



Points à identifier:

- Y a-t-il une corrélation avec l'âge?
- Le signal est-il plus fort dans certains sous-groupes?
- Quelles mesures préventives pourraient être mises en place?

Le taux d'attaque est un terme épidémiologique qui désigne la proportion de personnes dans un groupe défini qui développent la maladie ou l'infection étudiée pendant une période spécifique.

Dans le contexte de l'essai clinique d'un vaccin, le taux d'attaque représente :

- La proportion de participants qui ont contracté la maladie (ici, probablement COVID-19)
- Calculée comme le nombre de cas divisé par le nombre total de participants dans chaque groupe

Par exemple, dans notre graphique :

- Pour le groupe placebo : 162 cas sur 15 000 participants = 1,08%
- Pour le groupe vacciné : 8 cas sur 15 000 participants = 0,053%

C'est en comparant ces deux taux d'attaque qu'on peut calculer l'efficacité du vaccin :

- $\text{Efficacité} = (1 - (\text{taux d'attaque du groupe vacciné} / \text{taux d'attaque du groupe placebo})) \times 100\%$
- Dans ce cas : $(1 - (0,053\% / 1,08\%)) \times 100\% = 95\%$

Cette mesure est essentielle car elle permet d'évaluer directement l'impact protecteur du vaccin en conditions réelles d'exposition au pathogène, en tenant compte des tailles égales des groupes. Le diagramme à barres montre clairement la réduction significative du risque d'infection dans le groupe ayant reçu le vaccin.

Méthodes bayésiennes : flexibilité et puissance

Approches adaptatives pour les essais cliniques et le développement de procédés

Principe d'adaptation des protocoles

Les méthodes bayésiennes permettent d'ajuster dynamiquement les protocoles expérimentaux en fonction des données collectées en temps réel, améliorant la réactivité et la pertinence des décisions.

Applications concrètes

- Ajustements des doses
- Modification des critères d'inclusion
- Redimensionnement des cohortes

Applications industrielles

- Optimisation de bioprocédés
- Design of Experiments (DoE) adaptatif
- Contrôle qualité prédictif

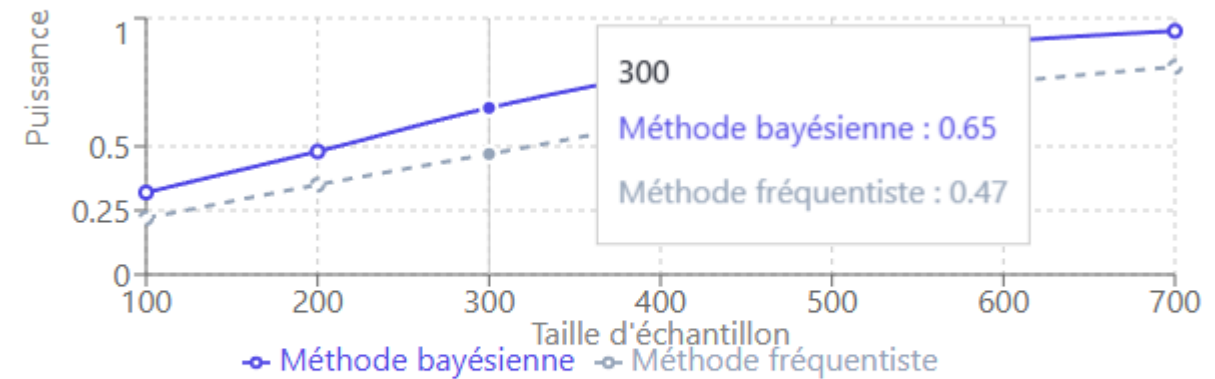
- L'approche **fréquentiste** est adaptée aux **tests stricts sur des échantillons massifs** sans besoin d'intégrer des croyances préalables.
- L'approche **bayésienne** est **particulièrement précieuse en économie expérimentale** car elle offre :
 - Plus de souplesse.
 - Une capacité d'adaptation dynamique.
 - Une meilleure intégration des connaissances a priori.
 - Une meilleure robustesse face à des échantillons limités et des comportements incertains.

L'approche bayésienne, tout en étant **extrêmement puissante**, n'est pas exempte de critiques philosophiques et pratiques.

Les critiques traditionnelles sont :

- **Dépendance au prior** : l'inférence peut être extrêmement sensible au choix subjectif du prior, surtout avec peu de données.
- **Modélisation forcée** : parfois, pour appliquer Bayes, il faut forcer une structuration probabiliste même quand l'incertitude est mal définie.
- **Complexité computationnelle** : la mise en œuvre numérique peut être très lourde (ex : MCMC, approximations variationnelles).

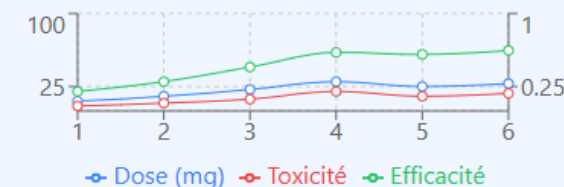
Simulation bayésienne vs approche fréquentiste



Gains de la méthode bayésienne:

- Puissance statistique supérieure à taille d'échantillon égale
- Réduction possible de ~35% de la taille d'échantillon
- Prise de décision plus précoce (analyse intermédiaire)

1 Ajustement des doses



Optimisation de la dose par modèle CRM (Continual Reassessment Method) bayésien, maximisant l'efficacité sous contrainte de toxicité.

2 Modification des critères d'inclusion

Phase initiale:

- Age: 18-65 ans
- Sans comorbidités
- Biomarqueur: tous

Après adaptation:

- Age: 18-75 ans
- Diabète contrôlé admis
- Biomarqueur: positif uniquement

Ajustement adaptatif des critères d'inclusion basé sur les distributions de probabilité a posteriori des réponses par sous-groupe.

Gain en puissance statistique: +27%

Bases statistiques des essais cliniques

L'hypothèse nulle (souvent notée H_0) est un concept fondamental en statistique inférentielle.

Formulation des hypothèses :

- Hypothèse nulle (H_0) : Postule qu'il n'y a pas d'effet ou de différence (ex : "le traitement n'a pas d'effet")
- Hypothèse alternative (H_1) : Postule qu'il existe un effet ou une différence (ex : "le traitement a un effet")

Logique du test : On part du principe que H_0 est vraie jusqu'à preuve du contraire

On recueille des données et on calcule une statistique de test. On détermine la probabilité (p-value) d'obtenir un résultat au moins aussi extrême si H_0 est vraie

Si cette probabilité est inférieure au seuil alpha (généralement 0,05), on rejette H_0

Décisions possibles :

- Rejeter H_0 : On conclut qu'il y a probablement un effet (résultat "statistiquement significatif")
- Ne pas rejeter H_0 : On ne peut pas conclure qu'il y a un effet (mais cela ne prouve pas l'absence d'effet)

Erreurs possibles :

- Erreur de type I (réglée par alpha) : Rejeter H_0 alors qu'elle est vraie (faux positif)
- Erreur de type II (liée à beta) : Ne pas rejeter H_0 alors qu'elle est fausse (faux négatif)
- Puissance (1-beta) : Probabilité de rejeter H_0 quand elle est effectivement fausse
- Limitations : Ne démontre jamais définitivement l'hypothèse nulle. La signification statistique n'implique pas nécessairement une pertinence clinique. Fortement influencée par la taille de l'échantillon. Dans les essais cliniques, ce cadre est utilisé pour évaluer si un nouveau traitement montre un effet suffisamment important pour être considéré comme non attribuable au hasard.

La puissance statistique représente la probabilité qu'un test statistique détecte un effet réel lorsqu'il existe.

En d'autres termes, c'est la capacité d'un test à rejeter correctement l'hypothèse nulle quand elle est fausse.

Plus concrètement, dans le contexte d'un essai clinique de vaccin par exemple : Si le vaccin a réellement un effet protecteur

La puissance statistique est la probabilité que l'étude conclue correctement à cet effet

La puissance statistique dépend de plusieurs facteurs :

- La taille de l'échantillon (nombre de participants)
- L'ampleur de l'effet à détecter
- La variabilité des données
- Le seuil de significativité choisi (généralement $\alpha = 0,05$) (**aussi considéré comme la probabilité d'un faux positif**)
- Une puissance de 80% ($\beta = 0,20$) est souvent considérée comme standard dans les essais cliniques, ce qui signifie que l'étude a 80% de chances de détecter l'effet recherché s'il existe réellement (**aussi considéré comme la probabilité d'un faux négatif**)
- Dans le graphique de la diapositive, on voit que l'approche bayésienne permet d'atteindre une puissance statistique plus élevée avec le même nombre de participants que l'approche fréquentiste classique, ce qui constitue un avantage majeur en termes d'efficacité des études.

Statistiques VS Probabilités

La Loi des Grands Nombres

Formellement, pour des variables aléatoires indépendantes et identiquement distribuées $X_1, X_2, \dots, X_n$ avec $E[X_i] = \mu$:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mu \quad (\text{presque sûrement})$$

Aspect	Probabilité	Statistique
Objet d'étude	Modéliser un phénomène incertain avant l'observation.	Déduire les caractéristiques du modèle inconnu à partir de données observées.
Sens de l'inférence	Déductif : part d'un modèle pour prédire des observations.	Inférentiel/Inductif : part d'observations pour estimer le modèle.
Exemple type	Connaître p d'une pièce biaisée, prédire que pile sort avec proba p .	Observer 100 lancers et estimer p à partir de la fréquence de piles.
Type de raisonnement	Théorique, a priori.	Empirique, a posteriori.

Probabilités : Que se passerait-il si l'on connaissait parfaitement les lois du hasard ?
Statistiques : Que pouvons-nous apprendre sur les lois du hasard en observant des résultats partiels et bruités ?

Dimension	Approche fréquentiste	Approche bayésienne
Définition de la probabilité	Limite d'une fréquence relative dans une infinité d'expériences répétées.	Degré subjectif de croyance, mis à jour avec l'information disponible.
Vision du paramètre (ex : moyenne, proportion)	Le paramètre est fixe mais inconnu.	Le paramètre est aléatoire : on en exprime une incertitude via une distribution.
Vision de l'échantillon	L'échantillon est aléatoire ; le paramètre est fixe.	L'échantillon est observé ; la distribution du paramètre est mise à jour.
Exemples typiques	p-value, tests d'hypothèses, intervalles de confiance.	Posterior distributions, Bayes Factors, credible intervals.

Statistiques VS Probabilités

En estimation

- Fréquentiste** : Estime un paramètre par une statistique (ex : moyenne), construit un **intervalle de confiance**.
L'intervalle contient la vraie valeur **dans X% des cas** si l'expérience était répétée.
- Bayésien** : Estime une **distribution a posteriori** du paramètre, construit un **intervalle de crédibilité**.
L'intervalle donne **directement la probabilité** que le paramètre soit dans l'intervalle, **étant donné les données**.

En test d'hypothèse

- Fréquentiste** : Rejette ou non une hypothèse nulle via une **p-value**.
- Bayésien** : Compare la probabilité des hypothèses via un **Bayes Factor** ou directement la distribution postérieure.

Exemple : Supposons que l'on teste l'efficacité d'un nouveau médicament.

- Fréquentiste** :
 - Hypothèse nulle : "Pas d'effet".
 - Si $p\text{-value} < 0.05$, on rejette H_0 .
 - Sinon, on ne peut pas conclure (on n'affirme jamais H_0 vraie).
- Bayésien** :
 - Modèle la probabilité que l'effet soit positif étant donné les données.
 - Peut conclure par exemple : "Il y a 95% de probabilité que l'effet du médicament soit supérieur à zéro."

Critères	Fréquentistes	Bayésiens
Neutralité a priori	Pas besoin de prior : estimation "objective".	Nécessite un prior : possible subjectivité, mais intégration de connaissance antérieure.
Simplicité computationnelle	En général plus simple, plus rapide pour tests simples.	Plus lourd (nécessite souvent MCMC, simulations), surtout pour modèles complexes.
Flexibilité	Protocoles rigides.	Grande flexibilité adaptative, surtout avec données séquentielles.
Interprétation intuitive	Difficile pour le grand public (ex : confusion sur p-value).	Plus intuitif pour l'utilisateur final (probabilité directe sur les hypothèses).
Acceptabilité réglementaire	Traditionnellement favorisé par FDA, EMA, etc.	Gagne du terrain (adaptive designs, essais bayésiens en oncologie).

Méthodes bayésiennes : flexibilité et puissance

Approches adaptatives pour les essais cliniques et le développement de procédés

Approche bayésienne de l’ajustement de doses

Fonctionnement technique : étapes

Étape	Description
1	Définir un prior : une croyance initiale sur la probabilité que chaque dose soit sûre et efficace.
2	Administrer une dose à un ou plusieurs patients.
3	Observer les résultats (réponse thérapeutique, toxicité).
4	Mettre à jour la distribution de croyance via la règle de Bayes (posterior).
5	Utiliser le posterior pour décider de la prochaine dose : augmenter, maintenir, ou réduire selon la probabilité d'atteindre un objectif thérapeutique (ex : taux de toxicité cible).
6	Répéter jusqu'à atteindre un critère de décision (stopping rule : saturation des données, seuil d'efficacité ou de toxicité atteint).

Modèles bayésiens typiques pour l'ajustement des doses

CRM (Continual Reassessment Method) :

- Méthode classique pour déterminer la **MTD (Maximum Tolerated Dose)**.
- Utilise un modèle probabiliste pour relier la dose au risque de toxicité.
- À chaque cohorte de patients, le modèle est mis à jour pour ajuster la dose suivante.

BLRM (Bayesian Logistic Regression Model) :

- Utilisé principalement dans les essais de phase I en oncologie.
- Modèle logistique bayésien reliant dose et probabilité de toxicité dose-limitante (DLT).
- Permet de modéliser la courbe dose-réponse avec incertitude.

Escalation with Overdose Control (EWOC) :

- Assure que la probabilité de donner une dose trop toxique reste en dessous d'un seuil prédéfini.
- Contrôle du risque à chaque étape de l'escalade.

Exemple simplifié : Escalade de doses en phase I

Situation

- Objectif : trouver une dose telle que 30% des patients présentent une toxicité dose-limitante (DLT).
- Doses testées : 10 mg, 20 mg, 40 mg, 80 mg.

Processus bayésien

1. **Prior** : On pense initialement que 10 mg = 5% DLT, 20 mg = 15%, 40 mg = 30%, 80 mg = 50%.
2. **1^{re} cohorte** (10 mg) : 0 DLT → posterior mis à jour → forte confiance sur sécurité à 10 mg → montée à 20 mg.
3. **2^e cohorte** (20 mg) : 1 DLT sur 3 → update → probabilité que 20 mg soit autour de 20-25% DLT → monte à 40 mg prudemment.
4. **3^e cohorte** (40 mg) : 2 DLT sur 3 → update → forte probabilité que 40 mg soit trop toxique → redescend à 20 mg ou explore 30 mg.

Avantage	Explication
Efficacité accrue	Permet d'atteindre plus rapidement la dose optimale sans exposer inutilement des patients à des doses inefficaces ou trop toxiques.
Flexibilité adaptative	Peut intégrer des informations en continu, y compris externes (ex : données d'autres essais).
Éthique renforcée	Réduit l'exposition des patients à des doses suboptimales.
Moins de patients nécessaires	Car décisions mieux informées à chaque étape.

Mais... Acceptabilité réglementaire : certaines autorités (FDA, EMA) exigent encore des validations supplémentaires pour accepter une

Pour aller plus loin : Probabilité conformationnelle

Qu'est-ce que la "probabilité conformationnelle" ?

Probabilité conformationnelle (conformational probability en anglais) est une approche théorique émergente qui tente de modéliser autrement l'incertitude, sans se reposer exclusivement sur les axiomes classiques de Kolmogorov (et donc indirectement sans dépendre strictement du formalisme bayésien).

En synthèse :

- Elle repose sur l'idée que l'incertitude est fondamentalement une propriété des structures d'information (les *formes* que prennent les données et leur organisation), pas uniquement des événements abstraits.
- Au lieu d'attribuer une probabilité à une hypothèse fixe, on attribue une structure dynamique de crédibilité qui évolue avec l'accumulation d'informations.
- Ce n'est ni purement fréquentiste, ni purement bayésien.

Critique plus large : émergence d'un "post-bayésianisme" ?

Il existe aujourd'hui plusieurs mouvements critiques du bayésianisme "pur" :

- **Inférence conformationnelle** (Karl Friston, en neurosciences) : notion que la cognition humaine minimise des "surprises" plus que maximiser une postérieure classique.
- **Entropic inference** (Ariel Caticha) : approche fondée sur la maximisation de l'entropie relative plutôt que sur l'actualisation stricte via Bayes.
- **Approches fiduciaires** (R. A. Fisher, ressuscitées) : inférence sans prior subjectif.

Tous posent, sous des angles différents, que l'évolution des croyances doit être moins rigide et plus intégrée dans des dynamiques d'information.

Critique	Bayésien traditionnel	Probabilité conformationnelle
Nature de l'incertitude	Incetitude quantifiée par une distribution sur des hypothèses fixes.	Incetitude est une propriété émergente de la structure d'information.
Evolution	Mise à jour des croyances par une règle fixe (Bayes' Rule).	Évolution dynamique et contextuelle des structures d'incertitude.
Rigidité	Dépendance forte aux modèles a priori.	Plus grande plasticité face à des informations non-conformes.
Sensibilité aux modèles	Forte (bad prior = bad inference).	Moins sensible car modèles évolutifs.
Question	Réponse	
L'approche bayésienne est-elle remise en question ?	Oui, partiellement, surtout sur son aspect rigidité / dépendance au prior.	
La probabilité conformationnelle est-elle une alternative crédible ?	Théoriquement prometteuse mais encore émergente et peu opérationnelle à large échelle.	
Quel est l'état actuel ?	Le Bayésianisme reste la méthode dominante mais son monopole conceptuel est challengé par des visions plus dynamiques de l'incertitude.	

R et SAS : l'arsenal statistique

Exemples de scripts simplifiés

R (analyse d'efficacité)

```
model <- survfit(Surv(time, status) ~ treatment, data = clinical_data)
plot(model)
```

La méthode O'Brien-Fleming est une approche statistique importante utilisée dans les essais cliniques pour gérer les analyses intermédiaires multiples. Voici plus de détails sur cette méthode :

La méthode O'Brien-Fleming est une technique de correction pour les tests multiples dans les analyses intermédiaires d'essais cliniques. Elle a été développée par Peter O'Brien et Thomas Fleming en 1979. Son objectif principal est de contrôler le taux d'erreur de type I (faux positifs) lorsque plusieurs analyses sont effectuées au cours d'un essai.

Caractéristiques principales :

1. Elle utilise un seuil alpha très conservateur au début de l'essai, qui devient progressivement moins strict à mesure que l'essai approche de sa conclusion.
2. Cette distribution asymétrique de l'alpha permet de dépenser très peu du budget d'erreur alpha global dans les premières analyses, réservant la majeure partie pour l'analyse finale.

Formule de base :

Pour k analyses avec un alpha global α , le seuil critique pour la j-ème analyse est approximativement :

$$\alpha_j \approx 2(1 - \Phi(Z_{\alpha/2}/\sqrt{j/k}))$$

où Φ est la fonction de distribution cumulative de la loi normale standard.

Avantages :

- Préserve la puissance statistique pour l'analyse finale
- Réduit le risque d'arrêter prématurément un essai sur un résultat faussement positif
- Largement acceptée par les agences réglementaires comme la FDA

Exemple pratique :

Dans un essai avec 4 analyses prévues et un alpha global de 0,05, les seuils critiques selon O'Brien-Fleming seraient approximativement :

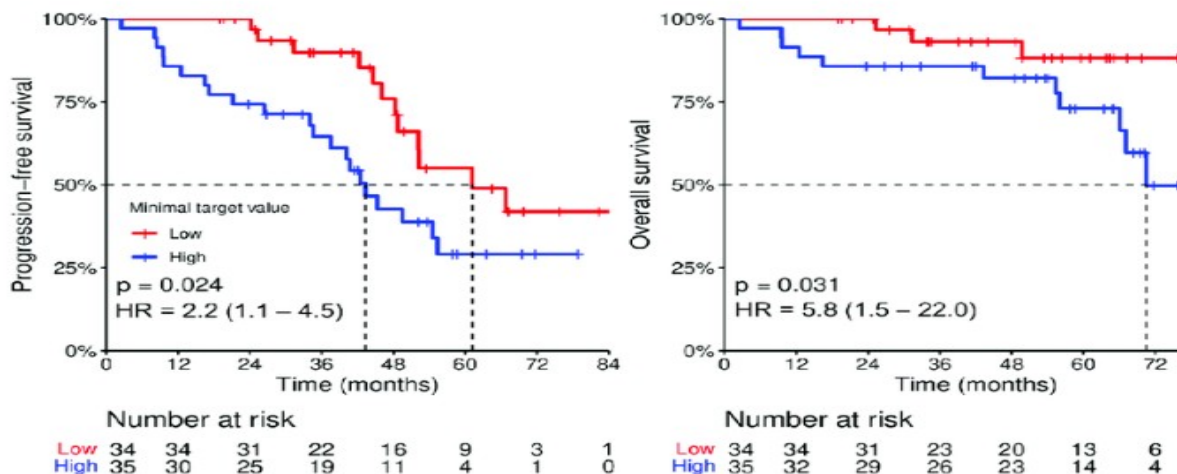
- 1ère analyse : $p < 0,0001$
- 2ème analyse : $p < 0,0013$
- 3ème analyse : $p < 0,0085$
- Analyse finale : $p < 0,041$

Tes

Cette méthode est implémentée dans les logiciels R (package 'gsDesign') et SAS (proc SEQDESIGN) pour faciliter la planification et l'analyse des essais avec analyses intermédiaires.

Mixed models

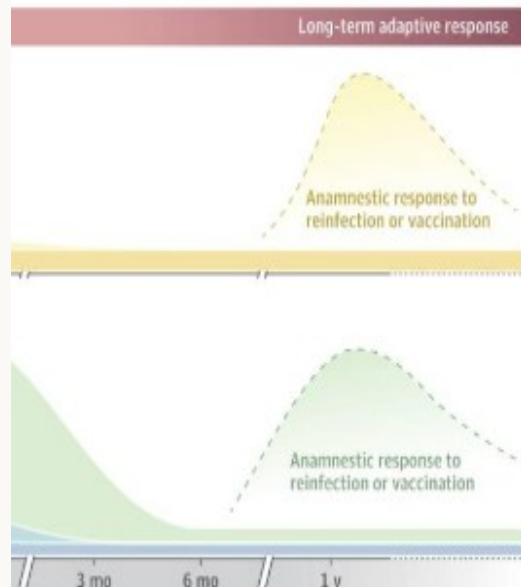
Conditional power



Courbes de Kaplan-Meier de la survie sans progression (gauche) et de la survie globale (droite) sous rituximab dans la leucémie lymphoïde chronique (Le Tilly et al., manuscrit en préparation) Comparaison de la survie de patients selon leur nadir de cible antigénique, estimée par un modèle TMDD, en rouge : nadir inférieur à la médiane (N = 34), en bleu : nadir supérieur à la médiane (N = 35)

Points méthodologiques

- Tests statistiques utilisés : log-rank test, tests t de Student, χ^2
- Gestion des données manquantes : méthodes d'imputation multiple, analyses de sensibilité
- Analyses intermédiaires multiples : contrôle du risque alpha (méthodes O'Brien-Fleming, Pocock)



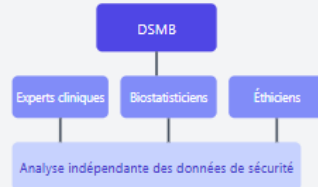
Gestion des données manquantes

- LOCF (Last Observation Carried Forward)
- Multiple imputation (MICE, Amelia)
- Imputation par la moyenne/médiane
- Méthodes hot-deck et cold-deck
- Régression (stochastique, prédictive)
- Maximum likelihood (FIML, EM)
- Pattern-mixture models

Data Safety Monitoring Boards

DSMB : gardiens de la sécurité

Schéma organisationnel



Structure indépendante composée d'experts qui examinent les données de sécurité des essais cliniques en cours pour protéger les participants.

Workflow

- 1 Collecte/analyse automatisée**
Systèmes automatisés de collecte et d'analyse continue des données de sécurité
- 2 Détection d'alertes**
Identification des signaux de sécurité et événements indésirables critiques
- 3 Processus de décision**
Évaluation et recommandations: continuation, modification ou arrêt de l'essai

Points éthiques

Balance bénéfice/risque

Évaluation continue pour garantir que les bénéfices potentiels justifient toujours les risques encourus

Challenges de l'accélération

Maintenir la rigueur d'évaluation face à la pression pour des résultats rapides

Transparence des données

Communication appropriée des résultats aux investigateurs et participants

Le plus grand défi est de prendre des **décisions en temps réel** avec des **données incomplètes**, tout en gardant **toujours à l'esprit la sécurité des participants** et **l'intégrité scientifique de l'étude**

Retour d'Expérience et Étude de Cas

Timeline comparée



Pfizer/BioNTech vs J&J : approches contrastées

Tableau comparatif

Aspect	Pfizer/BioNTech	Johnson & Johnson
Méthodes de recrutement	Centralisé - Plateformes digitales globales	Décentralisé - Réseaux hospitaliers locaux
Stratégies d'analyse	Analyses intermédiaires fréquentes Bayésien séquentiel	Analyses moins fréquentes Approche fréquentiste classique
Technologies de monitoring	ePRO avancé, IoT & wearables Surveillance à distance	Approche plus traditionnelle Visites sur site plus fréquentes
Résultats d'efficacité	95%	66%

Points différenciants

Pfizer/BioNTech

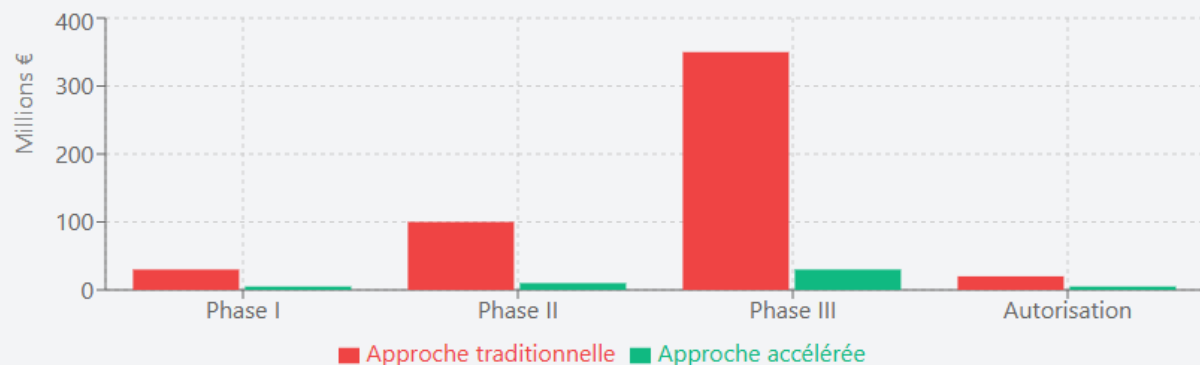
- Plateforme technologique ARNm innovante
- Rapidité exceptionnelle de développement
- Infrastructure numérique globale
- Analyse intérimaire précoce et fréquente
- Investissement massif en ressources
- Chaîne logistique complexe (ultra-froid)

Johnson & Johnson

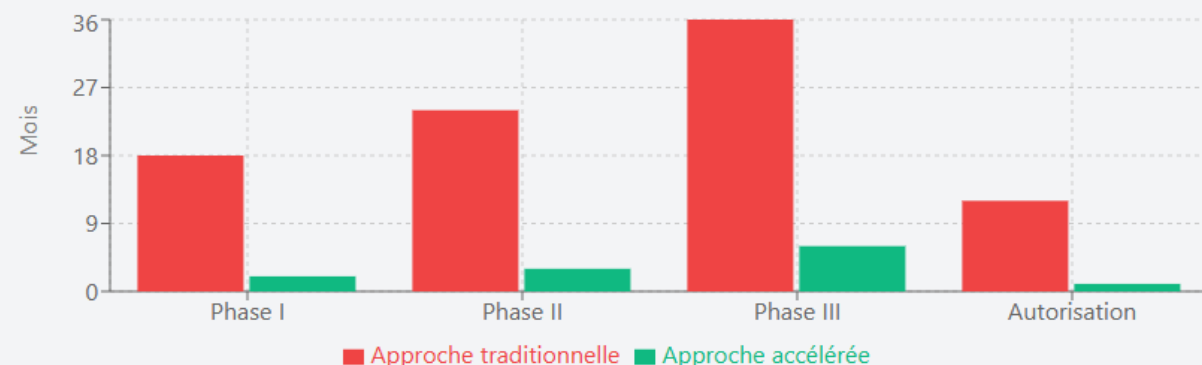
- Technologie à vecteur viral (adénovirus)
- Approche à dose unique
- Stabilité thermique supérieure
- Déploiement plus facile (logistique)
- Recrutement plus diversifié géographiquement
- Investissement dans l'équité vaccinale mondiale

Impact IT sur les Coûts

Coûts (millions €) - 1-30M\$ vs 100-500M\$



Délais (mois) - 5-10 ans vs 12 mois



Facteurs d'économie



Parallélisation phases

Réduction de 40% des délais par chevauchement des phases



Analyse temps réel

Détection précoce des tendances et décisions éclairées



Recrutement optimisé

Algorithmes prédictifs pour cibler les patients éligibles



Monitoring à distance

Réduction de 80% des visites de surveillance sur site

Transfert industriel



Scale-up bioproduction

- Modélisation prédictive pour transfert d'échelle
- Réduction de 60% des lots perdus
- Accélération time-to-market



Validation de procédés

- Validation continue vs validation rétrospective
- Qualification en temps réel des équipements
- Documentation électronique automatisée



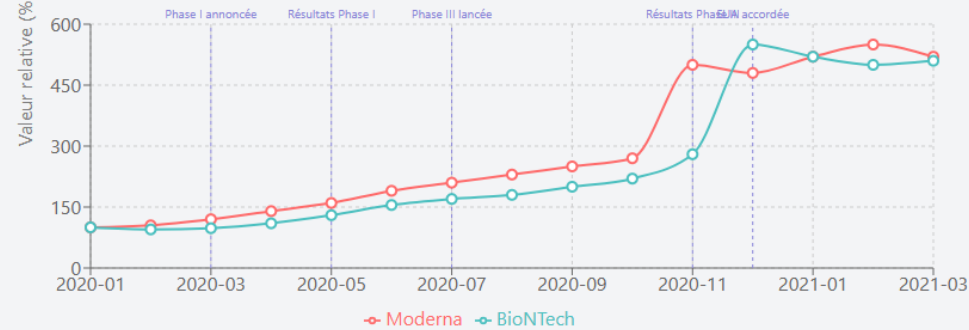
Contrôle qualité prédictif

- IA pour prédiction des déviations qualité
- Contrôle en ligne vs tests de libération
- Réduction de 70% des délais de libération de lots

Effet sur les Investisseurs

Confiance des marchés et communication

Corrélation annonces d'essais / cours boursiers



Étude de cas: Impact des communiqués

Moderna

+400%

Après annonce des résultats Phase 3 (94.5% efficacité)

Nov 2020:  x5

Capitalisation: \$20 milliards à \$100 milliards

BioNTech

+250%

Après autorisation d'urgence (EUA) par FDA

Déc 2020:  x3.5

Capitalisation: \$25 milliards à \$87.5 milliards

Stratégies de communication

🕒 Timing des annonces

- Publication en dehors des heures de marché
- Coordination avec évaluations réglementaires
- Espacement stratégique des nouvelles

📈 Gestion des attentes

- Prévisions conservatrices et dépassement
- Transparence sur les incertitudes
- Forward guidance clair avec marges d'erreur

📅 Communication scientifique vs financière

- Publications peer-reviewed vs communiqués
- Séparation temporelle annonces médicales/financières
- Formation des porte-paroles scientifiques
- Conformité SEC et réglementations boursières

Récapitulatif : Acquis majeurs



Validation du concept ARNm

Première validation clinique à grande échelle de la technologie ARNm



Accélération sans précédent

Processus clinique complet en moins d'un an vs 5-10 ans habituellement



Coopération internationale

Niveau inédit de collaboration entre laboratoires, régulateurs et gouvernements



Nouveau paradigme digital

Démonstration de la puissance des outils numériques pour les essais cliniques

Chiffres clés

70k+

Participants aux essais

Plus de 70,000 volontaires dans les essais de Phase III des vaccins ARNm

>90%

Efficacité démontrée

Protection remarquable contre les formes symptomatiques de COVID-19



Profil de sécurité

Effets secondaires principalement légers à modérés et transitoires

Surveillance continue post-autorisation

Défis à venir

Production à échelle mondiale

- Passage à l'échelle des procédés de fabrication
- Approvisionnement en matières premières
- Construction de nouvelles lignes de production

Chaîne du froid

- Maintien à -70°C pour certains vaccins
- Infrastructure de distribution globale
- Traçabilité et contrôle qualité

Déploiement équitable

- Accès pour les pays à revenus faibles/moyens
- Dispositifs COVAX et autres initiatives
- Transferts de technologie

Surveillance post-approbation

- Pharmacovigilance à grande échelle
- Évaluation de l'immunité à long terme
- Adaptation aux variants

Les essais cliniques accélérés des vaccins COVID-19 ont établi un nouveau paradigme qui redéfinira notre approche du développement des médicaments et vaccins pour les décennies à venir.

L'innovation réside autant dans la méthodologie que dans les produits eux-mêmes.



J201 à J300 : Le défi industriel

Capacités de production ET Transfert

- Plateformes ARN messenger vs vecteurs viraux
- Hub OMS « tech-transfer » express (Afrique du Sud, OMS, Unit-aid)
- Consortiums publics-privés pour modularité industrielle

Bottlenecks de l'approvisionnement

- Matériaux critiques : lipides, enzymes, bioréacteurs à usage unique
- Verre, seringues low dead-space (150 goulots identifiés UE)
- Logistique Froid : containers cryogéniques & suivi IoT

Main-d'OEUVRE ET CompÉTences

- Bootcamps GMP, partenariats écoles/industries
- e-learning 24/7 et tutoriels pratiques
- Pair-programming et formation accélérée (réalité augmentée)

Fill-Finish & Conditionnement

- Lignes 24/7 avec robots collaboratifs (cobots)
- Traçabilité RFID à la manière du luxe

Distribution & Collaboration

- Initiative COVAX (GAVI, CEPI, OMS) & nationalismes vaccinaux
- Usines satellites régionales pour résilience locale

Facteurs clés de succès

- Parallélisation phases (Toyota lean assembly)
- Analyse & dashboards en temps réel (Industrie 4.0)
- Recrutement optimisé & formations intensives
- Monitoring à distance IoT & wearables

Leçons ET Perspectives

- Diversification fournisseurs (semi-conducteurs)
- Agilité réglementaire : rolling reviews
- Infrastructures locales & écosystèmes hybrides



Jours 201-300 : Le défi industriel

Contexte de Production

Situation

- Vaccins validés cliniquement
- Demande mondiale sans précédent
- Contraintes techniques majeures

Chiffres

- Objectifs de production (milliards de doses)
- Capacités existantes vs nécessaires
- Investissements requis

Face à l'ampleur du défi, nous avons dû repenser entièrement notre chaîne de production. Ce qui prenait habituellement des années devait être accompli en quelques mois.

ERP : épine dorsale de la production

Outils ERP



Fonctionnalités clés

- Planification intégrée
- Gestion des matières premières
- Contrôle qualité continu
- Traçabilité complète

Cas biopharma

Suivi des lots

Traçabilité complète du batch, de la matière première au produit fini

Gestion des bioréacteurs

Monitoring temps réel et ajustements automatisés des paramètres critiques

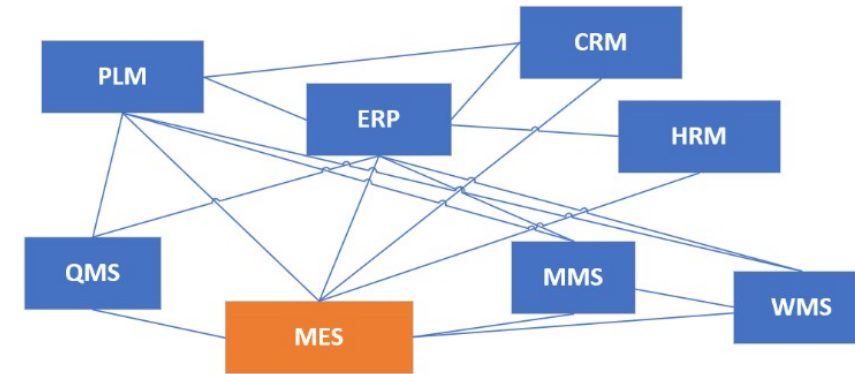
Libération paramétrique

Validation automatisée basée sur l'analyse des données de process

MES / MOM: Integration with surrounding systems

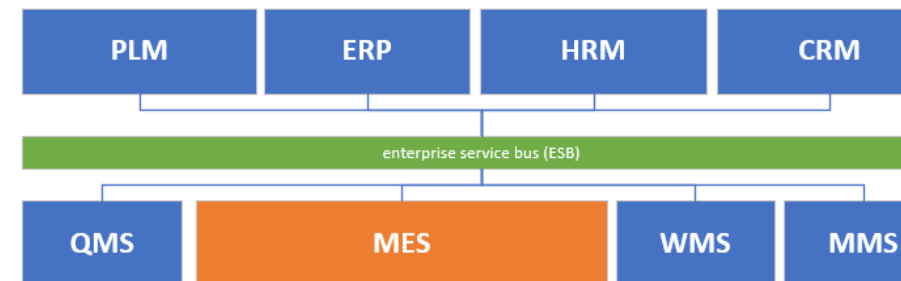
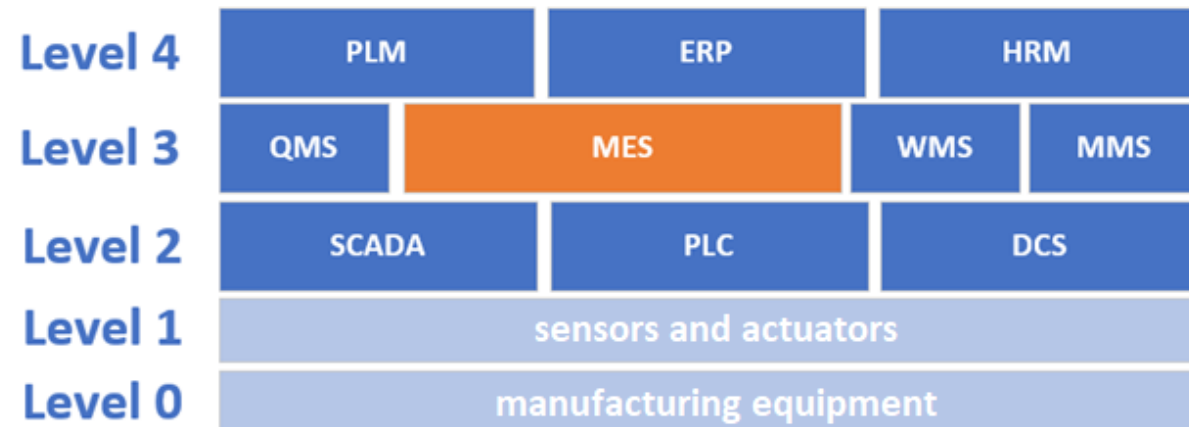
To be successful with MES / MOM implementation, it is important to make the integration with surrounding systems to ensure automated data transfer. This is a demanding part of the implementation, both from a technical and organizational point of view. Many implementation projects will fail because of their unmanaged integration with the surrounding systems. That is why it is very desirable to be well prepared for this part, which can be helped by the following article. Typical systems with which the MES system integrates are as follows:

ERP - Enterprise Resource Planning
PLM - Product Lifecycle Management
HRM - Human Resource Management
CRM - Customer Relationship Management
QMS (Quality Management System)
WMS - Warehouse Management System
CMMS - Computerized Maintenance Management System
SCADA - Supervisory Control And Data Acquisition
PLC - Programmable Logic Controller
DCS - Distributed Control Systems



Purdue model

In the 1990s, a reference model for enterprise system architecture was created under the name Purdue Enterprise Reference Architecture – more information on Wikipedia can be found here. This model defines vertical system architecture at 5 levels (Level 0 to Level 4) from physical processes, sensors and controllers, control systems, manufacturing information systems to enterprise business systems. This model was subsequently used as one of the key models in one of the most important standards for MES systems, namely ISA 95, where MES systems are placed at level 3.



Challenges faced by the big pharmas : Financial Performance is harder to achieve



« Productivity » of the R&D pipeline

- Less than 10% of drug candidates reach the market. Only a few will become block busters
- Drug development median price is above 1 billion \$ (1.3-2.8) and is a lengthy process (~10 years)
- Bias in physiopathological targets : Oncology, Metabolic disorders, cardiology.... But antibiotics R&D is lacking investment
- New Molecular Entities numbers are decreasing

Drug Development

- Discovery is outsourced to start-ups (scouting)
- Transition from small molecules (chemicals) to biologics, from stochastic discovery to intentional design

Clinical trials

- Outsourced to CRO, costs are on the rise
- Patient recruitment is hard, diversity is a challenge
- Quality / Quantity of clinical data is an issue

Scaling up for production

- Need of GxP qualified production line (not elastic)
- Transition from R&D to Production is lengthy
- Global Supply Chain disruption
- Quality issues

«Ageing » fixed size/production factories

Transition from small molecules to biologics requires new factories (very different processes)

- Retrofit is too expensive
- New factories must be digital factories with extended supply chain capabilities (interconnect factories and report to a central control tower)
- New factories are expensive and must be more flexible in terms of production lines
- Global Supply chains must be secured to avoid disruption of supply
- Relocation and de-concentration requires cheaper, flexible factories : ability to rent additional production capabilities

Procurement & Sourcing

- Global Supply Chains have been highly concentrated leading to major disruptions in Active Pharmaceutical Ingredients supply

Manufacturing

- GxP compliance is a heavy constraint used not to evolve
- Industry needs flexible production lines
- Quality is an issue linked to Sustainability

Cost of sales are too high

Powermap has been evolving. Patient voice must be heard while importance of KOL is decreasing. Government MOH are also key in leading price conversation

- The way Big Pharma are doing marketing must evolve leveraging digital marketing
- Sales organization are expensive to maintain and more regulations restrict access to KOL

Marketing & Sales

- Sales reps are underequipped when it comes to real time performance of their portfolio

Consumer engagement & Pharmacovigilance

- Differs depending on « over the counter » drugs or prescribed drugs.
- Even with prescribed drugs, big pharma needs patients to be ambassadors of their products.
- Depending on the regulations in place, national pharmacovigilance databases can help doing differential marketing

What if ?



LabOps are fully automated ...

- From configuration / experimental settings to deployment on the analysis line
- The exhaustive configuration of an experiment is saved in a secured database to track every experiences
- You can « fork » experiences, i.e. retry it the step before it failed.
- Outcomes of the experiences (data) are exhaustively collected and secured along with all the experience metadata

Clinical trials are virtualized...

- Software Defined Multicentric clinical trial IT can be deployed pressing one button
- Patient recruitment is augmented with teleconsultation leveraging social networks
- Clinical data is enriched with RWE and IOT/wearable devices and collected securely with all the required metadata

There's a continuum between Lab MVP and production

- The modular factory built as an extension of the lab
- Taking benefit of all the data collected during the R&D with complete lineage of different experiences

Software Defined GxP...

- GxP compliance processes are software defined and track all the necessary data for audit in a log analytics workspace
- New manufacturing parameters can be pass to the production line directly from the « pre production line » in the R&D lab

Modular factories cloud+edge

- You can deploy modular production line with one click (local IT deployed at the edge from the cloud)
- You can relocate plants on your local geographies + rent additional production capabilities with guaranteed production quality
- You can capture all the data from the production line and manage it from a central control tower
- Leverage AI algorithms to listen the the data stream and optimize manufacturing operations

Extend your control downstream to your suppliers

- Cloud + edge integration could extend your control tower capabilities to your tier 1 supplier (renting additional production capabilities)
- Quality management is automated end to end and can be securely shared to regulation institution

Industrial Metaverse continuum from lab to manufacturing to market access enables...

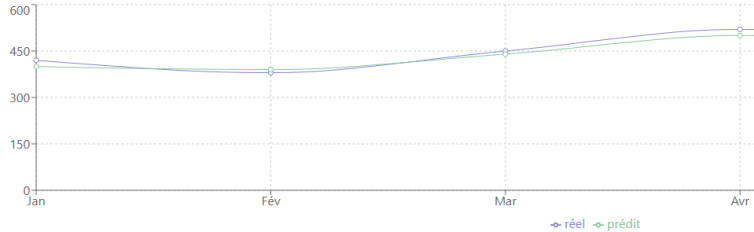
- One drug candidate can reassessed with new parameters and targetting a new physiopathological target (derisk the drug discovery and clinical qualification)
- Clinical trials for clinical validation can be deployed at less costs, and meta-cohorts can be built out of previous trials thanks to the systemic tracking of metadata + clinical data
- Plants can be « micro managed » from a central point, reconfiguring the production line in minutes and orchestrated with the inventories of API ... while guaranteeing/improving the quality end to end with extended supply

Algorithmes prédictifs et logistique

Optimisation de la chaîne logistique par l'intelligence artificielle

Modèles de prévision de la demande

Comparaison entre les valeurs réelles et prédites sur 6 mois

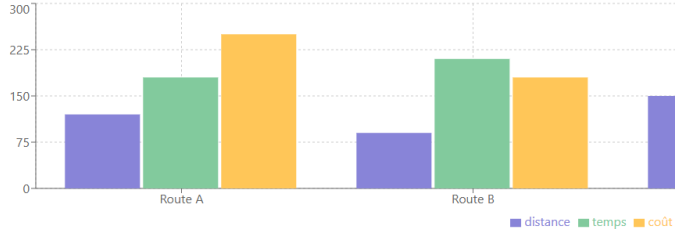


Avantages des modèles prédictifs :

- Réduction des stocks et des coûts associés
- Amélioration du taux de service client
- Anticipation des pics saisonniers
- Optimisation des ressources humaines et matérielles

Algorithmes de routage

Comparaison des paramètres des différentes routes

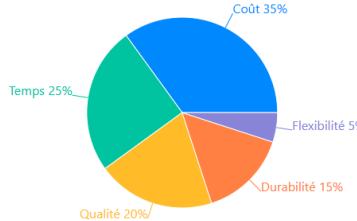


Types d'algorithmes :

- TSP (Problème du voyageur de commerce)
- VRP (Vehicle Routing Problem)
- Dijkstra et A* pour chemins optimaux
- Algorithmes génétiques pour l'optimisation multicontrainte

Optimisation multiparamétrique

Importance relative des paramètres d'optimisation



Techniques d'optimisation :

- Programmation linéaire et non-linéaire
- Algorithmes stochastiques (recuit simulé, colonies de fourmis)
- Optimisation par essaim particulaire
- Apprentissage par renforcement pour l'adaptation continue

Machine learning pour prédiction de demande

Modèles utilisés

- Séries temporelles (ARIMA, SARIMA)
- Réseaux de neurones (LSTM, GRU)
- Forêts aléatoires pour variables catégorielles
- Modèles hybrides combinant plusieurs approches

Variables explicatives

- Historique des ventes
- Saisonnalité et tendances
- Prix et promotions
- Événements externes (météo, vacances, etc.)
- Données concurrentielles

Intégration dans le processus de décision

Les prédictions de demande alimentent l'ensemble du système d'optimisation logistique, de l'approvisionnement à la distribution finale, en passant par la gestion des stocks et la planification des ressources.

Programmation linéaire pour allocation

Principes fondamentaux

La programmation linéaire est une technique mathématique d'optimisation où une fonction objectif linéaire est maximisée ou minimisée sous contraintes linéaires

Formulation générale :

Maximiser/Minimiser $Z = c_1x_1 + c_2x_2 + \dots + c_nx_n$

Sous contraintes :

$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1$

$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2$

...

$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m$

$x_1, x_2, \dots, x_n \geq 0$

Applications en logistique

- Allocation optimale des ressources
- Planification de la production
- Gestion des stocks multi-sites
- Affectation des moyens de transport
- Optimisation des emplacements de stockage

Algorithmes de résolution

- Méthode du simplexe
- Méthode des points intérieurs
- Décomposition de Dantzig-Wolfe
- Algorithme du Branch-and-Bound (pour les variables entières)

Cartes interactives



Capacités de production

Visualisation ArcGIS des installations et leur capacité productive



Corridors logistiques

Cartographie QGIS des voies d'approvisionnement et de distribution



Sites prioritaires de distribution

Identification géospatiale des points stratégiques de livraison

Applications technologiques

Optimisation des routes

Algorithmes d'itinéraires optimaux réduisant les délais et la consommation

Analyse de vulnérabilité

Évaluation des risques géographiques et climatiques sur la chaîne logistique

Modélisation d'impact

Prévisions des effets des perturbations sur la disponibilité des produits

Transfert aux produits biologiques

Chaîne du froid

Mapping des contraintes thermiques et points critiques de contrôle

Distribution contrôlée

Gestion géospatiale des produits sensibles à température régulée

Applications EBI

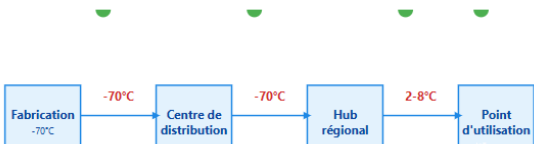
Solutions spécifiques pour enzymes, cultures biologiques et bioréactifs

Chaîne d'Approvisionnement

La chaîne du froid extrême

Infrastructure complète de la chaîne du froid ARNm

Schéma technique



Technologies clés

- Congélateurs ultra-basse température
- Conteneurs thermiques spécialisés
- Systèmes de surveillance IoT
- Logistique adaptée au froid extrême

Points critiques

- Transferts entre installations
- Pannes de courant
- Transport dernier kilomètre
- Formation du personnel

Challenges spécifiques

Maintien à -70°C vs 2-8°C

Exigence de température extrêmement basse pour les vaccins ARNm, comparée aux vaccins traditionnels nécessitant seulement une réfrigération standard.

Monitoring continu

Nécessité de surveiller constamment la température à chaque étape, sans interruptions ni zones aveugles dans la chaîne logistique.

Redondance des systèmes

Implémentation de systèmes de secours à chaque niveau pour prévenir les défaillances critiques et assurer la continuité.

Last-mile delivery

Complexité du maintien de la chaîne du froid dans la phase finale de livraison, particulièrement dans les zones reculées ou à ressources limitées.

Conséquences d'une rupture de la chaîne du froid

Perte d'efficacité

Risque sanitaire

Pertes financières

Innovations



Conteneurs spéciaux développés

Utilisation de matériaux à changement de phase et isolation avancée pour maintenir les températures ultra-basses pendant plusieurs jours sans alimentation externe.



IoT pour suivi en temps réel

Capteurs connectés transmettant données de température, localisation et conditions environnementales en continu vers des tableaux de bord centralisés.



Systèmes d'alerte précoce

Technologies prédictives identifiant les risques potentiels avant qu'ils n'affectent la chaîne du froid, permettant une intervention proactive.



Solutions de dernier kilomètre

Méthodes innovantes pour maintenir les conditions de température lors de la livraison finale, incluant des dispositifs portables et formations spécialisées.

Impact des innovations

99.8%

Fiabilité

-40%

Coûts logistiques

+60%

Couverture géographique

Tableau comparatif : Performances des différentes biotechs

Entreprise	Technologie	Cap. Boursière (Mar 2021)	Évolution Déc 2020-Mar 2021	CA 2020	CA 2021 (est.)
Moderna	ARNm	87 Mrd \$	+435%	803 M \$	18,4 Mrd \$
BioNTech	ARNm	43 Mrd \$	+328%	482 M \$	14,9 Mrd \$
AstraZeneca	Vecteur viral	130 Mrd \$	+19%	26,6 Mrd \$	32,7 Mrd \$
Johnson & Johnson	Vecteur viral	428 Mrd \$	+12%	82,6 Mrd \$	93,8 Mrd \$
CureVac	ARNm (échec phase 3)	18 Mrd \$	-32%	48 M \$	103 M \$

Leaders ARNm : Moderna & BioNTech

- ✓ Croissance explosive de la valorisation
- ✓ Transition de biotech R&D à entreprise commerciale
- ✓ Premiers bénéfices significatifs en 2021
- ✓ Pipeline de produits ARNm élargi post-Covid

Approche traditionnelle : AZ & J&J

- ✓ Croissance modérée de la valorisation
- ✓ Impact dilué (activités diversifiées)
- ✓ Stratégie "sans profit" pendant la pandémie
- ✓ Défis d'image liés aux effets secondaires

Échec relatif : CureVac

- X Résultats d'efficacité phase 3 décevants (47% vs >90% pour Moderna/BioNTech)
- X Retard dans le calendrier de développement et d'approbation
- X Correction boursière après l'échec clinique
- X Redirection stratégique vers vaccins de 2ème génération et autres applications ARNm

Points clés

Corrélation avec les premières livraisons

Les valorisations ont explosé en parallèle des premières livraisons réussies, démontrant la capacité des entreprises à tenir leurs promesses de production massive.

- Moderna : +50% après la première livraison
- BioNTech : +42% après début de distribution EU

Impact des contrats gouvernementaux

Les annonces de contrats massifs ont directement impacté les cours, créant un effet d'aubaine sans précédent pour les sociétés de biotech.

Moderna (US)	\$15,3 Mrd
BioNTech (EU)	\$12,9 Mrd
AstraZeneca (UK)	\$8,2 Mrd

Différenciation des acteurs

Une nette séparation s'est créée entre les gagnants (technologies ARNm avec forte efficacité) et les acteurs plus conventionnels ou moins performants.

GAGNANTS

PERDANTS RELATIFS

Analyse des facteurs de performance

Facteurs positifs

- Efficacité clinique > 90%
- Capacité de production à grande échelle
- Premiers sur le marché
- Flexibilité de la technologie ARNm

Facteurs négatifs

- Exigences logistiques complexes
- Efficacité réduite contre variants
- Retards de production
- Problèmes d'effets secondaires

J+300 : Vaccins déployés, nouveaux défis

Récapitulatif : Acquis majeurs

- Production massive établie à l'échelle mondiale
- Systèmes de distribution opérationnels dans plusieurs régions
- Premiers millions de vaccinations effectués avec succès

Innovation transférable

- Optimisation logistique pour les produits biologiques sensibles
- Déploiement de monitoring en temps réel des procédés
- Scaling rapide et agile des bioprocédés industriels

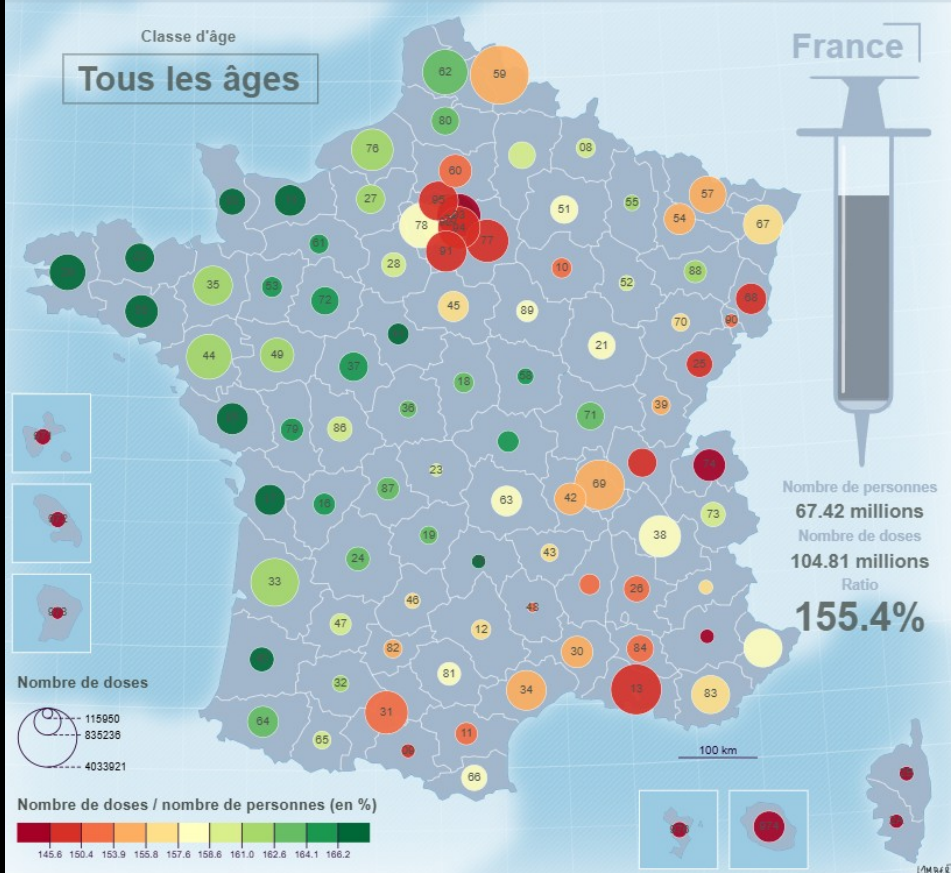
Transition

Introduction au suivi post-déploiement : surveillance de la durée de protection vaccinale, efficacité face aux nouveaux variants, et adaptation des stratégies industrielles pour anticiper les évolutions épidémiologiques.

**Covid-19 : la
nouvelle campagne
de vaccination a
débuté le 2 octobre
2023**



État de la vaccination dans votre département au 10 juillet 2023



J301 à J365 : ORGANISER LA VACCINATION ET LE SUIVI EPIDEMIOLOGIQUE

DEPLOIEMENTS MASSIFS

- Centres de vaccination, unités mobiles, pharmacies partenaires
- Réservation numérique et gestion des flux patients

Optimisation logistique

- Planification des livraisons et gestion de la chaîne du froid ultra-basse température
- Suivi en temps réel des stocks et des capacités de vaccination

Traçabilité complète

- Suivi des lots de vaccins de la production à l'injection
- Systèmes sécurisés de carnet vaccinal électronique

Suivi vaccinal dynamique

- Cartographie en temps réel des taux de couverture
- Ciblage des zones sous-vaccinées pour intervention rapide

Surveillance pharmacovigilance

- Détection en temps réel des événements indésirables
- Croisement automatisé des données pour alerter précocement

Apport de la bioinformatique

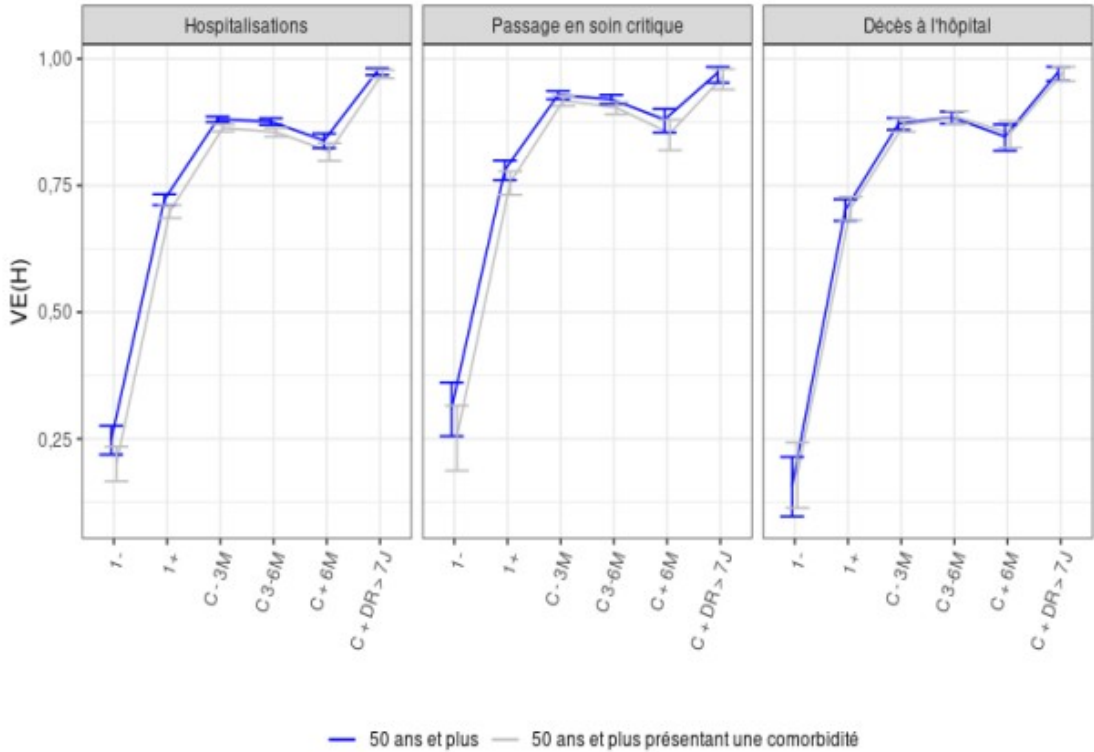
- Orchestration des données logistiques et médicales

Bilan des Résultats

Jours 301-365 : Le bilan d'une course sans précédent

Graphique 6 • Efficacité vaccinale contre les formes sévères de Covid-19 chez les personnes de 50 ans ou plus présentant une comorbidité et dans l'ensemble de la population, suivant l'ancienneté de la dose finale

Estimations du 1^{er} février au 14 novembre 2021



Lecture > La réduction de risque de développer une forme sévère donnant lieu (i) à une hospitalisation, (ii) à un passage en soins critiques, (iii) à un décès est estimée à (i) 84 %, (ii) 88 %, (iii) 85 % chez les personnes de 50 ans ou plus dont l'ancienneté de la deuxième injection est de plus de six mois, relativement aux non-vaccinés sur la période du 1^{er} février au 14 novembre 2021. Pour les personnes de 50 ans ou plus présentant une comorbidité, de (i) 82 %, (ii) 85 %, (iii) 85 %.

Source > SI-DEP-VAC-SI-SI-VIC, calculs DREES.

Statistiques clés

123M

Millions de doses administrées

92%

Efficacité en vie réelle vs essais cliniques



Profil de sécurité confirmés

Tableau comparatif

Plateforme	Efficacité	Conservation	Durabilité
ARNm	94-95%	Réfrigération spéciale	6-7 mois
Vecteur viral	66-81%	Réfrigération standard	3-6 mois
Protéique	89-90%	Réfrigération standard	12 mois

La rapidité sans précédent avec laquelle ces vaccins ont été développés, sans compromettre la rigueur scientifique, marque un tournant dans l'histoire de la médecine moderne. Les données recueillies démontrent un impact sanitaire considérable et promettent une nouvelle ère dans le développement vaccinal.

Suivi des Variants

La menace des variants : bioinformatique en alerte

Visualisations



Interface Nextstrain montrant l'évolution des variants

Pipeline de séqu

1

Extraction

Galaxy/Nextfl

Outils bioinformatiques

Variant Calling

- GATK
- LoFreq
- FreeBayes

Classification Pangolin

- Outils dynamiques
- Nomenclature standardisée
- Mise à jour continue

Parallèle industriel

Application	Méthode	Bénéfice
Suivi de souches de production	Séquençage NGS périodique	Stabilité
Détection de dérive génétique	Comparaison séquentielle	Prévent
Contrôle qualité microbiologique	Métagénomique ciblée	Détection

La classification Pangolin est un système de nomenclature et de classification des variants du SARS-CoV-2, le virus responsable de la COVID-19. Le nom "Pangolin" est un acronyme qui signifie "Phylogenetic Assignment of Named Global Outbreak LINEages" (Attribution phylogénétique des lignées d'épidémies mondiales nommées).

Voici les principales caractéristiques du système Pangolin :

- Objectif** : Fournir une méthode standardisée pour nommer et suivre les différentes lignées génétiques du SARS-CoV-2 à mesure qu'elles évoluent et se propagent mondialement.
- Fonctionnement** :
 - Utilise des analyses phylogénétiques (l'étude des relations évolutives entre différentes séquences génétiques)
 - Attribue des désignations alphanumériques aux variants (comme B.1.1.7, P.1, B.1.617.2)
 - Ces désignations reflètent la hiérarchie évolutive des variants
- Mise à jour dynamique** : Le système est constamment mis à jour pour refléter l'émergence de nouveaux variants à mesure que le virus évolue.
- Relation avec d'autres nomenclatures** :
 - Complémentaire au système de nomenclature de l'OMS qui utilise des lettres grecques (Alpha, Beta, Delta, Omicron, etc.)
 - Plus détaillé et technique que la nomenclature de l'OMS, destinée principalement à la communauté scientifique
- Importance** :
 - Essentiel pour la surveillance épidémiologique
 - Permet la communication standardisée entre chercheurs du monde entier
 - Facilite le suivi des mutations d'intérêt qui peuvent affecter la transmissibilité, la virulence ou l'échappement immunitaire
- Outil logiciel** : Pangolin est aussi le nom de l'outil bioinformatique qui implémente ce système de classification et permet aux chercheurs d'assigner une lignée à de nouvelles séquences de SARS-CoV-2.

Ce système a joué un rôle crucial dans la coordination mondiale de la réponse à la pandémie, permettant aux scientifiques de suivre et de communiquer efficacement sur l'évolution du virus.

Modèles Prédictifs

IA et Deep Learning : prévoir pour anticiper

Architectures de Deep Learning: LSTM, CNN, Transformer

Applications épidémiologiques



Prédiction d'échappement immunitaire

Modèles prédictifs des mutations permettant d'anticiper les mutations existantes.

F1-Score

Le F1-Score est une mesure qui combine deux autres métriques : la précision et le rappel.

- **Précision** : proportion d'identifications positives qui étaient correctes. Elle répond à la question "Parmi tous les éléments que j'ai classés comme positifs, combien le sont réellement?"
- **Rappel** : proportion d'éléments positifs réels qui ont été correctement identifiés. Il répond à la question "Parmi tous les éléments réellement positifs, combien ai-je réussi à identifier?"

Le F1-Score est la moyenne harmonique de ces deux valeurs :

$$F1 = 2 \times (\text{Précision} \times \text{Rappel}) / (\text{Précision} + \text{Rappel})$$

Il est particulièrement utile lorsque les classes sont déséquilibrées, car il donne une mesure plus équilibrée de la performance que la simple précision. La valeur du F1-Score est comprise entre 0 (pire) et 1 (meilleur).

Métriques d'évaluation

Précision

94.3%

Rappel

91.7%

F1-Score

93.0%

ROC-AUC

0.968

Performances moyennes sur les modèles de prédiction d'échappement immunitaire

Applications industrielles



ROC-AUC (Receiver Operating Characteristic - Area Under Curve)

La ROC-AUC est une métrique qui évalue la qualité d'un classificateur binaire à différents seuils de classification.

- **Courbe ROC** : représentation graphique qui illustre la performance d'un modèle de classification en traçant le taux de vrais positifs (sensibilité) en fonction du taux de faux positifs (1 - spécificité) à différents seuils de décision.
- **AUC (Area Under Curve)** : surface sous la courbe ROC. Cette valeur est comprise entre 0 et 1.

Interprétation de l'AUC :

- 0.5 : équivalent à une prédiction aléatoire (le modèle n'a aucune capacité de discrimination)
- 0.5 : meilleur qu'une prédiction aléatoire
- 1.0 : prédiction parfaite (le modèle distingue parfaitement les classes)

La ROC-AUC est particulièrement utile car elle est insensible au déséquilibre des classes et fournit une mesure agrégée de la performance du modèle à tous les seuils possibles.



Domaines d'application futurs

- › **Thérapies ARNm personnalisées**
Traitement du cancer basé sur le profil génétique individuel du patient
- › **Maladies génétiques rares**
Thérapies ciblées pour maladies monogéniques et ultraorphelines
- › **Immunothérapies**
Reprogrammation du système immunitaire pour cibler diverses pathologies



Technologies dérivées

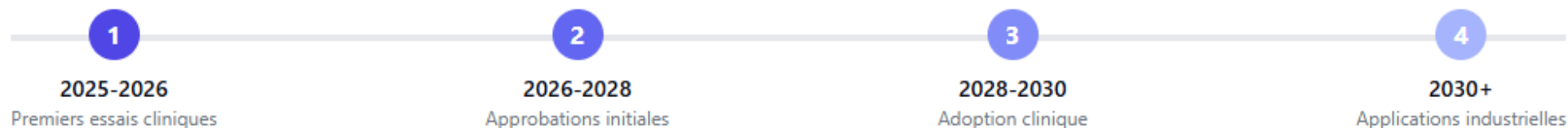
- › **Nouveaux systèmes de delivery**
Nanoparticules lipidiques de nouvelle génération et vecteurs ciblés
- › **Stabilisation à température ambiante**
Formulations lyophilisées et modifications chimiques pour éliminer la chaîne du froid
- › **Plateformes de production modulaires**
Systèmes automatisés de fabrication décentralisée et personnalisable



Impact en biologie industrielle

- › **Production d'enzymes industrielles**
Biosynthèse dirigée par ARNm pour catalyseurs biologiques spécifiques
- › **Biosenseurs ARN**
Détection ultraprécise de contaminants et monitoring de procédés en temps réel
- › **Catalyseurs biologiques programmables**
Ribozymes synthétiques pour la production de molécules complexes

Horizon d'implémentation



Réflexion Économique

J+365 : Bilan économique et perspectives

🏢 Analyse financière

Retour sur investissement

R&D investie	42,8 Mrd \$
Revenus générés	157,3 Mrd \$
ROI moyen	367%

Valorisation des technologies propriétaires

Les plateformes ARNm représentent désormais plus de 60% de la valorisation totale des acteurs clés, transformant radicalement leur modèle économique.

Durabilité des modèles économiques

Transition progressive vers des applications thérapeutiques diversifiées au-delà des vaccins, assurant des revenus plus stables et récurrents.

⚠️ Défis futurs

Adaptation aux variants

La nécessité d'actualiser continuellement les formulations représente un défi financier et logistique majeur, imposant des cycles R&D accélérés.

 Criticité: 83%

Équité d'accès global

Pression croissante pour établir des modèles de prix différenciés et des transferts technologiques vers les économies émergentes.

 Criticité: 67%

Valorisation de l'innovation de rupture

Les marchés financiers peinent encore à modéliser correctement la valeur à long terme des plateformes technologiques émergentes.

 Criticité: 75%

Le paradigme économique qui a émergé de cette année transformatrice dépasse largement la simple rentabilité à court terme. Nous assistons à la naissance d'un nouveau modèle industriel où l'agilité, l'innovation de rupture et la responsabilité sociale constituent désormais les véritables moteurs de création de valeur durable.

Mais qu'en reste-t-il 5 ans après ?

Organisation des Campagnes de Vaccination

Une opération logistique sans précédent



Infrastructure et Planification

- ✓ Méga-centres (stades, parcs des expositions)
- ✓ Sites mobiles et pop-up
- ✓ Réseau de pharmacies et cabinets médicaux



Systèmes d'information

- ✓ Portails nationaux de réservation (Doctolib, Santé.fr)
- ✓ Carnets de vaccination électroniques
- ✓ Systèmes intégrés de suivi des doses



Priorisation des Populations

- ✓ Calendriers basés sur la vulnérabilité
- ✓ Âge, comorbidités, travailleurs essentiels
- ✓ Adaptations régionales selon l'épidémiologie



Communication et Acceptabilité

- ✓ Campagnes de sensibilisation massive
- ✓ Lutte contre la désinformation
- ✓ Engagement des leaders communautaires



Logistique Ultra-froide (ARNm)

- ✓ Livraison synchronisée sous -70 °C
- ✓ Chaînes de froid spécifiques
- ✓ Gestion du dernier kilomètre



Indicateurs de Suivi

- ✓ Nombre de doses administrées par jour/semaine
- ✓ Taux de couverture par tranche d'âge et région
- ✓ Taux d'abandon (dose 1 vs dose 2)

Flux opérationnel simplifié



Réception & Stockage



Distribution



Vaccination



Suivi & Rappels



VIRUS DEFEATED. PLAYER 1
HAS LOST

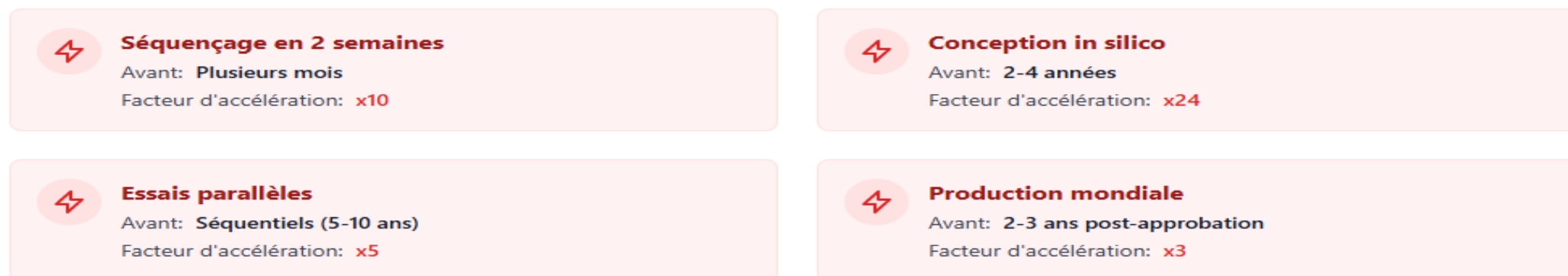
Récapitulatif Global

365 jours qui ont transformé la biopharma

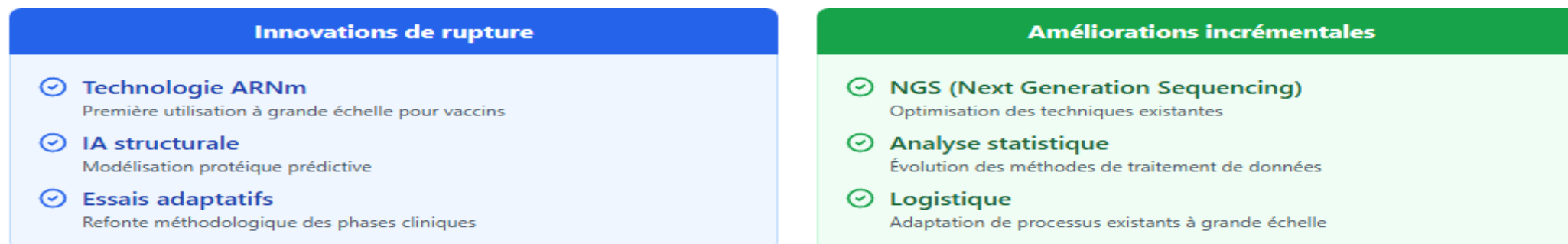
Timeline complète



Points de rupture



Classification des innovations



Synthèse Technologique

L'écosystème technologique construit en un an

Cartographie des outils et méthodes par phase

Phase 1: Découverte	Phase 2: Développement	Phase 3: Production & Distribution
> NGS ultrarapide > Bioinfo Cloud > Pangolin & Nextstrain	> IA structurale > Plateforme ARNm > Modélisation in silico	> Chaîne du froid -70°C > EDC adaptatifs > GIS / ERP intégrés

Classification par maturité

Technologies établies mais accélérées
1 NGS Séquençage nouvelle génération 2 ERP Planification ressources 3 GIS Systèmes d'information géographique
Technologies émergentes validées
1 EDC avancés Collecte électronique de données 2 Modèles bayésiens Analyse statistique adaptative 3 Cloud bioinformatique Analyses distribuées
Technologies de rupture développées
1 ARNm Plateforme vaccinale novatrice 2 IA structurale Prédiction conformationnelle 3 Essais adaptatifs Protocoles cliniques révolutionnés

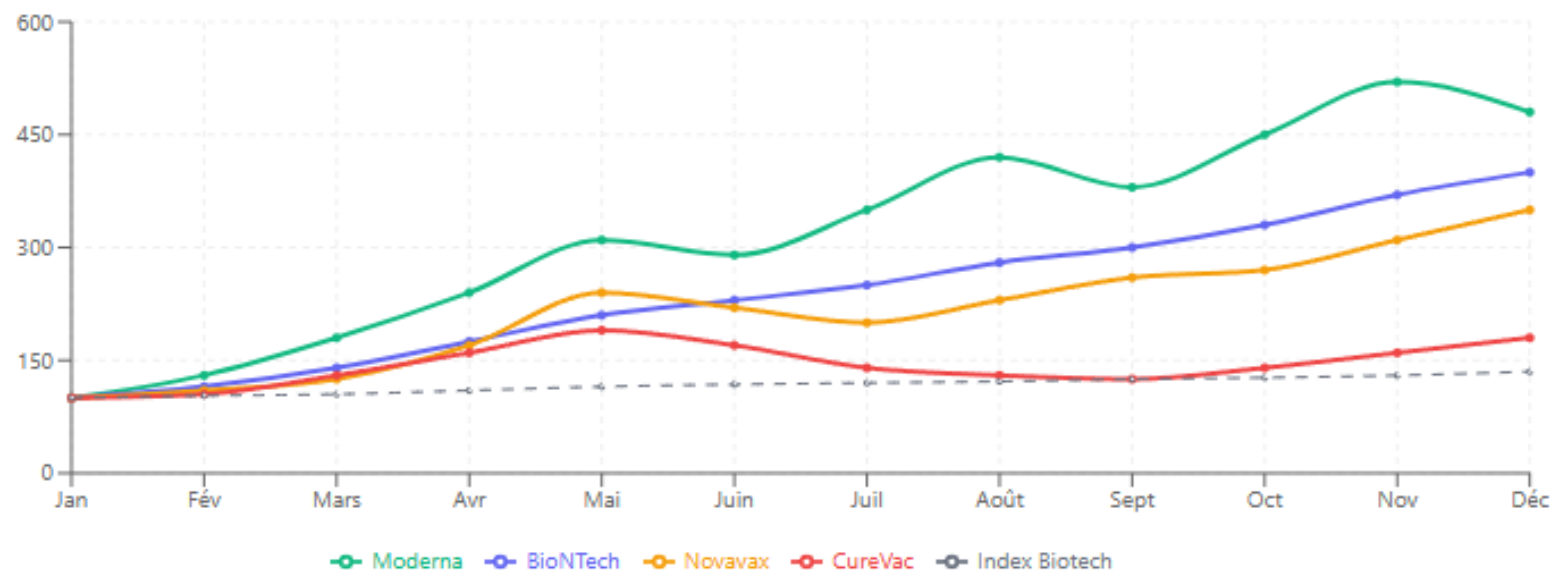
Potentiel de transfert

Haute valeur de transfert
🔗 Prédiction structurale Applications thérapeutiques multiples 🔗 Pipeline adaptatif Développement accéléré généralisable
Moyenne valeur de transfert
🕒 Monitoring temps réel Applications cliniques ciblées 🕒 Optimisation logistique Chaînes d'approvisionnement complexes
Valeur spécifique
🛡️ Formulation ARNm Applications très spécialisées 🛡️ Séquençage d'urgence Contextes pandémiques uniquement

Impact Financier Global

La plus grande création de valeur en biotech

Courbes comparées des principales biotechs (365 jours)



Chiffres clés

+200 Mrd \$

Capitalisation totale créée

32,6 Mrd \$

Investissements mobilisés (R&D, production)

+45%

Retours estimés à long terme (CAGR 5 ans)

Corrélation innovation/investissement



Pour chaque 1\$ investi en R&D avancée, le ratio de création de valeur atteint 6,2\$ en valorisation boursière.

Importance de l'innovation propriétaire



Les acteurs disposant de technologies propriétaires ont vu leur valorisation multipliée par 4-5x vs 1-2x pour les autres.

Valeur des plateformes vs produits



73% de la valorisation actuelle est attribuée aux plateformes technologiques et non aux produits spécifiques.



L'investissement en R&D et IT comme moteur de valorisation

Applications pour Ingénieurs EBI

Opportunités pour les ingénieurs en biologie industrielle

💡 Domaines d'application



Bioprocessing optimisé par IA

Utilisation d'algorithmes de deep learning pour optimiser les rendements et paramètres de bioréacteurs



Production d'ARNm pour applications industrielles

Adaptation de la technologie ARNm pour l'expression contrôlée d'enzymes et protéines industrielles



Traçabilité et chaîne d'approvisionnement

Systèmes intégrés pour le suivi en temps réel de produits biologiques sensibles avec contraintes strictes



Stabilisation de molécules biologiques

Techniques avancées de formulation et lyophilisation pour maintenir l'intégrité moléculaire

⚡ Compétences clés à développer



Bioinformatique appliquée

Maîtrise des outils d'analyse génomique et des pipelines d'analyse spécifiques au contexte industriel

Python

R

Biopython



Science des données et modélisation

Développement de modèles prédictifs et d'algorithmes d'optimisation pour les bioprocédés

TensorFlow

PyTorch

Scikit-learn



Interface biologie/informatique

Conception de systèmes intégrés combinant capteurs biologiques, IoT et traitement de données en temps réel

API REST

MQTT

Kafka



Optimisation de procédés par méthodes adaptatives

Application de techniques d'apprentissage par renforcement pour l'auto-ajustement des paramètres de production

Ray/RLlib

Bayesian Opt

DoE avancé