
**Biotechnologie — Séquençage
massivement parallèle —**

**Partie 2:
Évaluation de la qualité des données
de séquençage**

*Biotechnology — Massively parallel sequencing —
Part 2: Quality evaluation of sequencing data*
(<https://standards.iteh.ai>)
Document Preview

[ISO 20397-2:2021](https://standards.iteh.ai/catalog/standards/iso/db93fc69-ab5f-489f-b06c-7fb9e25b4762/iso-20397-2-2021)

<https://standards.iteh.ai/catalog/standards/iso/db93fc69-ab5f-489f-b06c-7fb9e25b4762/iso-20397-2-2021>



iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

[ISO 20397-2:2021](https://standards.iteh.ai/catalog/standards/iso/db93fc69-ab5f-489f-b06c-7fb9e25b4762/iso-20397-2-2021)

<https://standards.iteh.ai/catalog/standards/iso/db93fc69-ab5f-489f-b06c-7fb9e25b4762/iso-20397-2-2021>



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO 2021

Tous droits réservés. Sauf prescription différente ou nécessité dans le contexte de sa mise en œuvre, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie, ou la diffusion sur l'internet ou sur un intranet, sans autorisation écrite préalable. Une autorisation peut être demandée à l'ISO à l'adresse ci-après ou au comité membre de l'ISO dans le pays du demandeur.

ISO copyright office
Case postale 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Genève
Tél.: +41 22 749 01 11
E-mail: copyright@iso.org
Web: www.iso.org

Publié en Suisse

Sommaire

Page

Avant-propos.....	iv
Introduction.....	v
1 Domaine d'application	1
2 Références normatives	1
3 Termes et définitions	1
4 Données brutes	6
4.1 Généralités.....	6
4.2 Fichier de données brutes.....	7
4.3 Évaluation de la qualité des données brutes.....	7
4.3.1 Généralités.....	7
4.3.2 Statistiques élémentaires.....	7
4.3.3 Mesures de qualité.....	7
4.4 Prétraitement des données brutes.....	8
5 Alignement et cartographie des séquences	9
5.1 Généralités.....	9
5.2 Format des fichiers d'alignement et de cartographie.....	9
5.3 Contrôle qualité de l'alignement et de la cartographie des séquences.....	10
5.3.1 Statistiques sur les alignements de base.....	10
5.3.2 Indicateurs de qualité.....	11
5.3.3 Méthodes d'évaluation de la qualité d'alignement et de cartographie.....	12
5.4 Post-traitement de l'alignement.....	12
6 Détection de variants	12
6.1 Généralités.....	12
6.2 Fichier de données pour la détection de variants.....	12
6.3 Mesures de qualité lors de la détection de variants.....	12
6.4 Traitement des variants faux-positifs.....	13
6.5 Annotation de séquences.....	13
7 Validation	13
7.1 Généralités.....	13
7.2 Validation des mesures de qualité.....	14
8 Documentation	15
Annexe A (informative) Mesures de qualité applicables aux plateformes SMP	16
Annexe B (informative) Recommandations applicables à la couverture et aux lectures en fonction des applications	17
Annexe C (informative) Logiciel d'alignement et de cartographie des séquences	19
Bibliographie	20

Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est en général confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO participent également aux travaux. L'ISO collabore étroitement avec la Commission électrotechnique internationale (IEC) en ce qui concerne la normalisation électrotechnique.

Les procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/IEC, Partie 1. Il convient, en particulier, de prendre note des différents critères d'approbation requis pour les différents types de documents ISO. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/IEC, Partie 2 (voir www.iso.org/directives).

L'attention est attirée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO ne saurait être tenue pour responsable de ne pas avoir identifié de tels droits de propriété et averti de leur existence. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou dans la liste des déclarations de brevets reçues par l'ISO (voir www.iso.org/brevets).

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la nature volontaire des normes, la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'Organisation mondiale du commerce (OMC) concernant les obstacles techniques au commerce (OTC), voir www.iso.org/avant-propos.

Le présent document a été élaboré par le Comité technique ISO/TC 276, *Biotechnologie*.

Une liste de toutes les parties de la série ISO 20397 se trouve sur le site web de l'ISO.

Il convient que l'utilisateur adresse tout retour d'information ou toute question concernant le présent document à l'organisme national de normalisation de son pays. Une liste exhaustive desdits organismes se trouve à l'adresse www.iso.org/members.html.

Introduction

Le séquençage massivement parallèle (SMP) est une approche analytique de séquençage de l'acide nucléique à haut débit qui utilise un traitement massivement parallèle pour étudier des génomes entiers, des transcriptomes et des séquences ciblées d'acides nucléiques de différentes origines, en un laps de temps relativement court.

Le SMP est utilisé dans de nombreux domaines des sciences de la vie. Il permet une détermination et une analyse à haut débit de milliards de nucléotides. Du fait de la variabilité biologique des polymères d'acide désoxyribonucléique et d'acide ribonucléique à travers le vivant, la détermination précise de leurs séquences constitue un véritable défi. La qualité des séquences générées par SMP dépend de nombreux facteurs, notamment, entre autres, la qualité de l'échantillon, la préparation de la banque, le choix de la plateforme de lecture et la qualité des données de séquençage.

L'analyse des données de séquençage peut représenter de véritables défis bio-informatiques liés au stockage des données, au temps de calcul et à la précision de détection des variants. L'une des principales difficultés associées aux données de séquençage, trop souvent négligée, porte sur les mesures de contrôle qualité à tous les stades du pipeline de traitement des données, alors mêmes qu'ils sont essentiels à toute l'analyse en aval des données de séquences. Le contrôle qualité applicable au traitement et à l'analyse des données de séquençage de l'acide nucléique concerne trois niveaux distincts: données brutes, alignement et détection des variants. Le présent document fournit une liste d'éléments à prendre en compte lors de l'évaluation de la qualité des données de séquençage massivement parallèle, ainsi que les recommandations spécifiques à différentes plateformes SMP.

iTeh Standards (<https://standards.iteh.ai>) Document Preview

[ISO 20397-2:2021](https://standards.iteh.ai/catalog/standards/iso/db93fc69-ab5f-489f-b06c-7fb9e25b4762/iso-20397-2-2021)

<https://standards.iteh.ai/catalog/standards/iso/db93fc69-ab5f-489f-b06c-7fb9e25b4762/iso-20397-2-2021>