



HAL
open science

Programmer l'évolution : une faille dans la science

Giuseppe Longo

► **To cite this version:**

| Giuseppe Longo. Programmer l'évolution : une faille dans la science. 2021. hal-03319844

HAL Id: hal-03319844

<https://ens.hal.science/hal-03319844v1>

Submitted on 13 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Programmer l'évolution : une faille dans la science¹

Giuseppe Longo
CNRS – École Normale Supérieure, Paris
<https://www.di.ens.fr/~longo/>

Revue de l'ouvrage de Jennifer A. Doudna et Samuel H. Sternberg
"A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution"

Résumé

Jennifer Doudna, lauréate du prix Nobel, et Samuel Sternberg passent en revue les avancées récentes dans un domaine pionnier de la biologie moléculaire. Dans un style accessible et élégant, les auteurs présentent les succès et les défis d'une nouvelle technique de modification de l'ADN : CRISPR. Ils transmettent l'émotion de la découverte, la passion de la recherche et l'audace intellectuelle. Tout en admirant les compétences techniques des auteurs, qui comptent parmi les meilleurs chercheurs du domaine, cette revue souligne de manière critique les limites de leurs pratiques expérimentales, à savoir : un cadre théorique vague ou incomplet ; des cibles génétiques souvent inatteignables ; des effets hors cible ; des échecs antérieurs par d'autres formes de manipulation génétique, et, enfin, l'imprévisibilité intrinsèque de nombreuses conséquences phénotypiques d'une technique aussi puissante. En raison de ces préoccupations, l'approche des organismes et de l'évolution adoptée par les auteurs est remise en question dans le but de susciter un débat ouvert.

1 - Impression générale : vulgarisation et éthique.

Le livre examiné est un récit très efficace d'une aventure personnelle extraordinaire dans l'invention et l'utilisation des dernières techniques de manipulation génétique. Bien qu'il ait deux auteurs, il a été écrit à la première personne. Cela ajoute une touche personnelle à un style très lisible. En fait, on entrevoit la passion d'une chercheuse désintéressée et très compétente, plongée dans le monde difficile des techniques biochimiques. On saisit des moments de réussite joyeuse mais aussi de déception perplexe... En bref, ce livre exprime un bel esprit qui se consacre profondément au travail de laboratoire. L'auteur/narrateur prend le lecteur, même inexpérimenté, par la main dans un voyage difficile pour "découvrir" - ou plutôt inventer - le potentiel technique des mécanismes biologiques spécifiques à l'interaction entre les virus et les bactéries, puis l'étend à la manipulation de l'ADN dans les cellules eucaryotes. À cette fin, le livre contient des informations très intéressantes sur les virus et les bactéries, rendues accessibles à tous. Je ne commenterai pas davantage les nombreux détails fascinants illustrés, par exemple sur la façon dont les bactéries se défendent contre les virus, et comment les structures chimiques impliquées dans ce processus peuvent être reconstituées et utilisées dans les laboratoires grâce à la perspicacité et au travail...

L'ouvrage présente également les succès ainsi que de longues listes d'applications potentielles de la manipulation rendue possible par les nouvelles techniques d'édition de l'ADN : "les scientifiques peuvent désormais manipuler et modifier rationnellement le code génétique qui définit toutes les espèces de la planète, y compris la nôtre". Avant de discuter du cadre techno-scientifique proposé, passons directement à la dernière partie de l'ouvrage.

¹ En anglais, dans *Organisms. J. Biological Sciences*, July 2021.

Celle-ci aborde les questions éthiques liées au potentiel de la manipulation génétique chez l'homme - en particulier l'"amélioration" de l'espèce. Ici, malgré son enthousiasme pour les techniques auxquelles elle a contribué, l'auteure ne s'arrête pas au défi éthique posé par de telles manipulations. Avec une grande humanité et un souci intime, l'ouvrage présente les risques et dérives possibles de ces activités et propose des limites éthiques strictes aux manipulations chez l'homme. À cette fin, il nous fait vivre le drame de la violence possible faite à notre espèce, et l'on sent la sensibilité particulière d'une femme, l'auteure principale, confrontée à la manipulation du génome embryonnaire d'un futur enfant.

2 - Problèmes théoriques

Après avoir sincèrement apprécié les mérites du livre en termes d'écriture, de passion et d'éthique, nous passons maintenant à la critique de son contenu scientifique. Là encore, il faut saluer la grande honnêteté intellectuelle des auteurs. Sans hésiter, ils prennent le dogme central de la biologie moléculaire (Crick 1958, p. 11) comme pilier de leur cadre théorique. Aujourd'hui, ce n'est souvent pas le cas, même ceux qui s'appuient encore et de facto sur ce dogme, notamment dans la pratique de la biologie moléculaire en laboratoire, s'abstiennent le plus souvent de le mentionner. Si on leur demande, ils présentent souvent le dogme central comme une "figure de style" ou une "simplification" de la réalité. Nous saluons donc un choix courageux et précis qui ne nous laisse pas dans des cadres théoriques vagues et mal définis. Bien sûr, un problème se pose : comment interpréter ce dogme ? Bien que cela ne soit pas explicitement dit, il ne fait aucun doute que l'interprétation du livre fait référence à la version plus dure proposée par Watson dans les années 1960. Une telle version considère que l'ADN contient le codage *complet* de l'information génétique, donc la transmission héréditaire. On ne peut pas reprocher aux auteurs un peu de flou à cet égard, car la notion d'"(in)complétude", claire et précise en mathématiques, est inhabituelle dans les sciences naturelles. L'article fondateur de 1935 d'Einstein, Podolsky et Rosen (connu sous le nom de paradoxe EPR), qui traitait de l'"incomplétude" de la mécanique quantique et fournissait une critique très rigoureuse et constructive de ses fondements, constitue une exception. Tout porte à croire que Doudna et Sternberg considèrent systématiquement l'ADN comme complet dans son potentiel ontogénétique prescriptif. En conséquence, l'écriture dans les gènes contient l'ensemble complet des instructions, elle prescrit l'ontogénèse et est au cœur de la phylogénèse.

Cependant, un certain flou apparaît rapidement : les notions d'"information (génétique)" et de "programme" sont aussi omniprésentes qu'indéfinies. Puisqu'il s'agit d'informations codées sur des bases de données discrètes (la structure chimique de l'ADN), on nous fait croire qu'il s'agit d'informations de Shannon (transmission) et/ou de Turing-Kolmogorov (traitement). Comme il est typique de la biologie, cela manque de toute référence précise à d'autres notions d'information. N'entrons pas ici dans la diversité de ces deux notions. Pour de bonnes raisons, celles-ci traitent de manière duale de la relation entre les notions d'entropie et de complexité, donc d'information qui est habituellement considérée comme négentropie (Longo 2019). En fait, l'absence de corrélation entre la "complexité" d'un organisme, quelle que soit sa définition, et son ADN, ne semble pas préoccuper les partisans de l'approche géocentrique. Bien que les auteurs considèrent l'ADN comme un codage complet de l'organisme, ils rappellent que, par exemple, le génome est des centaines de fois plus grand chez les plantes que chez les humains. Notons qu'en 1999, le directeur du projet du génome humain, Francis Collins, écrivait qu'il s'attendait à trouver 80 000 gènes chez l'homme, considérant, non sans fierté, que le beaucoup moins complexe *C. elegans* (un

ver microscopique de 1 000 cellules) en comptait 16 000. Deux ans plus tard, il reconnaissait qu'il semblait y avoir 25 000 gènes chez l'homme, ou, comme il l'a affirmé plus tard avec d'autres auteurs, 21 000. La notion de programme génétique est encore plus vague.

Lorsqu'une tentative a été faite par quelques biologistes utilisant le langage le plus adéquat pour la manipulation des chaînes de caractères et la réécriture des termes ou "term-editing" (le lambda-calcul de Church, qui est ma spécialité depuis longtemps (Barendregt 1984 ; Kreisel 1982)), l'utilisation de la récursion était encore abusive (voir Longo 2018 ; 2019 pour une critique et les sources). En somme, la biologie moléculaire main stream a tendance à se référer de manière floue à des notions précises telles que l'information et le programme, alors que ces notions sont mathématiquement engagées dans des formes fortes et spécifiques de traitement de l'information ainsi que de "détermination" (ce qui et comment détermine quoi). Cela filtre implicitement dans les vues, les expériences et l'interprétation des mesures. C'est plutôt insuffisant pour un texte aussi riche en descriptions rigoureuses de virus et de bactéries, qui vise une présentation globale, et qui appelle une définition claire de termes si libéralement utilisés dans la discipline (y compris la notion fondatrice de gène).

En fait, qu'est-ce qu'un gène ? Dans son livre, *The Century of the Gene*, Evelyn Fox-Keller note que la notion de gène a changé cinq fois au cours du 20^e siècle. En fait, le "gène" n'est pas défini dans le livre de Doudna et Sternberg. Cependant, le lecteur est amené à penser qu'ils considèrent qu'il s'agit d'un segment d'ADN à associer non seulement à une protéine mais aussi à un phénotype. Ceci est en contradiction avec leur reconnaissance du fait que certains phénotypes sont le résultat d'un réseau d'expression génétique, comme c'est le cas suite aussi à des phénomènes identifiés de longue date tels que "l'épissage alternatif" (Leff et al. 1986 ; voir aussi Brett et al. 2001 ; Nilsen & Graveley 2010). Ces initiations alternatives de la transcription et de la traduction (de Klerk & 'tHoen 2015) appellent à une révision de la vision "dogmatique" de la correspondance d'un ARNm à une protéine chez les eucaryotes (Mouilleron et al. 2016). Cette complexité supplémentaire va au-delà du concept de réseaux dans la relation génotype-phénotype (Brunet et al. 2018 ; 2020). Un exemple particulièrement parlant concerne les "gènes superposés". Ce phénomène a été découvert dans les années 1970 grâce au tout premier séquençage d'un génome d'ADN (Barrell et al. 1976) et a été négligé depuis. Aujourd'hui encore, certains chercheurs (Schlub et Holmes 2020) le considèrent comme une caractéristique typique des virus, tandis que beaucoup commencent à le reconnaître comme une caractéristique très pertinente parmi la catégorie générale des "protéines alternatives" dans les organismes cellulaires (Mouilleron et al. 2016 ; Brunet et al. 2018 ; Pavesi et al. 2018 ; Meydan et al. 2019). Globalement, il est clair que ces phénomènes réfutent l'idée que les gènes sont des segments d'ADN avec un début et une fin précis, comme des instructions conçues pour un logiciel. En effet, le projet ENCODE a déjà mis en évidence "les modèles complexes de régulation dispersée et de transcription omniprésente" et a proposé de définir un gène comme "une union de séquences génomiques codant pour un ensemble cohérent de produits fonctionnels potentiellement superposés." Pourtant, les chercheurs impliqués sont conscients que leur "définition contourne les complexités de la régulation et de la transcription en supprimant complètement la première [la régulation] de la définition" (Gerstein et al. 2007).

En résumé, la signification exacte non seulement de "information" et de "programme" mais aussi de "gène" n'est pas claire. Souvent, le flou de ces notions laisse place à l'attribution d'un pouvoir extraordinaire aux "gènes". Tout est dans l'information génétique et élaboré par le programme génétique. Le programme comme l'information sont entièrement inscrits dans

les gènes. Certes, les auteurs soulignent que "chez un individu, toutes les cellules somatiques ont le même ADN." Cependant, la contribution du contexte dans le contrôle de l'expression des gènes n'est jamais évoquée - peut-être parce que le mentionner remettrait en cause le rôle moteur de l'ADN dans la détermination du phénotype. Par conséquent, on suppose qu'un programme très détaillé contrôle l'expression génétique dans l'ADN lui-même, du zygote à l'adulte. Cela signifie également que l'être humain est inscrit principalement dans les 5 000 gènes en plus de ceux du *C. Elegans*, qui contribuent de manière causale à ce que chaque cellule prenne des formes et des fonctions très différentes, des cellules cardiaques aux neurones et aux cellules hépatiques. L'édition de ce programme permettrait à l'organisme d'être complètement piloté dans l'écosystème par la volonté rationnelle de l'homme, ce qui est éthiquement acceptable et même nécessaire, selon les auteurs, au moins chez les plantes et les animaux.

Une autre lacune théorique du livre est l'utilisation implicite d'une autre propriété essentielle au déterminisme géocentrique proposé : la stéréospécificité exacte des interactions macromoléculaires et, par conséquent, de toutes les cascades allant de l'ADN aux fonctions des protéines jusqu'aux phénotypes. Monod, dans son livre de 1970 *Le hasard et la nécessité*, reconnaît avec une grande cohérence intellectuelle que cette propriété est "nécessaire à la transmission de l'information." Plus fort encore, Monod affirme que "la cellule est un mécanisme cartésien", une chaîne d'engrenages et de poulies comme dans une montre. La stéréospécificité macromoléculaire dans une cellule, aussi exacte que les "algèbres booléennes ... dans nos ordinateurs", dit-il, nous fait comprendre comment le traitement et la transmission de l'information génétique contenue dans l'ADN peuvent fonctionner. Le premier problème qui découle d'un tel postulat est que la physico-chimie traite depuis longtemps les interactions entre macromolécules de manière statistique. Les interactions moléculaires dans une cellule ne font pas exception, comme cela a été constaté pour l'expression génétique dès 1983 par Kupiec (1983), voir aussi (Kupiec 2010). Depuis, la stochasticité de toutes les étapes de l'expression génétique, de la transcription à la traduction en passant par l'épissage alternatif, a été largement confirmée (voir Elowitz 2002 ; Paldi 2003 ; Raj & Oudenaarden 2008 ; Waks et al. 2011 et plus récemment Boersma et al. 2019).

D'une manière générale, les interactions macromoléculaires sont stochastiques, elles doivent être données en probabilités, et ces probabilités dépendent du contexte. Il existe de nombreuses références qui justifient ce principe théorique fort, mais elles sont négligées par le géocentrisme dominant. En fait, le tableau change complètement si l'on considère que, dans cet esprit, presque chaque "gène" est transcrit dans presque chaque cellule. Chelly et al. (1989) l'ont souligné il y a longtemps et cela a été largement confirmé depuis : c'est une question de probabilités différentes (voir aussi les références ci-dessus sur la stochasticité). De plus, les torsions et les pressions exercées sur la chromatine modifient les sites d'accès à l'ADN, altérant ainsi son expression (Cortini et al. 2016). Il s'agit certainement d'une question cruciale dans l'embryogenèse, même si elle ne s'applique guère aux ordinateurs. De même, beaucoup mettent en avant la "diversité cellulaire non génétique" et "le rôle de la structure des réseaux de régulation et du bruit moléculaire" (Balazsi et al. 2011). Comme le souligne (Braun 2015) : "Le génome ne détermine pas l'état cellulaire ordonné. Il participe plutôt à ce processus en fournissant un ensemble de contraintes sur le spectre des modes de régulation, qui sont analogues aux conditions limites dans les systèmes dynamiques physiques." Il s'agit clairement d'un changement de perspective radical par rapport à l'approche géocentrique : dans ce cadre, les "conditions limites" et leurs modifications, bien que toujours pertinentes

pour la dynamique, nécessitent un type d'analyse différent. En général, aucune composante de la dynamique n'est "complète". De plus, en physique, une différence dans les conditions aux limites peut induire une différence dans la dynamique ou dans son résultat. Cependant, les conditions aux limites sont analysées différemment des "causes" de la dynamique elle-même. En d'autres termes, ces dernières sont clairement (mathématiquement) distinctes des conditions limites et sont généralement et magnifiquement formulées en termes de lois de conservation ou de symétries, de sorte que la notion de cause peut même être évitée (une pierre tombe pour des raisons de symétrie selon la théorie de la relativité).

En physique, cependant, les conditions limites sont censées être préétablies par rapport au processus prévu. En biologie, au contraire, ces "conditions limites" sont des contraintes co-construites. Elles dépendent également du processus contraint qui les produit : même l'ADN, cette trace physico-chimique fondamentale de l'histoire, subit une reconstruction constante. C'est une contrainte massive pour la dynamique et la construction des macromolécules. Il se modifie dynamiquement et s'applique de façon différentielle au cours de l'ontogenèse, ainsi que, de façon spectaculaire, au cours de l'embryogenèse. Plus généralement, les processus moléculaires, cellulaires et organisationnels reconstruisent continuellement les membranes, les microtubules et les autres composants cellulaires, ainsi que toutes les parties fonctionnelles de l'organisme. Ces derniers constituent des contraintes qui contribuent à la dynamique biologique à tous les niveaux d'organisation. Si tel est le cas, ils affectent également les nombreux réseaux macromoléculaires qui, bien que hautement improbables du point de vue de la physique, existent et fonctionnent, mais uniquement dans les cellules vivantes, avec une histoire. La notion originale de "fermeture de contraintes" de Montévil & Mossio (2015) introduit élégamment l'approche suggérée ici. Bien sûr, la modification de l'une de ces contraintes, en particulier une contrainte aussi importante que l'ADN, entraîne un changement. Cependant, c'est parce que le changement des contraintes s'avère re-canaliser les processus macromoléculaires, qui, en soi, sont au moins non linéaires ou, plus généralement, stochastiques.

Bien entendu, cette analyse s'écarte du déterminisme de Doudna et Sterner fondé sur le programme génétique, le dogme central et l'idée (malheureusement implicite) que les interactions stéréospécifiques macromoléculaires sont exactes. Ces hypothèses théoriques ne sont pas des simplifications par souci de vulgarisation. Au contraire, elles sont au cœur de la perspective du livre. Ces fondements fragiles sapent tout l'édifice conceptuel du géocentrisme strict, qui a été présenté au lecteur comme la seule façon de penser. L'approche théorique différente que nous suivons ici, telle que proposée par beaucoup d'auteurs et discutée par Soto et al. (2016), offre une autre perspective lors de l'analyse des preuves et des promesses faites dans le livre quant au rôle que CRISPR peut jouer dans la "reprogrammation" du vivant.

3 - Théories et preuves empiriques

En science, comme l'a observé Boltzmann, il n'y a rien de plus pratique qu'une bonne théorie. Les preuves empiriques peuvent-elles falsifier l'approche géocentrique du livre ? Je pense que oui, mais ce n'est pas si évident. Longo et Mossio (2020) présentent une analogie étroite entre la vision géocentrique et la perspective géocentrique, ptolémaïque, du système planétaire. Ils reconnaissent notamment l'extraordinaire progrès dans la connaissance du ciel dû à la grande astronomie et aux mathématiques islamiques du 8e au 14e siècle. Les astronomes de langue arabe ont décrit tous les corps célestes visibles et leurs mouvements,

notamment le système planétaire, dans une perspective géocentrique. Aucune preuve empirique ne pouvait falsifier leur récit des mouvements des planètes puisque, mathématiquement, tout nombre fini de points d'une ellipse autour du Soleil peut être interpolé par suffisamment d'épicycles centrés sur la Terre. Un changement de perspective, en fait, un changement métaphysique, était nécessaire afin de considérer les planètes du point de vue du Soleil. Seul un changement spectaculaire des principes théoriques pouvait alors falsifier la perspective géocentrique, comme l'invention du premier principe fondamental de conservation de la physique, à savoir l'inertie, par Galilée. Les "mouvements rétrogrades" des planètes, si bien décrits en arabe, sont alors devenus totalement impossibles en l'absence de masses dans tous les centres des épicycles, notamment après les travaux de Newton. Notez que l'inertie est un principe limite. Il ne s'applique jamais en pratique puisque les mouvements visibles sont toujours contraints par les gravitations et les frictions. Cependant, elle permet de comprendre tous les mouvements physiques en même temps et d'analyser ce qui les contraint : les gravitations et les frictions - depuis Galilée. En un sens, le mouvement inertiel est un état "par défaut" de la matière inerte. Ci-dessous, nous ferons référence à un "état par défaut" propre aux organismes vivants, en suivant Soto et al. (2016).

La pertinence du changement de perspective et de l'invention d'un "principe de conservation", l'inertie, est devenue évidente lorsque le nouveau cadre théorique a permis d'unifier la chute des pommes et les mouvements planétaires (Newton, Hamilton), évitant ainsi les descriptions ad hoc et les épicycles sur les épicycles. Cela rappelle l'écriture alphabétique ad hoc du programme d'ADN du zygote qui est censé permettre à chaque cellule de se différencier en neurone ou en leucocyte *parce que* les gènes contrôlent l'expression des gènes, l'un sur l'autre et ils ont la *propriété descriptive*, intrinsèque. Dans un contexte dominé par Monod, Jacob et Lwoff, la découverte du contrôle épigénétique de l'expression génétique par Barbara McClintock n'a pas été citée au cours de 20 ans ou plus (Fox-Keller 2003). Bien sûr, certains épicycles existent, par exemple, les satellites stationnaires autour de la Lune ou les satellites des planètes par rapport au Soleil.

En référence au livre examiné, les auteurs expliquent en outre que l'écriture alphabétique des gènes, avec son contrôle total de l'ontogenèse, est "aussi modifiable qu'un simple morceau de texte." Par conséquent, le destin de l'embryon peut être programmé selon notre volonté rationnelle, du moins pour de très nombreux traits. Nous pouvons "imaginer que le génome humain est un grand logiciel". En tant que lecteur du livre, je comprends l'enthousiasme d'un bio-chimiste talentueux qui voit soudain, en laboratoire, l'explosion de son pouvoir combinatoire sur les séquences de bases d'ADN. Pourtant, en tant que théoricien, je suis radicalement en désaccord avec la perte du sens de la vie organique dans un contexte historique qu'une telle position transmet au lecteur. Existe-t-il des preuves empiriques confirmant au moins certaines réalisations effectives de la perspective de la programmation géocentrique ? Oui, et les auteurs fournissent de longues listes de résultats et des listes encore plus longues d'applications futures et potentielles. Quel est alors le problème, au moins avec les résultats ? En fait, il y en a plusieurs.

Tout d'abord, comme dans le cas des astronomes islamiques, certaines applications peuvent fonctionner et s'avérer très utiles. Nous devons à Ibn Yunus (Égypte, vers 1000 après J.-C.) et à de nombreux autres grands scientifiques islamiques des avancées majeures dans le domaine de la trigonométrie sphérique et des observations célestes qui ont abouti aux Tables Alfonsines (Espagne catholique, 1483), qui ont été utilisées avec succès et à grande échelle pour la navigation. Cependant, généraliser leur point de vue et leurs promesses serait une

erreur majeure, sans parler de leurs prédictions mêlées à l'astrologie. Aujourd'hui, elles sont comparables à celles de (Plomin 2019), où l'on prétend également que le comportement humain est inscrit dans l'ADN d'un nouveau-né.

De plus, il faut considérer que les observations et les expériences en biologie (moléculaire) souffrent de la plus grave crise d'irreproductibilité (Begley & Ionidis 2014). En effet, la biologie est le lieu théorique de la diversité, de la variabilité et de la spécificité historique des organismes, qui résultent d'une histoire phylo- et onto-génétique. Cela signifie que l'on ne peut pas (facilement) généraliser des cas individuels (ou pas de la même manière qu'en physique, (Montévil, 2019)). En tant que discipline, la biologie moléculaire subit une forte pression pour " publier ou périr ", ce qui est désastreux pour la perspicacité et l'intégrité scientifique critique et chronophage (Longo 2014) et produit des résultats dont la validité temporelle est la plus courte (della Briotta et al. 2015).

Deuxièmement, "la mesure en biologie est méthodisée par la théorie", comme analysé de près dans (Montévil 2019). L'arrière-plan théorique flou de l'information et de la programmation contribue à rendre les résultats et les données trop souvent peu fiables ou incertains lorsqu'il s'agit de les interpréter. Avant la théorie de Newton, les astronomes avaient rencontré de gros problèmes avec les données sur les orbites képlériennes des planètes dont les irrégularités étaient dues aux interactions gravitationnelles planétaires. Jusqu'à la théorie d'Einstein, les mesures du périhélie de Mercure étaient incompréhensibles. Les données ne parlent pas d'elles-mêmes, et encore moins les très grands ensembles de données, car elles contiennent nécessairement beaucoup de corrélations fallacieuses (Calude&Longo 2017). En ce qui concerne notre objet d'étude, la "réécriture" de l'ADN peut non seulement atteindre son but et modifier le phénotype prévu, mais aussi brouiller d'autres parties de l'ADN et ainsi affecter l'organisme. Les changements induits par CRISPR dans l'ADN peuvent être plus nombreux que ceux attendus. Les différents changements génétiques peuvent être dus à la diversité des modifications des nucléotides dans la séquence cible, ainsi qu'à un spectre variable de sites qui ont été modifiés. Comme les effets indésirables peuvent provenir à la fois des sites cibles et des sites hors cible, la détection et la mesure des modifications non intentionnelles ou hors cible peuvent être beaucoup plus difficiles que celles des modifications des sites cibles. En fait, il s'avère que c'est le cas (Chaudhari et al. 2020 ; Höijer et al. 2020, Modrzejewskiet al. 2020) car le nombre et l'emplacement des changements de nucléotides sont inconnus, en particulier s'ils se produisent avec des probabilités plus faibles mais non nulles dans des sites non spécifiques. De plus, les changements peuvent ne pas dépendre de la séquence d'acide nucléique modifiée. Ils peuvent plutôt dépendre de l'échelle de la modification induite (par exemple, le niveau de l'organisme ou de l'écosystème), ainsi que de son moment et de sa durée (temporaire ou permanente) (Adikusuma et al. 2018). Les théories informationnelles de l'édition exacte macromoléculaire des codes alphabétiques ne permettent pas de voir ces phénomènes ni de les interpréter.

Les observations critiques augmentent avec le temps, y compris les remarques sur la faible efficacité de la réparation des mutations, les taux élevés de mosaïcisme et la possibilité de résultats d'édition involontaires pouvant avoir des conséquences pathologiques (National Academies of Sciences, 2020 ; Alanis-Lobato et al. 2021 et références 10-14 dans ce document). Récemment, Leibowitz et al. (2021) ont montré que "l'édition CRISPR-Cas9 génère des défauts structuraux du noyau, des micronoyaux et des ponts chromosomiques, qui initient un processus mutationnel appelé chromothripsie. La chromothripsie est un réarrangement chromosomique extensif limité à un ou quelques chromosomes qui peut causer

des maladies congénitales humaines et des cancers. Ces résultats démontrent que la chromothripsie est une conséquence sur cible jusqu'alors non appréciée des DSB générées par CRISPR-Cas9."

Nous sommes loin de l'affirmation des auteurs concernant "la capacité remarquable de réécrire le code de la vie avec une précision chirurgicale et une simplicité étonnante". En effet, les techniques inventées par les auteurs et leurs collaborateurs modifient l'ADN, peuvent guider la production d'une molécule fonctionnelle spécifique, et induire, entre autres, un "gain de fonction" au niveau cellulaire. Cependant, leurs effets hors cible ou sur cible non appréciés, et leurs conséquences enchevêtrées et non-compositionnelles sur les différents niveaux d'organisation d'un organisme - qui sont intégrés dans un écosystème - sont loin d'être sous contrôle. Nous pouvons fortement influencer l'évolution, mais pas la contrôler. En fait, nous pouvons réussir à modifier quelques contraintes des processus complexes, mais nous ne parvenons jamais à les contrôler totalement. Nous pouvons certainement agir sur la nature, mais avec prudence.

En résumé, la technologie CRISPR modifie bien l'ADN, mais où, et avec quelles conséquences dans le temps ? La croyance que nous pouvons couper précisément les interactions macromoléculaires est une illusion qui appartient au mythe de la cellule comme "mécanisme cartésien" avec des ordinateurs et des logiciels remplaçant les horloges de Descartes. Par conséquent, la question clé consiste à passer d'une perspective géocentrique à une vision centrée sur l'organisme dans sa relation avec l'écosystème, où l'ADN représente une contrainte interne et historique fondamentale, au sens de (Montévil & Mossio 2015). Je crois et j'espère que la remarquable invention technique de CRISPR pourra être utilisée de manière judicieuse pour la connaissance et les thérapies, au moins pour les maladies monogénétiques rares. La plupart des pathologies, cependant, même lorsque l'ADN joue un rôle clé, sont dues à la déformation d'un large réseau d'expressions génétiques et d'activités moléculaires qui interagissent dans un contexte organisationnel et écosystémique.

4 - Précédents

Les attentes exubérantes de CRISPR ont des précédents importants dans la vision géocentrique dominante. Revisiter quelques-uns d'entre eux peut aider à comprendre les limites des promesses actuelles. Sur la base de mon expérience personnelle indirecte, je me référerai aux géothérapies du cancer, attendues depuis environ un siècle et promises depuis au moins 50 ans, comme l'âge de la théorie de la mutation somatique du cancer (SMT). Cette théorie considère le cancer comme un problème entièrement génétique et compte explicitement sur CRISPR pour le résoudre. Depuis 1971, des projets généreusement financés annoncent la victoire finale contre le cancer grâce à des thérapies génétiques capables de "reprogrammer" l'ADN déprogrammé". La "guerre contre le cancer" de l'ancien président américain Richard Nixon visait à fournir ces thérapies pour 1976, année du bicentenaire de la Révolution américaine. En 2000, l'avancée technologique majeure que constitue le "décodage" du génome humain est considérée comme un outil supplémentaire pour résoudre l'énigme du cancer et, une fois de plus, permettre les thérapies génétiques. Hanahan & Weinberg (2000), avec plus de 20 000 citations en quelques années, et de nombreux autres auteurs, ont promis des thérapies génétiques pour "éliminer la souffrance et la mort dues au cancer d'ici 2015", selon les termes du directeur de l'Institut national du cancer de l'époque, Andrew von Eschenbach (2003). En effet, d'ici quelques années, les analyses d'ADN devraient avoir permis de poser un diagnostic et d'établir un pronostic.

Beaucoup d'entre nous, malheureusement, ont eu une expérience directe ou indirecte de cette maladie qui menace nos vies. Nous savons donc qu'en 2021, seul l'histologiste au microscope optique pourra reconnaître si un cancer est primaire, métastatique, bénin ou malin. En outre, il n'existe aucune thérapie plausible du cancer basée sur les gènes (voir Baker 2014 ; Huang 2014 ; Maeda & Katami 2018). Finalement, Weinberg (2014), dans une autocritique sévère de son approche précédente (voir l'article de 2000 ci-dessus avec Hanahan), reconnaît que "le séquençage du génome est également arrivé à maturité et a documenté des myriades de mutations affectant les génomes des cellules cancéreuses individuelles." De plus, "63 à 69 % de toutes les mutations somatiques [ne sont] pas détectables dans toutes les régions tumorales... Des signatures d'expression génétique de bon et de mauvais pronostic ont été détectées dans différentes régions d'une même tumeur" (Gerlinger et al. 2012). "Le séquençage a révélé que les cellules saines de tous les tissus portent de lourdes charges mutationnelles et que les mutations ne sont pas exceptionnelles, mais normales" (Mustjoki, Young 2021). Versteeg (2015) mentionne également des tumeurs sans mutations, tandis que Gatenby (2017) observe que "les cellules cancéreuses peuvent présenter un état apparemment paradoxal dans lequel leur charge mutationnelle est similaire et peut-être même inférieure à celle des cellules normales adjacentes." Sur cette base, Gatenby émet l'hypothèse que le tissu et l'environnement de l'organisme pilotent le processus, en suivant (Sonnenschein & Soto 1999). De plus, comme ose l'admettre Weinberg (2014), "la plupart des cancérigènes humains ne sont en fait pas mutagènes." Quarante ans d'analyses contradictoires sur l'amiante (Huang et al. 2011), plus les preuves susmentionnées, ont ouvert à l'idée que, lorsque la charge mutationnelle fréquente et lourde dans le cancer se produit, elle est le plus souvent une conséquence plutôt qu'une cause de la perturbation du contrôle cellulaire de la reproduction (see also (Mally, Chipman 2002). Ce phénomène apporte une diversité spécifique au cancer et conduit à le considérer comme un problème systémique (Bizzarri 2014; Baker 2021). Enfin, face à la confusion et profusion mutationnelle dans le cancer, (Weinberg 2014) parle de "complexité infinie", ainsi certains parient désormais sur le Big Data pour que les machines réparent l'échec humain dans la compréhension de l'étiologie du cancer. Malheureusement, les mathématiques montrent que c'est un non-sens (Calude & Longo 2017 ; Montévil & Longo 2018). Malgré cet échec, trop nombreux sont ceux qui – évitant le plus souvent toute référence explicite au dogme central, voire niant son rôle dans des conversations privées – continuent de faire ou de financer des recherches uniquement sur les mutations cancérigènes, les oncogènes, les proto-oncogènes ou les onco-suppresseurs (Kato et al. 2016 ; Rohan et al. 2018).

Avec une perspective organisationnelle plus robuste, la théorie du champ d'organisation tissulaire (TOFT) (Sonnenschein & Soto 1999) nous permet de comprendre pourquoi les cellules mutées d'un tissu cancéreux peuvent se normaliser fonctionnellement lorsqu'elles sont transférées dans un tissu sain. Par exemple, les cellules d'un néoplasme mammaire relocalisées dans un stroma de glande mammaire sain, se normalisent fonctionnellement (Maffini et al. 2005 ; Soto & Sonnenschein 2011). La TOFT se concentre sur la défaillance de la relation triangulaire tissu/organisme/environnement dans la formation du cancer. Elle souligne également le rôle des perturbateurs endocriniens et d'autres causes écosystémiques qui affectent le contrôle tissulaire et organisationnel de la reproduction des cellules somatiques. Au contraire, l'accent mis sur l'ADN dans l'étude et le traitement du cancer détourne l'attention et la recherche des causes environnementales, qui sont rarement mentionnées par les tenants de la SMT. En ce sens, l'environnement n'est pas mentionné une

seule fois dans ce livre, malgré une centaine de références au "cancer". En fait, la recherche d'une "solution miracle" génétique a dominé financièrement pendant 50 ans. Cela a largement exclu les autres voies de recherche et minimisé les analyses environnementales.

5 - Remarques sur la méthode

Certains pourront observer que j'ai mentionné le manque fréquent de fiabilité ou d'irreproductibilité des résultats expérimentaux dans la perspective que je critique, alors que j'ai attribué plus de validité aux preuves empiriques qui s'accumulent sur mon point de vue. Cela dépend d'analyses théoriques explicites. En effet, j'ai souligné dans plusieurs écrits, souvent en collaboration avec des biologistes, l'incohérence ou l'incomplétude du déterminisme géocentrique. Ces lacunes théoriques résultent de notions vagues ou incohérentes du gène, de l'information et du programme (voir (Longo 2019) pour une synthèse sur l'utilisation abusive de "information" et "programme"), ainsi que de leur causalité ou déterminisme implicite. Nonobstant, les expériences sont conçues sur la base de ces notions vagues ou implicites et de leurs fortes conséquences. Parmi celles-ci figure l'idée (jugée "nécessaire" par Monod en 1970) d'interactions macromoléculaires exactes au cœur d'immenses réseaux macromoléculaires, disions-nous, qui seraient conçus comme des circuits électroniques et élaboreraient des "algèbres booléennes" (et cela ne veut même pas dire "en moyenne"). De telles idées sont diffusées dans les manuels universitaires de biologie moléculaire et façonnent les esprits pour toujours. Cela m'a conduit à émettre davantage de doutes a priori sur les expériences et les mesures effectuées dans le cadre du géocentrisme informationnel. En effet, comme le soulignait Einstein en physique, la théorie décide des observables et des paramètres pertinents. Elle propose des outils et des méthodes de mesure, ainsi que des interprétations des données, comme mentionné plus haut à propos des orbites des planètes : les théories vagues ou incohérentes sapent les mesures, les méthodes et les interprétations.

Or, la biologie souffre encore plus de ces biais car la spécificité historique et contextuelle des organismes nécessite des connaissances et des mesures à la fois diachroniques et synchroniques, comme nous l'avons mentionné plus haut - voir aussi (Longo 2017 ; Montévil 2019 ; Montévil & Mossio 2020). En conséquence, un cadre théorique plus explicite, bien défini et robuste justifie une plus grande confiance dans les résultats empiriques. Par exemple, bien qu'elles se soient divisées en au moins deux approches différentes (Gould 2002), les théories de l'évolution font aujourd'hui un usage fantastique des empreintes génétiques en paléontologie. Souvent, il s'agit de l'ADN mitochondrial, qui permet de reconstruire les chemins phylogénétiques dans des perspectives théoriquement bien construites. Dans le cas du cancer, après 50 ans d'échecs des promesses de thérapies génétiques basées sur la SMT, la TOFT a été explicitement basée sur le premier principe de Darwin (l'hérédité comme "descendance avec modification"), interprétée comme un "état par défaut" ("reproduction avec variation") pour tous les organismes suffisamment nourris, et appliquée aux cellules somatiques sous des contraintes différentielles massives (contraignant la reproduction ainsi que la motilité de manière variable, selon le contexte). Cela semble plus convaincant que les principes de la SMT, indépendamment des échecs empiriques de cette dernière. En fait, la SMT fait implicitement référence au Dogme Central et à son ensemble de notions biologiquement floues d'information et de programme. La TOFT fait référence à Darwin et, aujourd'hui, à une théorie de plus en plus robuste d'une "fermeture des contraintes" en biologie. Son cadre théorique ne dépend plus des théories de l'information ou de la

programmation de Shannon, ni de celles de Turing, Church et Gödel (voir (Longo 2018) pour une critique de la " Gödelite" affectant certains biologistes). La TOFT permet une compréhension pertinente des perturbateurs endocriniens en tant que cancérigènes (Sweeney et al. 2015 ; Paulose et al. 2015) et outils de prévention, ouvrant ainsi la voie à de nouveaux paradigmes thérapeutiques (Baker 2014 ; Bizzarri et al. 2014 ; Proietti et al. 2019), tels que la réversion tumorale.

Deuxièmement, je considère que les "résultats négatifs" sont particulièrement intéressants en science, car ils ont toujours ouvert la voie à de nouveaux chemins de construction de la connaissance (Longo 2018). Au niveau théorique, l'aléatoire, en particulier, est subtilement lié à l'indécidabilité, s'il est compris comme une imprévisibilité dans la théorie visée (Calude & Longo 2016). S'il est bien défini, il fournit ainsi une limite précise à la connaissance. Or, la construction de l'indécidabilité est le résultat "négatif", qui est l'origine et le pilier de la théorie de la calculabilité ou "élaboration de l'information" (Gödel, Church & Turing dans les années 1930), si souvent citée dans la biologie moléculaire dominante. Il convient de noter qu'en biologie, le caractère aléatoire n'est pas un "bruit" (Bravi & Longo 2015 ; Calude & Longo 2016). Il s'agit plutôt d'une composante essentielle de la production de la variabilité et de la diversité, et donc de l'adaptabilité et de la stabilité des organismes et des écosystèmes (un bug typique de la "théorie de l'information" en biologie est qu'elle ne peut pas distinguer l'aléatoire du bruit – sauf par une notion de "incompressibilité", insensée en génétique). En d'autres termes, si l'on peut "faire quelque chose" ou mieux comprendre grâce à un aperçu des limites de la connaissance, telles que l'imprévisibilité (le caractère aléatoire), je considère cela comme un progrès théorique majeur. J'insiste sur le fait que les notions de programme et de calcul, qui ont changé le monde, ont été définies dans les années 1930 pour démontrer l'incalculabilité et l'indécidabilité – pour démontrer qu'il n'y a pas un algorithme, un programme, il faut dire précisément ce qu'on entend par algorithme, par programme. Bref, il s'agissait de clarifier les limites de la connaissance et de la praxis au lieu de prétendre à la complétude théorique de l'analyse de telle ou telle composante concourant à un processus, voire d'en démontrer l'imprévisibilité. Une telle méthode est donc fondamentale pour renforcer le cadre de la connaissance et pour s'ouvrir à de nouvelles théories et applications – et a permis de donner une notion robuste de aléatoire. Pour ces raisons, la reconnaissance de la stochasticité de l'expression génétique et des interactions macromoléculaires, canalisées par les contraintes biologiques, est une voie méthodologique convaincante, qui s'ajoute à l'évidence empirique. En fait, étant donné les énormes oscillations enthalpiques des macromolécules (non cristallisées) dans une cellule à une température viable, elle est également convaincante sur le plan empirique, mais aussi plus robuste sur le plan théorique que les vagues théories qui envisagent que l'information génétique programmable détermine entièrement les processus biologiques, modulo du bruit dispersé.

Ce changement de perspective suggère des dualités fondamentales. Par exemple, les réseaux moléculaires physiques et hautement improbables d'une cellule ne déterminent pas complètement les activités et les composants cellulaires ascendants. Au contraire, ils sont rendus possibles par les contraintes cellulaires mêmes qu'ils produisent (Montévil & Mossio 2015). En effet, il n'y a pas de génération spontanée, des molécules à la vie, à l'exception de la "singularité" totalement inconnue à l'origine de la vie. La vie existante et même artificielle est le résultat d'une histoire, où chaque trajectoire phylogénétique est déclenchée par des événements rares (Longo 2017). Par conséquent, nous devons nous concentrer sur la manière

de comprendre et d'agir sur les contraintes, y compris la plus fondamentale : L'ADN. De cette façon, nous pouvons contribuer à canaliser les processus en modifiant des contraintes de nature diverse. En référence à la discussion précédente, un exemple typique est la "réversion tumorale" (Bizzarri et al. 2014 ; Proietti et al. 2019 ; Kuchling et al. 2020 ; Sonnenschein & Soto 2020). Une telle approche totalement différente contraste avec des décennies d'affirmations et d'échecs concernant la "réécriture du programme génétique brouillé de la tumeur." En outre, je pense que cette approche peut également mettre en lumière notre relation avec l'écosystème : nous avons surtout agi et agissons sur lui en modifiant les contraintes de ses processus - avec l'efficacité et les limites de compréhension et de prédiction propres à ce type d'actions.

En ce qui concerne la théorisation, Weyl (1949) souligne que le principal enseignement méthodologique de la théorie de la relativité, à partir de la relativité de Galilée, consiste à passer de la perspective "subjective-absolue" (si semblable aux approches géocentrique et géocentrique) à la perspective "relative-objective". La construction de l'objectivité scientifique passe par l'analyse des invariants, c'est-à-dire de ce qui est stable par rapport aux transformations des systèmes de référence ; en biologie, cela devrait signifier une stabilité par rapport à une "relativisation" des niveaux d'organisation et des échelles, pour les intégrer. Tout en considérant l'ADN comme une contrainte interne étonnamment importante pour la dynamique cellulaire, nous devons être capables de passer du point de vue de l'ADN à celui de l'organisme et de l'écosystème et vice versa. Ensuite, nous devons comprendre leur intégration et leurs rôles respectifs dans la structure de la détermination biologique (Noble et al. 2019).

Comme indiqué au début de cette note, j'ai beaucoup apprécié que le livre rende certains principes théoriques explicites. Je l'ai également critiqué pour en avoir laissé d'autres implicites. Malgré mon admiration pour le talent expérimental et la perspicacité des auteurs, je tenais à exprimer mon désaccord avec le cadre de pensée biologique qu'ils proposent. Ibn Yunus (Égypte, vers 1000 après J.-C.) devrait-il recevoir le prix Nobel pour sa contribution à l'astronomie ? Certainement oui, malgré les lacunes de sa vision théorique. Cependant, je pense que nous devrions maintenant étudier plus avant la pertinence pratique et théorique de l'analogue du principe asymptotique d'inertie de Galilée en biologie organique, l'état par défaut de la "reproduction avec variation", une application du premier principe de l'évolution de Darwin, la "descendance avec modification", que Darwin considérait comme omniprésente dans toutes les espèces (et qu'il a longuement discuté dans quatre des six premiers chapitres de *On the Origin of Species*). Notez que la "reproduction avec variation" des cellules somatiques dans un tissu (sain) est un état limite, comme le mouvement inertiel en physique. C'est parce que la reproduction dans les cellules somatiques est toujours (mais différemment) contrainte. En posant ce principe darwinien pour toutes les cellules, y compris les cellules somatiques, on suit les traces de 150 ans de microbiologie et on peut mieux comprendre ce qui les contraint dans un organisme, ainsi que l'échec de ces contraintes dans le contrôle de la reproduction cellulaire, comme cela semble se produire principalement dans le cas du cancer (Soto & Sonnenschein 2011). Ce principe devrait se combiner avec la vision unificatrice des organismes comme une "fermeture de contraintes", appliquée à tous les niveaux d'organisation. Tous deux exigent des scientifiques qu'ils précisent les contraintes de la dynamique moléculaire largement brownienne ou chaotique, ainsi que de la reproduction et la motilité des cellules, c'est-à-dire de leurs activités fonctionnelles dans un organisme (Montévil & Mossio 2015 ; Soto et al. 2016 ; Bizzarri et al. 2020).

Je suis persuadé que la biologie des organismes obtiendra d'autres résultats pertinents. Les connaissances et les techniques générées par les travaux des auteurs et de nombreuses autres personnes sur CRISPR y ont contribué et pourraient y contribuer encore. Un exemple très intéressant a déjà été fourni par des études fondamentales, où " les études basées sur CRISPR ont révélé de manière surprenante que... les effets sur l'expression des gènes qui ne sont pas médiés par la transcription de l'ARN elle-même... se produisent dans de nombreux loci qui produisent des lncRNA (non codants longs ARN) ainsi que dans de nombreux loci qui codent des ARNm " (Engreitz et al. 2016 ; Engreitz et al. 2019, p.237). À la suite également des travaux de Cortini et al. (2016), Ramdas & Shivashankar (2015), et d'autres, cela confirme que les actions physico-chimiques et dépendantes du contexte, y compris la structure des lncRNA, peuvent avoir un rôle régulateur clé, bien au-delà de l'approche informationnelle géocentrique. La compréhension par des théories solides, au lieu de vagues "métaphores", et par leur contrepartie expérimentale, tout en encadrant également les résultats remarquables obtenus par les auteurs, devrait être une composante essentielle de la connaissance, bien avant d'agir sur la nature.

Références (Giuseppe Longo's co-authored papers can be found at <https://www.di.ens.fr/users/longo/download.html>)

- Adikusuma, F., Piltz, S., Corbett, M.A. *et al.* (2018) Large deletions induced by Cas9 cleavage. *Nature* **560**, E8–E9. <https://doi.org/10.1038/s41586-018-0380-z>
- Alanis-Lobato, G. et al. (2021) Frequent loss-of-heterozygosity in CRISPR-Cas9–edited early human embryos, *Pnas*, April: <https://www.pnas.org/content/early/2021/04/09/2004832117>.
- Balazsi G, Oudenaarde A., Collins J. (2011) Cellular Decision Making and Biological Noise: From Microbes to Mammals. *Cell* 144, March 18.
- Barendregt, H. (1984) *The Lambda-Calculus: its Syntax, its Semantics*. Amsterdam: North-Holland.
- Begley, C., & Ioannidis, J. (2014). Reproducibility in science. *Circulation Research*, 116, 116–126. doi:10.1161/CIRCRESAHA.114.303819.
- Baker S. G. (2014) Recognizing Paradigm Instability in Theories of Carcinogenesis. *British Journal of Medicine and Medical Research*, 4(5): 1149-1163.
- Baker S. G. (2021) The case for a cancer paradox initiative, *Carcinogenesis*, <https://doi.org/10.1093/carcin/bgab052>
- Barrell BG, Air GM, Hutchison CA III (1976) Overlapping genes in bacteriophage fX174. *Nature* 264: 34-41.
- Bizzarri, M. (2014) System Biology for Understanding Cancer Biology, *Curr Synthetic Sys Biol*, 2:1
- Bizzarri, M, Cucina, A (2014) Tumor and the microenvironment: a chance to reframe the paradigm of carcinogenesis. *Biomed Res Intl* :934038.
- Bizzarri M, Giuliani A, Minini M, Monti N, Cucina A. (2020) Constraints Shape Cell Function and Morphology by Canalizing the Developmental Path along the Waddington's Landscape. *Bioessays*. Apr; 42(4):e1900108. Epub 2020 Feb 27. PMID: 32105359.
- Boersma S, Khuperkar D, Verhagen BMP, Sonneveld S, Grimm JB, Lavis LD, Tanenbaum ME (2019) Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding. *Cell* 178 : 458-472.
- Braun E. (2015) The unforeseen challenge: from genotype-to-phenotype in cell populations, *Rep. Prog. Phys.* 78, 036602.
- Bravi B., Longo G. (2015) The Unconventionality of Nature: Biology, from Noise to Functional Randomness. *Unconventional Computation and Natural Computation*, Springer LNCS 9252, Calude, Dinneen (Eds.), pp 3-34.
- Brett D., Pospisil H., Valcárcel J., Reich L., Bork P. (2001) Alternative splicing and genome complexity, *Nature Genetics*, 30.
- della Briotta Parolo P, Kumar Pan R, Ghosh R, Huberman BA, Kaski K, Fortunato S (2015) Attention decay in science, *Journal of Informetrics* 9:734-745
- Brunet M, Levesque SA, Hunting DJ, Cohen AA , Roucou, X (2018) Recognition of the polycistronic nature of human genes is critical to understanding the genotype- phenotype relationship. *Genome Res* 28 : 609-624.
- Brunet M, Leblanc S, Roucou X (2020) Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs. *Exp Cell Res* 393, 112057

- Calude C., Longo G. (2016) Classical, Quantum and Biological Randomness as Relative Unpredictability. Special issue of *Natural Computing*, 15 (2):263–278
- Calude C, Longo G. (2017). The deluge of Spurious Correlations in Big Data. *Foundations of Science* 22: 595–612.
- Casali M & Merlin F (2020) Rethinking the role of chance in the explanation of cell differentiation. In : *Phenotypic switching*, (Levine H, Jolly M, Kulkarni P, Nanjundiah V, eds), pp. 23-51, Elsevier.
- Chaudari HG, Penterman J, Whitton HJ, Spencer SJ, Flanagan N, Lei Zhang MC, Huang E, Khedkar AS, Toomey JM, Shearer CA, Needham AW, Ho TW, Kulman JD, Cradick TJ, Kernitsky A (2020) Evaluation of Homology-Independent CRISPR-Cas9 Off-Target Assessment Methods. *The CRISPR Journal* 3: 440-453
- Chelly J, J P Concordet, J C Kaplan, and A Kahn (1989) Illegitimate transcription: transcription of any gene in any cell type. *PNAS* April 1, 86 (8) 2617-2621
- Collins, F.S. (1999). Medical and societal consequences of the human genome project. *New England Journal of Medicine* 341, 28–37.
- Cortini, R., M. Barbi, B. Caré, C. Lavelle, A. Lesne, J. Mozziconacci and J.M. Victor (2016) The physics of epigenetics. *Reviews of Modern Physics* 88, 025002
- Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297: 1183-1186.
- Engreitz J, Abudayyeh O., Gootenberg J. & Zhang F. (2019) CRISPR Tools for Systematic Studies of RNA Regulation in *RNA Worlds*. *New Tools for Deep Exploration* (Cech, Steitz eds).
- Engreitz J.M. Haines J.M., Perez E.M., Munson G., Chen J., Kane M., McDonel P.E., Guttman M. & Lander E.S. (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing, *NATURE* | Vol 539 | 17, November.
- von Eschenbach A. (2003) NCI Sets Goal Of Eliminating Suffering And Death Due To Cancer By 2015, *J Natl Med Assoc.*, 95:637-639.
- Fox-Keller E. (2003) *A Feeling for the Organism Life and Work of Barbara McClintock*, Freeman, New York.
- Gatenby, RA. (2017). Is the Genetic Paradigm of Cancer Complete?, *Radiology*, 284:1–3.
- Gerlinger, M et al (22 authors) (2012) Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing, *Engl J Med* 366; 10, march 8.
- Gerstein M. et al. (2007) What is a gene, post-ENCODE? History and updated definition, *Genome Research*, 17: 669-681.
- Gould S.-J. (2002). *The Structure of Evolutionary Theory*, Harvard U. Press.
- Hanahan, D & Weinberg RA (2000) The hallmarks of cancer. *Cell*,100, 57–70.
- Höijer I, Johansson J, Gudmundsson S, Chin C-S, Bunikis I, Häggqvist S, Emmanouilidou A, Wilbe M, den Hoed M, Bondeson M-L, Feuk L, Gyllensten U, Ameer A (2020) Amplification-free long-read sequencing reveals unforeseen CRISPR-Cas9 off-target activity. *Genome Biol* 21:290
- Honegger K & de Bivort B (2018) Stochasticity, individuality and behavior. *Curr Biol* 28: R1-R16
- Huang, S. (2014) The war on cancer: lessons from the war on terror. *Front Oncol.* 4:293.
- Huang, S, Jaurand, MC, Kamp, D, Whysner, J & Heil T (2011) Role of Mutagenicity in Asbestos Fiber-Induced Carcinogenicity and Other Diseases, *J Toxicol Environ Health B Crit Rev.* Jan-Jun; 14(1-4): 179–245.
- Kato S, Lippman SM, Flaherty KT, Kurzrock R. (2016) The Conundrum of Genetic "Drivers" in Benign Conditions. *J Natl Cancer Inst.* 2016 Apr 7; 108(8):djw036. PMID: 27059373; PMCID: PMC5017937.
- de Klerk E, t Hoen PAC (2015) Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends in Genetics* 31:128-139.
- Kuchling F, Friston K, Georgiev G, Levin M. (2020) Morphogenesis as Bayesian inference: A variational approach to pattern formation and control in complex biological systems. *Phys Life Rev.* Jul;33:88-108. doi: 10.1016/j.plrev.2019.06.001. Epub 2019Jun 12. PMID: 31320316.
- Kupiec J.-J. (1983) A probabilistic theory for cell differentiation, embryonic mortality and DNA C-value paradox. *Specul Sci Technol* 6, 471-478 (reprinted in *Organisms*, n. 2, 2020).
- Kupiec J-J (2010) On the lack of specificity of proteins and its consequences for a theory of biological organization. *Prog Biophys Mol Biol* 102:45-52.
- Kreisel, G. (1982). Four letters to G. L. <http://www.di.ens.fr/users/longo/files/FourLettersKreisel.pdf>
- Leff SE, Rosenfeld MG, Evans RM (1986) Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Ann Rev Biochem* 55:1091-117.
- Leibowitz M.L. et al. (2021). Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nature Genetics* | VOL 53 | JUNE 2021 | 895–905 |
- Longo, G. (2014). Science, Problem solving and bibliometrics, in *Bibliometrics: Use and Abuse in the Review of Research Performance*, W. Blockmans et al. (eds), Portland Press.
- Longo, G. (2017). How Future Depends on Past Histories and Rare Events in Systems of Life, *Foundations of Science*, pp. 1-32.
- Longo, G (2018). Interfaces of Incompleteness. In Minati, G, Abram, M & Pessa, E (Eds.) *Systemics of Incompleteness and Quasi-systems*, Springer, New York, NY.

- Longo, G. (2018t). Letter to Alan Turing. In *Theory, Culture and Society, Special Issue on Transversal Posthumanities*. Füller, Braidotti (eds).
- Longo, G. (2019). Information at the Threshold of Interpretation, Science as Human Construction of Sense. In Bertolaso, M. & Sterpetti, F. (Eds.) *Will Science Remain Human?* pp. 67-100, Springer, New York.
- Longo, G. and Montévil M. (2014). *Perspectives on Organisms: Biological Time, Symmetries and Singularities*. Dordrecht: Springer.
- Longo G. and Mossio M. (2020) Geocentrism vs genocentrism: theories without metaphors, metaphors without theories. *Interdisciplinary Science Reviews*, 45 (3), pp. 380-405.
- Maffini, MV, Calabro, JM, Soto & AM, Sonnenschein, C. (2005) Stromal regulation of neoplastic development: Age-dependent normalization of neoplastic mammary cells by mammary stroma. *Am. J. Pathol.* 167, 1405-1410.
- Maeda H., Khatami M. (2018) Analyses of repeated failures in cancer therapy for solid tumors: poor tumor-selective drug delivery, low therapeutic efficacy and unsustainable costs *Clin Transl Med.* 2018; 7: 11.
- Meydan S, Marks J, Klepacki D, Sharma V, Baranov, PV, Firth AE, Margus T, Kefi A, Vazquez-Laslop N, Mankin AS (2019) Retapamulin-Assisted Ribosome Profiling Reveals the Alternative Bacterial Proteome. *Molecular Cell* 74: 481-493.e6
- Mally A, Chipman JK. (2002) Non-genotoxic carcinogens: early effects on gap junctions, cell proliferation and apoptosis in the rat. *Toxicology.* Dec 2;180(3):233-48. doi: 10.1016/s0300-483x(02)00393-1. PMID: 12393293.
- Montévil M (2019). Measurement in biology is methodized by theory. *Biology & Philosophy.* June, 34:35.
- Montévil M. & Longo G. (2018) Big Data and biological knowledge. In *Predictability and the Unpredictable. Life, Evolution and Behaviour*, (Ceccarelli & Frezza, eds) CNR Edizioni, Roma.
- Montévil, M. & Mossio, M. (2015). Closure of constraints in biological organisation. *Journal of Theoretical Biology*, vol. 372: 179-191.
- Montévil, M & Mossio, M (2020) The Identity of Organisms in Scientific Practice: Integrating Historical and Relational Conceptions. *Front. Physiol.* 11:611.
- Modrzejewski D, Hartung F, Lehnert H, Sprink T, Kohl C, Keilwagen J, Wilhelm R (2020) Which Factors Affect the Occurrence of Off-Target Effects Caused by the Use of CRISPR/Cas: A Systematic Review in Plants. *Front. Plant Sci.* 11:574959.
- Mouilleron H, Delcourt V, Roucou X (2016) Death of a dogma: eukaryotic mRNAs can code for more than one protein. *Nucl Ac Res* 44:14-23
- Mustjoki S., Young N. (2021) Somatic Mutations in “Benign” Disease, *N Engl J Med.* May 27.
- National Academy of Sciences (2020). *Heritable Human Genome Editing*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25665>.
- Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463: 457-463.
- Noble R., Tasaki K., Noble P., Noble D. (2019) Biological Relativity requires circular causality but not symmetry of causation: so, where, what and when are the boundaries? *Frontiers in Physiology* 10: 827.
- Paulose, T., Speroni, L., Sonnenschein, C., and Soto, A.M. (2015) Estrogens in the wrong place at the wrong time: fetal BPA exposure and mammary cancer. *Reprod Toxicol* 54: 58-65.
- Pavesi, A., Vianelli, A., Chirico, N., Bao, Y., Blinkova, O., Belshaw, R., Firth, A., Karlin, D. (2018) Overlapping genes and the proteins they encode differ significantly in their sequence composition from non-overlapping genes. *PLoS one* 13, e0202513.
- Paldi, A. 2003. Stochastic gene expression during cell differentiation: order from disorder? *Cell Mol. Life Sci.*, 60, 1775-1779.
- Plomin, R. (2019) *Blueprint: How DNA Makes Us Who We Are*. MIT Press.
- Proietti S, Cucina A, Pensotti A, Biava PM, Minini M, Monti N, Catizone A, Ricci G, Leonetti E, Harrath AH, Alwasel SH, Bizzarri M. (2019) Active Fraction from Embryo Fish Extracts Induces Reversion of the Malignant Invasive Phenotype in Breast Cancer through Down-regulation of TCTP and Modulation of E-cadherin/ β -catenin Pathway. *Int J Mol Sci.* Apr 30;20(9):2151. PMID: 31052313; PMCID: PMC6539734.
- Raj A & van Oudenaarden A. (2008) Stochastic gene expression and its consequences. *Cell* 135:216-226.
- Rohan TE, Miller CA, Li T, Wang Y, Loudig O, Ginsberg M, Glass A, Mardis E. (2018) Somatic mutations in benign breast disease tissue and risk of subsequent invasive breast cancer. *Br J Cancer.* Jun; 118(12): 1662-1664. Epub 2018 Jun 6. PMID: 29872146; PMCID: PMC6008400.
- Ramdas NM, Shivashankar GV. (2015) Cytoskeletal control of nuclear morphology and chromatin organization. *J Mol Biol.* Feb 13;427(3):695-706. Epub 2014 Oct 2. PMID: 25281900.
- Schlub T, Holmes EC (2020) Properties and abundance of overlapping genes in viruses. *Virus evolution* 6(1): veaa009
- Sonnenschein, C., Soto, A.M., (1999) *The society of cells: cancer and control of cell proliferation*. Bios Scientific Pub Limited.
- Sonnenschein & Soto (2020) Over a century of cancer research: Inconvenient truths and promising leads. *PLoS Biol* 18(4): e3000670. Published: April 1, 2020 *PLOS Biology*.

- Soto A., Longo G., Noble D. (eds.) (2016) *From the century of the genome to the century of the organism: New theoretical approaches*, a Special issue of *Progress in Biophysics and Mol. Biology*, Vol. 122, 1, Elsevier.
- Soto AM, Sonnenschein C. (2011) The tissue organization field theory of cancer: A testable replacement for the somatic mutation theory. *BioEssays* 33:332–340.
- Sweeney MF, Hasan N, Soto AM, Sonnenschein C (2015) Environmental endocrine disruptors: Effects on the human male reproductive system. *Rev Endocr Metab Disord* 16:341-357
- Versteg, R. (2014) Cancer: Tumours outside the mutation box, *Nature* vol. 506, pages 438–439
- Waks Z, Klein AM, Silver PA (2011) Cell-to-cell variability of alternative RNA splicing. *Molecular Systems Biology*, 7:506
- Weinberg, R. (2014). Coming Full Circle - from endless complexity to simplicity and back again. *Cell* 157, March 27.
- Weyl, H. 1949. *Philosophy of Mathematics and of Natural Sciences*, Princeton U. Press, New Jersey.