



Les impasses de la sociogénomique

Nicolas Robette

► To cite this version:

Nicolas Robette. Les impasses de la sociogénomique. Population (édition française), 2022, 77 (2), pp.191-227. 10.3917/popu.2202.0191 . hal-03806637

HAL Id: hal-03806637

<https://hal.science/hal-03806637>

Submitted on 7 Oct 2022

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



Nicolas ROBETTE*

Les impasses de la sociogénomique

Depuis deux décennies, la collecte de données génétiques de plus en plus riches a pris de l'ampleur, offrant des opportunités nouvelles à la recherche. Si l'apport de ces données est indéniable dans le domaine de la santé, l'est-il également dans celui des sciences de la population ? Cet article passe en revue de manière critique les hypothèses sur lesquelles repose la sociogénomique, un nouveau champ d'étude qui explore le lien entre génétique et comportements humains, lien débattu de longue date, comme en témoignent les écrits d'Alfred Jacquard et de Daniel Courgeau.

Depuis une quinzaine d'années, des sociologues et des démographes ont entrepris d'intégrer des données génétiques à leurs analyses empiriques. Ils s'appuient pour cela sur le développement récent des technologies de séquençage et de génotypage à haut débit⁽¹⁾, qui permettent de cartographier le génome entier. Ils s'inscrivent également dans le prolongement d'une spécialité de recherche, la génétique comportementale, qui a pris son essor dans les années 1970. Ces travaux, qui se sont notamment intéressés au déterminisme génétique de l'intelligence, ont fait l'objet de critiques radicales, y compris dans les rangs des démographes (Jacquard, 1978 ; Vetta et Courgeau, 2003 ; Courgeau, 2017). Toutefois, ces critiques ne semblent pas toujours avoir été entendues, et l'apparition de nouvelles données et techniques fait surgir de nouveaux problèmes, comme l'indiquent les appels à la prudence des sociétés savantes en génétique humaine (SFGH, 2010 ; Risch, 2016 ; SFGH, 2018).

Cet article se situe dans le prolongement de ce travail critique. On revient tout d'abord sur le développement de la sociogénomique, entendue ici comme la combinaison de la sociologie/démographie et de la génétique (partie I). La partie II présente un état des lieux des limites des concepts et modèles théoriques utilisés par les sociogénomistes, et la partie III les limites des données et méthodes d'analyse. La partie IV fait état d'un bilan provisoire des travaux de sociogénomique et de leurs apports à la connaissance en sociologie et

(1) Les termes suivis d'un astérisque sont définis dans le glossaire du document annexe A.

* CREST-ENSAE, Institut Polytechnique de Paris.

Correspondance : nicolas.robette@uvsq.fr

démographie. Enfin, la conclusion interroge les facteurs favorisant le développement de cette spécialité de recherche, alors même que, sur le plan scientifique, elle semble particulièrement fragile (partie V).

I. Des origines de la génétique à la sociogénomique

La naissance de la génétique est souvent associée aux travaux de Francis Galton et de Gregor Mendel au XIX^e siècle. L'article « The Correlation between Relatives on the Supposition of Mendelian Inheritance », publié par Ronald Fisher en 1918, marque quant à lui une étape décisive dans le développement de la génétique quantitative⁽²⁾. Fisher y pose les bases du modèle « polygénique additif », selon lequel les variations d'un phénotype* sont le résultat de la somme des effets d'un grand nombre de gènes, chaque gène ayant individuellement un effet infinitésimal, et de facteurs environnementaux (partie II.2). Du point de vue mathématique, Fisher propose de décomposer la variance⁽³⁾ du phénotype en la somme de la variance des gènes et de la variance de l'environnement. Il ouvre ainsi la possibilité de calculer l'« héritabilité » d'un phénotype, définie comme la part de la variance du phénotype attribuable à la variance des gènes (partie II.1).

Comprendre comment les gènes contribuent à des traits complexes a été depuis lors au centre de nombreuses recherches et reste une question majeure aujourd'hui. Les applications à l'homme des méthodes de la génétique quantitative se sont multipliées à partir des années 1970 en médecine, mais également par l'intermédiaire du développement d'une nouvelle spécialité de recherche, la génétique comportementale (*behavior genetics* ; Panofsky, 2011 et 2014), qui entend étudier ce que les comportements sociaux doivent au patrimoine génétique des individus. Fortement investie par des psychologues, la génétique comportementale s'intéresse par exemple aux traits de personnalité, aux attitudes sociales et aux maladies mentales, comme la schizophrénie, mais aussi à l'intelligence (via le quotient intellectuel), dont il s'agit de quantifier l'héritabilité. Plus récemment, l'étude de l'héritabilité des traits* comportementaux a été reprise dans d'autres disciplines de sciences humaines, avec le développement de courants de recherche en criminologie (criminologie biosociale ; Larrègue, 2016, 2017, 2018a), en sciences politiques (*genopolitics* ; Larrègue, 2018b) ou en économie (*genoeconomics* ; Benjamin *et al.*, 2012), et l'étude de traits aussi divers que la délinquance, les comportements électoraux ou les revenus.

Dans les années 2000, les technologies de séquençage et de génotypage à haut débit se développent très rapidement. On peut dès lors étudier

(2) Voir les communications du colloque « Un siècle de Fisher », qui s'est tenu à Paris les 12 et 13 septembre 2019 (<https://1siecledefisher.sciencesconf.org/> et, pour les captations des présentations, <https://sfg.igh.cnrs.fr/1-siecle-de-fisher.html>).

(3) Fisher introduit à cette occasion le terme « variance », qui désigne le carré de l'écart type.

simultanément un grand nombre de marqueurs génétiques* chez un grand nombre de sujets, pour un coût qui décroît très vite, ce qui permet l'apparition de nouvelles sources de données et de nouvelles techniques d'analyse. C'est dans ce contexte de rapides progrès technologiques et statistiques que la combinaison de la sociologie (ou de la démographie) et de la génétique, la « sociogénomique », apparaît et se diffuse, à partir de la fin des années 2000. Lorsqu'il est utilisé pour la première fois, le terme de « sociogénomique » désigne l'étude de « la vie sociale en termes moléculaires », qui s'applique aussi bien aux êtres humains qu'aux animaux (Robinson *et al.*, 2005). Dans son ouvrage *Social by Nature. The Promise and Peril of Sociogenomics* (2018), Catherine Bliss appelle « sociogénomique » (ou, le plus souvent, « génomique sociale ») l'application des méthodes génomiques à la recherche en sciences sociales pour trouver les causes génétiques des phénomènes sociaux (à l'exception des maladies). Cet article adopte la définition plus étroite de la « sociogénomique » proposée par Mills et Tropf (2020), qui restreint le périmètre disciplinaire à la sociologie⁽⁴⁾. Ce label est affiché et revendiqué par quelques dizaines de chercheurs, principalement aux États-Unis et au Royaume-Uni, parmi lesquels une quinzaine sont particulièrement actifs. Les recherches empiriques des sociogénomistes se concentrent pour le moment sur un nombre relativement réduit d'objets d'études : principalement le niveau d'éducation⁽⁵⁾, la mobilité sociale, la fécondité, ainsi que quelques comportements considérés comme à risque (tabagisme, alcoolisme, partenaires sexuels multiples).

Les sociogénomistes ne publient que quelques articles dans les années 2000, mais leurs travaux se multiplient la décennie suivante, ce qui tient probablement au développement de nouveaux outils comme les « scores de risques polygéniques » (partie III.4). Ceux-ci trouvent souvent des débouchés dans des revues de génétique et/ou de biologie. Mais ils sont également publiés dans de grandes revues de sociologie – avec un numéro spécial dans l'*American Journal of Sociology* en 2008, des articles ponctuels dans la même revue ainsi que dans *American Sociological Review*, *Social Forces*, *Journal of Marriage & Family*, *Social Science and Medicine*, *Sociological Science* – et de démographie (*Demography*, *Population Studies*), font l'objet de séances dans les congrès de la Population Association of America. Ils bénéficient aussi parfois de financements conséquents, comme le projet « Sociogenome »⁽⁶⁾, dirigé par Melinda Mills à l'université d'Oxford qui s'appuie sur de larges subventions européennes (BSA, 2017). Si la plupart des sociogénomistes n'occupent pas de positions centrales dans le champ de la sociologie, on observe quelques exceptions,

(4) La démographie, dans le contexte académique anglo-saxon, n'est pas aussi nettement séparée de la sociologie qu'en France.

(5) La plupart du temps objectivé à partir du nombre d'années d'études. On peut se demander si cet objet n'est pas le prolongement des travaux de la génétique comportementale sur le quotient intellectuel qui, ayant suscité de nombreuses critiques, sont actuellement délégitimés.

(6) <https://www.sociogenome.org>

comme Dalton Conley (Princeton), Melinda Mills (Oxford), Michael Shanahan (Zurich) ou Jeremy Freese (Stanford), et la spécialité attire un nombre croissant de jeunes chercheurs.

Les arguments que ces « entrepreneurs de génétique » (Shostak et Beckfield, 2015, p. 98) développent pour convaincre de la fécondité de l'intégration de la sociologie/démographie et de la génétique dessinent une sorte de programme de recherche, qui se décline en plusieurs axes (partie IV). Tout d'abord, à un niveau très général, la sociogénomique doit permettre d'améliorer la compréhension des phénomènes sociaux. Elle « ajoute une pièce incroyablement grande au puzzle » (Mills in BSA, 2017, p. 17), afin « d'approfondir la connaissance des structures sociales dans lesquelles nous vivons et des mécanismes qui donnent naissance à ces structures » (Bearman, 2008). Cela implique de se déprendre de la conception durkheimienne des faits sociaux (expliquer le social par le social), considérée comme une entrave (Udry, 2000), et de se « réapproprier » la biologie (Fuller in BSA, 2017, p. 20). Plus concrètement, les interactions entre gènes et environnement – autrement dit la manière dont les effets de l'environnement sont médiés par les prédispositions génétiques et *vice versa* – sont l'une des pistes les plus explorées dans les travaux des sociogénomistes. Par ailleurs, l'ajout de variables sur les caractéristiques génétiques des individus dans les analyses statistiques par régressions est susceptible de produire des modèles économétriques plus robustes, moins biaisés. Enfin, la connaissance des prédispositions génétiques des individus pourrait, dans certains cas, guider les politiques publiques.

II. Les limites des concepts et des modèles théoriques

Les travaux des sociogénomistes mobilisent très souvent un concept, l'héritabilité, et reposent tous sur un cadre théorique, le modèle polygénique additif, qui sont largement discutés et font l'objet d'un certain nombre de critiques.

1. L'héritabilité

Le concept d'héritabilité est omniprésent dans les travaux des sociogénomistes, soit que l'héritabilité d'un phénomène social soit directement mesurée, soit que l'héritabilité mesurée dans des travaux antérieurs serve d'argument à la pertinence de l'intégration d'informations génétiques dans les analyses.

L'origine du terme « héritabilité » est difficile à retracer. Il est souvent attribué à Lush et son ouvrage intitulé *Animal Breeding Plans* (1937). Mais la formulation mathématique de l'héritabilité est antérieure et est l'œuvre de Ronald Fisher en 1918. Celui-ci propose de décomposer la variance du phénotype P en la somme de la variance des gènes G et de la variance de l'environnement E . Il souligne l'importance du rapport $\text{var}(G)/\text{var}(P)$, qui est précisément ce que nous appelons

maintenant l'héritabilité, autrement dit la part de la variabilité d'un phénotype qui est attribuable à la variabilité génétique⁽⁷⁾.

Le modèle de Fisher a d'abord été appliqué à des traits physiques humains mesurables tels que la taille ou le poids, mais aussi à un autre trait quantitatif, le quotient intellectuel (QI), en utilisant les résultats des tests de QI. C'est à partir de la fin des années 1960 que l'héritabilité du QI devient largement débattue, notamment après la publication des travaux de Jensen (1969). En utilisant des données sur le QI recueillies dans différentes études, Jensen estime que l'héritabilité du QI est d'environ 80 %. Il en conclut que les différences d'intelligence entre groupes sociaux sont en grande partie d'origine génétique et que les politiques éducatives visant à réduire les inégalités seraient, de ce fait, inefficaces. Ce type de raisonnement est notoirement prolongé dans *The Bell Curve* de Herrnstein et Murray (1994) ou dans les travaux de l'éminent psychologue Robert Plomin (2018). Parallèlement, l'étude de l'héritabilité du QI fait l'objet de nombreuses critiques théoriques, méthodologiques, morales et politiques (Jacquard, 1978 ; Kempthorne, 1978 ; Lewontin *et al.*, 1984).

Les études sur l'héritabilité des traits cognitifs, culturels et sociaux se sont multipliées depuis les années 1970, notamment grâce au développement de la génétique comportementale. L'estimation de l'héritabilité de différentes maladies humaines courantes a aussi suscité un intérêt considérable. Le diabète est un exemple de ces maladies qui a été au centre de nombreuses études pour en comprendre la contribution génétique (Genin et Clerget-Darpoux, 2015b).

Les sociogénomistes quantifient également l'héritabilité de traits. Par exemple, à partir de données sur des jumelles au Royaume-Uni, Tropf *et al.* (2015) établissent que la variation de l'âge au premier enfant est expliquée pour 26 % par les prédispositions génétiques, pour 14 % par l'environnement partagé dans les fratries et pour 60 % par l'environnement non partagé ou par les erreurs de mesure. Autre exemple, Baier et Van Winkle (2020) constatent que l'héritabilité des performances scolaires est moins forte pour les enfants dont les parents sont séparés, et en concluent que des politiques éducatives pourraient cibler spécifiquement ce type d'enfants pour les aider à réaliser leur potentiel génétique.

Cependant, quelle que soit leur discipline d'appartenance, les travaux mobilisant des mesures d'héritabilité sont bien souvent marqués par de mauvais usages et des interprétations erronées.

Population versus individu

Tout d'abord, le concept d'héritabilité porte sur une population, mais on le confond souvent avec le concept individuel d'hérédité, i.e. le degré auquel

(7) Cette formulation est celle de l'héritabilité « au sens large ». Cependant, la variabilité génétique G peut encore être décomposée en ses composantes additives, dominantes et épistatiques*. Le rapport entre la variabilité génétique additive (qui correspond à l'addition des effets moyens des deux allèles* de chaque locus génétique*) et la variabilité du phénotype est appelé héritabilité « au sens étroit ». Il est couramment utilisé pour la sélection des animaux et des plantes.

un trait donné d'un individu est causé par des facteurs génétiques. La variabilité d'un trait dans une population n'est pas équivalente au déterminisme de ce trait. Dans un cas, on s'intéresse aux variations et, dans l'autre, à la valeur. Dans un cas, la mesure porte sur une population et, dans l'autre, sur un individu. Par conséquent, un trait peut avoir une héritabilité nulle tout en étant héréditaire. Par exemple, les mammifères ont deux oreilles, ce qui est un caractère héréditaire car contrôlé par des gènes et transmis au fil des générations. Pourtant, la très faible variabilité du nombre d'oreilles selon les individus est d'origine accidentelle, donc environnementale et non génétique : l'héritabilité de ce trait est nulle (de Vienne, 2019).

Mesure locale versus universelle

En outre, l'héritabilité n'est pas une mesure universelle, elle est spécifique à la population étudiée. C'est une mesure locale, dans l'espace et dans le temps. En effet, elle dépend de la fréquence des marqueurs génétiques, de la variabilité de l'environnement et de la variabilité du phénotype. Or ces trois éléments varient d'une population à l'autre. Par conséquent, un même groupe d'individus, avec le même patrimoine génétique aura, pour un trait donné, une héritabilité différente selon qu'on le place dans un contexte où l'environnement est constant ou variable.

Variance versus causes

Mais surtout, l'héritabilité ne dit rien des causes, ni des mécanismes ou des origines des différences entre populations. Analyse de variance et analyse des causes sont déconnectées (Lewontin, 1974)⁽⁸⁾, comme l'illustre l'exemple du divorce :

« Le divorce est héritable, mais s'attend-on vraiment à ce que les études de jumeaux sur les processus conjugaux nous conduisent à une explication génétique du divorce ? La question n'est pas de savoir s'il s'agit de facteurs environnementaux ou génétiques ; en fait, comme nous ne le soulignerons jamais assez, le mariage, le divorce et leurs causes sont tout aussi héréditaires que n'importe quoi d'autre. Mais cette héritabilité ne signifie pas que l'un ou l'autre est un processus biologique qui nécessite une analyse génétique [...] ils n'ont pas d'étiologie génétique spécifique. » (Turkheimer *et al.*, 2014, p.18)

L'erreur biologique

Par ailleurs, la mesure de l'héritabilité implique de séparer la contribution des gènes de celle de l'environnement. Elle est utilisée de manière routinisée et utile pour la sélection végétale et animale (pour prédire l'efficacité de cette sélection). Mais c'est dans le cadre de dispositifs expérimentaux permettant

(8) L'analyse de la variance (Anova), en tant qu'outil statistique, n'est de façon générale pas appropriée à la mesure de l'efficacité causale. Elle mesure une efficacité relative et non absolue, et un effet sur les variations et non sur le niveau du trait (Northcott, 2008).

de contrôler l'environnement, ce qui est impossible dans la nature et dans le cas des êtres humains. La compréhension des facteurs environnementaux qui contribuent au développement des traits humains est très partielle et on ne peut jamais être sûr d'avoir identifié l'ensemble de ces facteurs. On ne peut donc pas non plus savoir si les facteurs environnementaux importants sont stables ou s'ils ont une forte variabilité. Par conséquent, l'héritabilité est susceptible de refléter avant tout la variabilité de l'environnement (Moore, 2006).

Plus fondamentalement, la biologie contemporaine a démontré que les traits sont le produit des interactions entre facteurs génétiques et non génétiques à chaque moment du développement (Moore et Shenk, 2017). Les gènes font partie d'un « système développemental d'influences » (Gottlieb, 2001, p. 6126). Les facteurs génétiques, les protéines, les cellules, les organes, les organismes, les populations d'individus, les facteurs culturels et les autres aspects de l'environnement coopèrent pour produire les traits des êtres vivants au cours de leur développement (Moore, 2013).

Finalement, si l'on part du constat que n'importe quel phénotype a un contenu génétique (Rutter, 2002), alors savoir si la variation d'un comportement humain est influencée par les gènes est une question uniquement rhétorique (Courgeau, 2017). En outre, l'héritabilité ne donne pas la cause du fait qu'un individu ait un trait donné. Elle ne permet pas de savoir si des facteurs génétiques contribuent à ce trait, ou quelles sont les influences relatives des gènes et de l'environnement. Elle ne fournit pas non plus d'information sur la transmission d'un trait entre les parents et leurs enfants⁽⁹⁾.

De nombreux auteurs ont souligné l'impasse que constitue l'héritabilité pour la génétique humaine, au moins depuis les années 1970. Lewontin suggère dès 1974 « d'arrêter la recherche sans fin de meilleures méthodes pour estimer des quantités inutiles » (Lewontin, 1974, p. 525), quand Jacquard souligne en 1978 que « la complexité des mathématiques utilisées pour répondre ne suffit pas à donner du sens à une question absurde [...], dénuée de toute signification » (Jacquard, 1978, p. 241). Plus récemment, Moore note que « plutôt que de passer des heures innombrables à analyser comment différentes sources de variation [...] expliquent les variations dans les traits étudiés, notre temps et notre énergie seraient mieux employés à explorer en premier lieu les causes de ces traits » (Moore, 2006, p. 350-351).

2. Le modèle polygénique additif

Le concept d'héritabilité repose sur l'idée que les phénotypes peuvent être décrits comme la somme d'une composante génétique et d'une composante

(9) La philosophe des sciences Evelyn Fox Keller (2010) fait l'hypothèse que l'omniprésence et la persistance des mauvais usages et d'interprétations erronées, même chez les auteurs les plus compétents et prudents, tient en partie à la polysémie des termes utilisés et, en particulier, aux inévitables glissements sémantiques entre la définition de sens commun de l'« *heritability* » (caractère de ce qui peut s'hériter, c'est-à-dire se transmettre d'une génération à l'autre) et sa définition scientifique (rapport entre la variabilité génétique et la variabilité du phénotype).

non génétique (ou environnementale), et que la composante génétique implique un grand nombre de gènes dont les effets s'additionnent et qui apportent chacun une contribution infinitésimale à la variation du phénotype. Ce modèle, appelé « modèle polygénique additif » (ou « modèle infinitésimal »), trouve son origine dans les observations de Galton (1877), dans leur analyse par Pearson (1898), puis dans l'interprétation de Fisher (1918). Il est au fondement de nombreuses approches en génétique humaine et forme le socle de tous les outils présentés dans la partie III. Il repose sur un certain nombre d'hypothèses. Or celles-ci ne sont pas toujours nettement énoncées, et encore moins vérifiées. Courgeau (2017) en identifie cinq principales, sur lesquelles reposent les calculs d'héritabilité :

- H1. Les gènes agissent additivement (i.e. leurs effets s'additionnent).
- H2. La ségrégation des gènes* se fait de façon indépendante.
- H3. L'environnement est indépendant des gènes et aléatoire.
- H4. La population est en équilibre de Hardy-Weinberg, c'est-à-dire qu'il n'y a pas d'individus consanguins, de migration, de mutation, ni de sélection.
- H5. Le nombre de gènes est supposé infini (ce qui simplifie les calculs).

Cependant :

- Il existe des effets de dominance* (interactions entre allèles*) et d'épistasie (interactions entre gènes), donc les effets des gènes ne sont pas seulement additifs (H1).
- Les gènes ne connaissent pas une ségrégation indépendante, en particulier lorsqu'ils sont situés sur un même chromosome. Courgeau (2017) note que Fisher ne savait pas, en 1918, que les gènes étaient répartis sur 23 paires de chromosomes (H2).
- Dans le cadre des expériences sur les animaux ou les plantes, il est envisageable de contrôler les effets de l'environnement (ou certains d'entre eux). Pour les populations humaines, en revanche, cela est impossible, de sorte que l'exposition à l'environnement n'est pas aléatoire. De plus, les facteurs génétiques et environnementaux interagissent et ne se transmettent pas de manière indépendante, notamment du fait de phénomènes épigénétiques* – gènes soumis à empreinte*, méthylation*, etc. (Génin et Clerget-Darpoux, 2015b) (H3).
- Du fait des unions homogames (voir *infra*), H4 n'est pas vérifiée (Courgeau, 2017), d'autant que les formules proposées par Fisher pour les corrélations entre apparentés sont incorrectes (Vetta, 1976).
- On estime généralement à environ 20 000 le nombre de gènes codant les protéines (H5).

Génin et Clerget-Darpoux (2015b) mentionnent aussi le fait que le modèle polygénique additif suppose que les facteurs génétiques et environnementaux sont nombreux et ont chacun une faible contribution, autrement dit qu'aucun facteur

généétique ou environnemental n'a une contribution majeure. Or il existe bien des facteurs environnementaux exerçant des effets majeurs, comme c'est le cas de la durée des études sur l'âge au premier enfant ou du régime alimentaire et de l'activité physique sur le diabète de type II (Génin et Clerget-Darpoux, 2015b).

Finalement, on le voit, aucune des hypothèses sur lesquelles repose le modèle n'est vérifiée. C'est pourquoi des voix de plus en plus nombreuses appellent à dépasser le modèle polygénique additif et à un profond changement de paradigme en génétique humaine, « plus proche du véritable système biologique » (Nelson *et al.*, 2013, p. 671). Le « modèle omnigénique » de Boyle *et al.* (2017) est une proposition qui va dans ce sens. Les auteurs prennent note du fait que, dans les études d'association à l'échelle du génome, les associations statistiques entre les marqueurs génétiques et les maladies identifient un grand nombre de gènes dispersés dans le génome, y compris de nombreux gènes sans lien évident avec la maladie. Cela va à l'encontre de l'attente selon laquelle les marqueurs causaux seraient regroupés dans les principales voies (*pathways*) liées à la maladie. Boyle *et al.* suggèrent que les réseaux de régulation des gènes sont tellement interconnectés que tous les gènes sont susceptibles d'influencer les fonctions des gènes centraux de la maladie. De ce fait, une distinction est faite entre les gènes régulateurs et les gènes centraux. Mais surtout, selon le modèle omnigénique, la majeure partie de l'héritabilité s'explique par l'effet de gènes situés en dehors des voies centrales.

III. Les limites des données et des méthodes d'analyse

Les méthodes utilisées par les sociogénomistes, comme par les chercheurs en génétique comportementale ou dans d'autres disciplines apparentées, reposent sur un certain nombre de présupposés conceptuels et statistiques dont il convient d'examiner la validité.

1. Études de jumeaux

Les premières estimations de l'héritabilité ont été dérivées de données empiriques sur les corrélations phénotypiques entre parents, et différentes approches ont été proposées pour comparer ces corrélations entre différents types de parents (Tenesa et Haley, 2013). Parmi les différents types de parents, le plus exploitable en génétique humaine est la comparaison des vrais et faux jumeaux. En 1876, Galton a proposé l'utilisation de jumeaux pour distinguer les facteurs génétiques et environnementaux dans l'expression d'un trait. Mais ce n'est qu'au début du vingtième siècle qu'émerge l'idée qu'il existe deux types de jumeaux : les monozygotes (MZ, qui ont 100 % de leur patrimoine génétique en commun) et les dizygotes (DZ, qui ne partagent que 50 % de leur génotype). En 1924, Siemens a publié la première étude comparant la similarité entre les jumeaux MZ et DZ. En 1960, Falconer a montré que l'héritabilité peut être simplement estimée à partir

de la différence entre les taux de concordance, soit le taux de similarité pour la caractéristique étudiée, MZ et DZ. Dès lors, les « *twin studies* » se sont rapidement multipliées pour étudier les traits humains (Polderman *et al.*, 2015). L'étude sur l'héritabilité de l'âge au premier enfant mentionnée dans la partie précédente s'appuyait précisément sur des données de jumeaux (Tropf *et al.*, 2015).

Ces études de jumeaux consistent, pour un trait donné, à mesurer la ressemblance parmi des paires de jumeaux monozygotes et parmi des paires de jumeaux dizygotes. Si la ressemblance est plus forte parmi les jumeaux monozygotes, on considère que le trait étudié est – au moins pour une part – déterminé par les gènes. Des modèles statistiques permettent de quantifier cette part attribuée à l'héritabilité du trait.

Les études de jumeaux reposent sur un ensemble d'hypothèses qui ne sont pas toujours mentionnées, en particulier les suivantes (Joseph, 2013 ; Burt et Simons, 2014) :

- H1. Les chercheurs peuvent déterminer de façon fiable et précise si les jumeaux sont monozygotes (MZ) ou dizygotes (DZ).
- H2. Les gènes des jumeaux MZ sont identiques à 100 %, ceux des jumeaux DZ à 50 %.
- H3. Le pourcentage de gènes partagés par des jumeaux reste le même tout au long de la vie.
- H4. La variation phénotypique peut être décomposée en une composante génétique, une composante environnementale partagée par les jumeaux et une composante environnementale non partagée.
- H5. Les gènes concernés exercent un effet additif.
- H6. Les parents se mettent en couple de manière aléatoire pour le trait étudié, autrement dit il n'y a pas d'homogamie (« *random mating* », dite aussi « *no assortative mating assumption* »).
- H7. Les environnements des jumeaux MZ ne sont pas plus similaires que ceux des jumeaux DZ (dite « *equal environment assumption* »).

Or une partie (au moins) de ces hypothèses ne sont pas vérifiées. On a vu dans la partie précédente que la stricte additivité des effets génétiques n'est pas recevable (H5). La possibilité de décomposer la variation du phénotype en une composante génétique et une composante environnementale implique l'absence d'interactions entre gènes et environnement, postulat dont on a vu également l'irrecevabilité (H4). L'homogamie est un résultat bien établi, en génétique humaine (un individu tend à ressembler génétiquement plus à son conjoint qu'à un individu pris au hasard dans la population, voir Conley et Fletcher, 2017) comme en sciences sociales (H6). Or l'homogamie sociale est associée à l'homogamie génétique, qui accroît la ressemblance génétique entre les enfants. Dans le contexte des études de jumeaux, cela signifie que l'homogamie accroît la ressemblance entre jumeaux DZ relativement aux jumeaux MZ, donc biaise à la baisse l'estimation de l'héritabilité.

Par ailleurs, les recherches récentes en génétique montrent que les génotypes des jumeaux monozygotes ne sont pas identiques à 100 % (H2) et que l'identité génétique entre jumeaux n'est pas constante tout au long de la vie (H3), ces discordances étant suffisantes pour rendre les estimations de l'héritabilité peu fiables (Charney, 2012).

L'hypothèse d'*equal environment* (H7) est sans doute la plus discutée. Depuis les années 1960, les preuves empiriques s'accumulent, montrant que les jumeaux monozygotes vivent dans des environnements sociaux plus similaires que les dizygotes. Par exemple, ils ont plus de chances d'être traités de la même manière par leurs parents, d'avoir les mêmes amis, d'être dans la même classe, de passer du temps ensemble, d'être plus attachés l'un à l'autre, etc. (Joseph, 2013 ; Burt et Simons, 2014).

De plus, l'environnement prénatal (intra-utérin) des jumeaux monozygotes et dizygotes est différent : les environnements prénataux des jumeaux MZ (qui partagent souvent le même placenta) sont plus semblables que ceux des jumeaux DZ (qui ne partagent jamais le même placenta). Or on sait que cet environnement a des conséquences très importantes sur de nombreux aspects du développement (Charney, 2017).

La plupart des défenseurs des études de jumeaux reconnaissent que les environnements des jumeaux MZ sont plus similaires que ceux des jumeaux DZ. Mais certains d'entre eux reformulent l'hypothèse pour la rendre moins contraignante : il suffirait que les environnements soient « égaux » dans les caractéristiques directement liées au trait étudié (*trait-relevant equal environment assumption*). Ce faisant, ils tendraient à faire porter la charge de la preuve sur les critiques, auxquels il incomberait de démontrer que l'hypothèse, dans cette version révisée, n'est pas valide.

De nombreux travaux tentent de montrer que la violation de l'hypothèse d'environnements égaux n'engendre pas de biais dans les calculs d'héritabilité. Parmi eux, les plus rigoureux et convaincants sont sans doute ceux utilisant des données sur les jumeaux « mal classés »⁽¹⁰⁾ (Conley *et al.*, 2013). Cependant, ils n'ont pas une portée générale, dans le sens où ils analysent des données et des traits particuliers, et reposent implicitement sur la validité de l'ensemble des autres hypothèses sous-jacentes aux études de jumeaux. Guo (1999) montre, à l'inverse, que l'estimation d'héritabilité peut être relativement élevée même en l'absence complète de facteurs génétiques lorsque les environnements des jumeaux MZ sont plus ressemblants que ceux des jumeaux DZ.

2. Études d'associations pangénomiques (GWAS)

À partir des années 2000, les technologies de séquençage et de génotypage à haut débit se développent. On peut dès lors étudier simultanément un grand nombre de marqueurs génétiques, chez un grand nombre de sujets, pour un

(10) Jumeaux MZ classés par erreur comme jumeaux DZ et *vice versa*.

coût qui décroît extrêmement vite⁽¹¹⁾. Ce progrès technologique permet l'apparition des « études d'associations pangénomiques » (*genome-wide association studies*, ou GWAS). Leur principe consiste à cartographier (i.e. « génotyper »), pour chaque individu étudié, un nombre très élevé⁽¹²⁾ de polymorphismes nucléotidiques* (*single nucleotide polymorphism*, ou SNP). On réalise ensuite un test statistique d'association pour chaque polymorphisme, afin de détecter l'ensemble des « hits », i.e. des polymorphismes dont l'association statistique avec le trait étudié est significative, c'est-à-dire ceux qui sont significativement plus souvent présents chez les individus ayant le trait.

Dans le domaine de la sociogénomique, par exemple, à partir d'une étude d'associations pangénomiques, Barban *et al.* identifient douze loci* influençant les comportements reproductifs, « ce qui permet de mieux comprendre ces traits complexes » (Barban *et al.*, 2016, p. 1).

Hypothèses

Les études d'associations pangénomiques s'appuient sur le modèle polygénique additif, mais également sur quelques hypothèses qui lui sont propres, notamment les suivantes (Charney, 2012) :

- H1. Les individus ont un ADN identique dans tous les tissus et cellules de leur corps (à l'exception des cellules germinales et de certaines cellules du système immunitaire).
- H2. La présence d'un gène particulier (polymorphisme ou mutation) implique qu'il est activé, c'est-à-dire qu'il peut être transcrit d'une manière associée à ce polymorphisme ou mutation. Par conséquent, les deux mêmes polymorphismes chez deux individus donnés auront la même capacité à être transcrits de la même manière (i.e. ils seront tous deux activés).

Or, comme l'explique Charney (2012), les individus n'ont pas un ADN identique dans toutes les cellules de leur corps, du fait du phénomène répandu de mosaïcisme* (H1) ; la présence d'un allèle particulier n'implique pas qu'il puisse être transcrit de la manière associée à cet allèle, car il peut être épigénétiquement réduit au silence (H2).

Par ailleurs, du fait que l'on mesure l'association statistique entre les SNP et un phénotype à partir de l'estimation de modèles de régression, les hypothèses sous-jacentes à ces modèles s'appliquent, mais certaines d'entre elles ne sont pas valides dans le cas des GWAS (Angers *et al.*, 2019). En particulier, l'effet d'un SNP n'est pas linéaire en termes de nombre d'allèles mineures (hypothèse de linéarité) si un allèle est dominant ou si l'allèle causal est récessif. Les données utilisées sont rarement des échantillons tirés aléatoirement dans

(11) Le coût du génotypage d'un génome humain est ainsi passé de 10 millions de dollars à environ 1 000 dollars entre 2007 et 2015.

(<https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>).

(12) Plusieurs centaines de milliers, voire plus d'un million.

une population (hypothèse d'échantillonnage aléatoire). Si un SNP est corrélé avec des facteurs environnementaux non observés, le paramètre estimé sera biaisé de telle manière qu'une cause environnementale sera attribuée au génotype de façon erronée⁽¹³⁾ (hypothèse d'exogénéité des erreurs). La même hypothèse sera violée si un SNP n'est pas causal mais est corrélé à un SNP causal (Angers *et al.*, 2019).

Faux positifs et répliquabilité

Du fait du très grand nombre d'associations testées, le risque de « faux positifs » – i.e. que les associations statistiquement significatives soient le fait du hasard – est important, même s'il est limité par l'utilisation de la correction de Bonferroni, c'est-à-dire l'adoption d'un seuil de significativité très bas pour corriger le biais lié aux tests multiples⁽¹⁴⁾.

La répliquabilité peut aussi être un problème. En effet, d'une étude à l'autre portant sur un même phénotype, on peine parfois à retrouver les mêmes associations significatives. Par exemple, Smith (2019) rassemble les résultats de 17 GWAS sur l'intelligence et constate que plus de 87 % des 2 335 SNP identifiés ne l'ont été qu'une seule fois.

Données

Les études d'associations pangénomiques s'appuient sur des données qui manquent beaucoup de diversité, sur les plans démographiques, géographiques et de l'ascendance : elles sont très largement biaisées en faveur des individus blancs, d'origine européenne⁽¹⁵⁾. Les échantillons sont également souvent non représentatifs, avec une surreprésentation des femmes, des personnes âgées et des classes supérieures (Mills et Rahal, 2019). Ces biais de diversité et d'échantillonnage peuvent avoir des conséquences problématiques sur la fiabilité des résultats produits (Mills et Tropf, 2020).

Par ailleurs, du fait de la nécessité statistique de disposer d'un grand nombre d'observations, les chercheurs ont tendance à combiner des bases de données d'origines diverses (enquêtes épidémiologiques, banques d'ADN, données d'entreprises privées commercialisant des tests génétiques, etc.), ce qui peut compromettre la qualité des données finalement analysées (Barton *et al.*, 2019).

De plus, peu des données utilisées ont été collectées avec des visées sociologiques ou démographiques, et les indicateurs disponibles pour mesurer des phénomènes sociaux, tels que la réussite scolaire, ont souvent une précision et une pertinence limitées (Mill et Tropf, 2020), d'autant plus qu'ils doivent être présents dans toutes les GWAS combinées⁽¹⁶⁾.

(13) Ce qui renvoie au problème de stratification de la population discuté *infra*.

(14) Le seuil couramment utilisé est $p < 5 \times 10^{-8}$.

(15) Cela concerne 88 % des individus dans les GWAS en 2017, selon Mills et Rahal (2019).

(16) Cela explique sans doute le fait que l'ensemble des travaux de sociogénomique sur la réussite scolaire utilisent un indicateur rudimentaire comme le nombre d'années d'études.

Enfin, les études d'associations pangénomiques se concentrent sur un seul type de polymorphisme génétique, les polymorphismes nucléotidiques (SNP). Ces marqueurs génétiques ont été retenus parce que ce sont les plus fréquents. Mais il en existe de nombreux autres, comme la variabilité du nombre de copies, les copies multiples de segments de gènes, les gènes entiers ou même les chromosomes, qui tous sont susceptibles d'exercer des effets sur les phénotypes (Charney, 2013). Le choix, dans les GWAS, de se concentrer sur les SNP renvoie à une raison pragmatique – la facilité de collecte – et non scientifique, et l'information sur les associations entre gènes et phénotypes est finalement très pauvre.

3. La méthode *Genome-based restricted maximum likelihood* (GREML)

Le nombre d'études d'association pangénomique a explosé durant les quinze dernières années, engendrant le développement de nouvelles méthodes d'analyse, notamment pour estimer l'héritabilité à partir des polymorphismes nucléotidiques recensés dans ces études, donc sans s'appuyer sur des études de jumeaux (Speed *et al.*, 2020). Parmi ces méthodes, GREML* (*Genome-based restricted maximum likelihood*) est sans doute la plus mobilisée. Elle permet d'étudier la similarité génétique entre des individus non apparentés (Yang *et al.*, 2011).

Cependant, du point de vue de la génétique des populations, l'apparement ne se réduit pas aux liens de parentés : GREML doit faire face au biais de stratification de la population (voir partie III.4). Or, pour le moment, les techniques mises en œuvre pour corriger ce biais ne sont pas adaptées (Browning et Browning, 2011, 2013 ; Janss *et al.*, 2012).

GREML est également sensible à l'échantillonnage et aux erreurs de mesure du phénotype ; les estimations sont biaisées et les écarts types inexacts : la méthode ne produit pas une estimation fiable et stable de l'héritabilité (Charney, 2013 ; Kumar *et al.*, 2016a, 2016b).

En outre, les mesures d'héritabilité obtenues à partir des études de jumeaux sont généralement élevées, alors qu'elles le sont beaucoup moins à partir des études d'associations pangénomiques. Par exemple, l'héritabilité de l'âge au premier enfant est estimée à environ 30 % à partir d'études de jumeaux (Briley *et al.*, 2017) et à 15 % à partir de la méthode GREML (Tropf *et al.*, 2015b), voire 0,9 % à partir d'un score de risque polygénique (partie III.4.) (Mills *et al.*, 2018). L'écart entre ces résultats donne lieu à des débats fournis sur l'origine de cette « héritabilité manquante » (*missing heritability*, Manolio *et al.*, 2009), qui est souvent interprétée comme un artefact de l'imperfection transitoire des données et des méthodes statistiques employées.

4. Les scores de risque polygénique

En 2007, Wray *et al.* ont proposé un nouvel outil de prédiction pour les cliniciens : le « score de risque polygénique »* (*polygenic risk score*, ou PGS). Un PGS est une variable quantitative qui résume la prédisposition génétique

d'un individu pour un trait donné. Il est calculé à partir de la combinaison linéaire des SNP présents chez l'individu, pondérée par la taille des effets des SNP mesurés dans la GWAS (Dudbridge, 2013). Il fait l'hypothèse que chaque individu possède un certain degré de prédisposition génétique, résultant de petites contributions de nombreux marqueurs génétiques.

Des applications logicielles calculant le PGS individuel pour de nombreuses maladies et destinées à aider la décision clinique ont été développées. Le nombre d'articles vantant les avantages de l'utilisation des PGS pour différentes maladies complexes a augmenté de façon exponentielle au cours de la dernière décennie⁽¹⁷⁾. Les PGS sont également très mobilisés par les sociogénomistes⁽¹⁸⁾. Par exemple, Mills *et al.* (2018) montrent que les femmes qui ont une prédisposition génétique à avoir un premier enfant tardivement présentent également un décalage de l'ensemble de leur période reproductive, avec une apparition plus tardive des premières règles et de la ménopause. Domingue *et al.* (2015) observent que les individus ayant une prédisposition génétique élevée à la réussite scolaire font en moyenne plus d'années d'études, même en neutralisant l'effet du milieu social. Ils en concluent que la prédisposition génétique exerce bien un effet causal.

Les PGS sont également utilisés par les sociogénomistes pour étudier les interactions entre gènes et environnement. Par exemple, Schmitz et Conley (2016) analysent un score de risque polygénique pour le tabagisme parmi les vétérans du Vietnam. Ils montrent que, parmi les vétérans ayant des prédispositions génétiques élevées, le tabagisme est significativement atténué chez ceux ayant fait des études après la guerre. Conley *et al.* (2015) s'interrogent sur la manière dont les gènes interviennent dans la relation entre le niveau d'études des parents et celui de leurs enfants. Ils constatent que la corrélation entre l'éducation des parents et celle des enfants est attribuable pour un sixième à la transmission génétique et cinq sixièmes à l'héritage social ; que le score de risque polygénique parental n'a pas d'effet significatif sur la réussite scolaire des enfants quand on contrôle par le score de risque des enfants ; que les prédispositions génétiques des enfants ne sont pas modérées par les caractéristiques sociodémographiques des parents. Ils concluent à la coexistence de deux systèmes d'héritage distincts, l'un génétique et l'autre social.

S'appuyant sur les résultats des études d'associations pangénomiques, les scores de risque polygéniques en héritent les limites. Mais leur utilisation fait aussi émerger d'autres problèmes.

Sensibilité

Tout d'abord, les PGS dépendent beaucoup des différentes décisions prises lors de leur calcul : utilisation de l'ensemble des SNP ou uniquement des plus

(17) 2 339 publications apparaissent sur *Pubmed* lors de la recherche de « Polygenic Risk Score(s) » le 20 octobre 2021. La première date de 2010 ; plus des 2/3 ont été publiées au cours des trois dernières années ; le taux de croissance annuel moyen du nombre de publications est de 75 %.

(18) On peut même raisonnablement penser que le développement des PGS a marqué un tournant dans l'utilisation des données génomiques par les sociologues et démographes.

fortement associés au phénotype, choix d'un seuil de significativité, « élagage selon le déséquilibre de liaison » (*LD-pruning**), etc. (Ware *et al.*, 2017).

Euro-centrisme

Ensuite, les PGS sont très majoritairement calculés à partir de populations d'origine européenne, reflétant en cela le biais d'échantillonnage des GWAS. Pour les autres populations, leurs prédictions sont bien moins performantes, voire très médiocres, comme c'est le cas pour les populations d'ascendance africaine (Martin *et al.*, 2019). L'utilisation fait donc courir le risque d'accroître les inégalités entre les populations étudiées par les GWAS et les autres, par exemple dans le domaine de la santé (Martin *et al.*, 2019).

Stratification de la population

La stratification de la population peut être définie comme la présence d'une différence systématique dans la fréquence des allèles entre sous-populations, du fait d'ascendances différentes. Ce phénomène, omniprésent pour l'espèce humaine, peut biaiser les résultats : les associations observées entre gènes et phénotypes peuvent être (au moins en partie) fallacieuses parce qu'elles reflètent aussi des différences de structures génétiques. Dit autrement, un facteur environnemental associé au phénotype peut différer d'une sous-population à l'autre. Ce facteur étant associé à la fois au phénotype et aux variations génétiques de la population, il devient un facteur de confusion.

Un exemple classique de ce biais concerne l'aptitude à manger avec des baguettes (Lander et Schork, 1994). Il existe peut-être des marqueurs génétiques qui affectent la capacité à manipuler des baguettes. Mais, à l'échelle de la population mondiale, la majeure partie de la variation de ce trait est due à des différences de contexte culturel, donc à des différences environnementales et non génétiques. Une GWAS identifierait des marqueurs associés avec l'aptitude à manger avec des baguettes. Mais ces associations seraient fallacieuses, car reflétant en réalité des différences génétiques entre les personnes originaires d'Asie orientale et celles originaires du reste du monde.

Or le biais introduit par la stratification de la population peut être petit au niveau d'un locus individuel, mais devenir très important lorsqu'on agrège des milliers de loci, comme on le fait quand on calcule un PGS (Barton *et al.*, 2019).

Il existe des méthodes pour tenter de « contrôler » ce biais, comme l'analyse en composantes principales, mais elles sont insuffisantes pour l'éliminer (Dandine-Roulland *et al.*, 2016 ; Curtis, 2018 ; Berg *et al.*, 2019 ; Haworth *et al.*, 2019 ; Sohail *et al.*, 2019), y compris sur des populations relativement homogènes⁽¹⁹⁾. Et même à l'intérieur d'une même ascendance (*ancestry*), il existe des effets de composition, selon le lieu de naissance (Haworth *et al.*, 2019), le sexe et l'âge (Mostafavi *et al.*, 2020), l'ethnie (Freese *et al.*, 2019) ou la classe sociale (Abdellaoui *et al.*, 2019 ; Richardson et Jones, 2019).

(19) Voir Kerminen *et al.* (2019) sur l'exemple de la Finlande.

Finalement, aucune population naturelle n'est sans structure – c'est-à-dire homogène, génétiquement aléatoire – et on ne peut pas contrôler l'environnement de façon expérimentale comme lorsqu'on étudie les plantes ou les animaux (Barton *et al.*, 2019). Aucune technique statistique ne semble capable de neutraliser efficacement la complexité de la stratification de la population. De ce fait, les corrélations observées entre les PGS et les traits étudiés sont en partie fallacieuses (Richardson, 2017).

5. Études « gène candidat »

Pour étudier les interactions entre gènes et environnement, les sociogénomistes utilisent des scores de risques polygéniques, mais aussi une autre approche, plus ancienne, appelée « gène candidat » (*candidate gene studies*). Celle-ci consiste à cibler certains marqueurs génétiques dont on connaît la fonction physiologique et dont on fait l'hypothèse qu'ils ont un effet sur le trait étudié. Si le trait est significativement plus souvent présent chez les individus possédant les marqueurs génétiques, l'hypothèse est validée.

Guo *et al.* (2008) étudient le nombre de partenaires sexuels comme comportement à risque parmi les jeunes hommes blancs. Ils montrent que le génotype 9R/9R exerce un effet protecteur, mais que cet effet tend à disparaître dans les établissements scolaires où la proportion d'élèves commençant à avoir des relations sexuelles précoces est plus élevée. Pescosolido *et al.* (2008) mettent en évidence que le risque de dépendance alcoolique des femmes est peu affecté par le gène GABRA2, et que cette influence génétique est atténuée par le soutien familial mais accentuée par la déprivation durant l'enfance.

Les études « gène candidat » souffrent d'un problème majeur, celui de la répliquabilité. En effet, la plupart des associations entre gènes et phénotypes mises en évidence n'ont pas pu être répliquées sur de nouvelles données. Cette non-répliquabilité est souvent expliquée par le manque de puissance statistique, qui serait à l'origine de « faux positifs » parmi les associations observées (Chabris *et al.*, 2012).

Une autre faiblesse de cette approche est le problème de la « surabondance », autrement dit le fait que des phénotypes nombreux et divers ont été associés aux mêmes gènes. Charney et English (2012) montrent que quatre gènes (MAOA, 5-HTT, DRD2 et DRD4) sont à l'origine d'un grand nombre de travaux et recensent dans un tableau de 15 pages les phénotypes ayant été « expliqués » par l'un de ces gènes, de l'âge à la première relation sexuelle au comportement de vote, en passant par l'alcoolisme, le cancer colorectal, la créativité, l'appartenance à un gang, le syndrome de Tourette, l'ouverture d'esprit ou l'éjaculation précoce. Une telle variété de phénotypes associés à ces quatre gènes montre que ces derniers portent finalement peu d'information sur les mécanismes biologiques susceptibles de véritablement expliquer l'existence des phénotypes.

En outre, les études « gène candidat » sont le plus souvent réduites à de simples études d'association statistique, alors même que des méthodes plus avancées existent pour tester des modèles génétiques (Clerget-Darpoux *et al.*, 1988).

IV. Bilan et perspectives des travaux de sociogénomique

Au-delà des limites conceptuelles et statistiques des approches mobilisées, quel bilan peut-on tirer de la contribution des travaux de sociogénomique à la connaissance en sociologie ou en démographie ?

1. De l'association statistique aux mécanismes

Tout d'abord, le fossé est immense entre un « *hit* » et son interprétation biologique, i.e. entre l'identification d'un SNP associé au trait et la compréhension des mécanismes produisant le trait. Cela est valable en médecine et *a fortiori* en sociologie et en démographie, où le chemin qui mène des gènes au phénomène étudié est sans doute plus long et sinueux. Combler un tant soit peu ce fossé nécessiterait l'usage et la combinaison de données et de méthodes variées (Bourgain *et al.*, 2007).

Certes, dans quelques cas, les gènes identifiés par les sociogénomistes ont des fonctions biologiques déjà connues. C'est le cas de certains gènes statistiquement associés aux comportements de fécondité, dont on sait qu'ils sont biologiquement associés à des processus liés à la fécondité, comme la croissance des ovaires, la production d'œstrogène ou la stimulation hormonale (Barban *et al.*, 2016). Mais, outre le fait que ces résultats sont rares en sociogénomique et quelque peu tautologiques, on peut se demander s'ils ne représentent pas plus un progrès pour la biologie ou la médecine que pour la sociologie et la démographie. Par ailleurs, ils ne constituent qu'un tout petit pas vers la compréhension des mécanismes reliant les gènes, les autres composantes biologiques, les facteurs sociodémographiques et les processus étudiés (fécondité, réussite scolaire, etc.), qui semblent largement hors de portée à l'heure actuelle.

2. Purification des effets

Selon les sociogénomistes, la prise en compte de facteurs génétiques dans les modèles économétriques est susceptible d'éliminer les facteurs de confusion génétiques, pour obtenir des effets causaux des variables sociodémographiques plus « purs », moins biaisés, avec des écarts types réduits et, finalement, d'améliorer la « puissance prédictive » des modèles mobilisés en étude de la population, que ce soit en sociologie ou en démographie (Guo *et al.*, 2008 ; Conley *et al.*, 2014 ; Conley, 2016 ; Freese, 2018).

Dans une optique strictement prédictive, la prise en compte de données génomiques est légitime. Cependant, on constate que les performances

prédictives des scores de risques polygéniques sont faibles. La plus-value est donc limitée. Certains sociogénomistes répondent à cette critique en pointant le fait que les facteurs sociodémographiques ont également de faibles performances prédictives. L'argument n'est guère convaincant, pour plusieurs raisons. Tout d'abord, certains facteurs sociodémographiques exercent en fait un effet majeur. Par exemple, chez les femmes, la durée des études influe très fortement sur l'âge au premier enfant : la corrélation est de 54 % aux États-Unis (Marini, 1984), quand le PGS explique moins de 1 % des variations de l'âge au premier enfant (Mills *et al.*, 2018). De même, l'origine sociale pèse fortement sur la réussite scolaire : dans les pays l'OCDE, le nombre d'années d'études des parents explique environ 30 % des variations du nombre d'années d'études des enfants (OCDE, 2019), contre 2 % à 3 % pour le PGS (Conley *et al.*, 2015). En outre, ces deux catégories de facteurs n'ont pas le même statut. Alors que le PGS, en tant que mesure d'une prédisposition génétique, est supposé capturer l'effet global de la génétique, voire de la nature par opposition à l'environnement, un facteur sociodémographique n'a pas la prétention de mesurer l'effet du « social », ni même d'une de ses dimensions. Il n'en est toujours qu'un indicateur partiel et imparfait⁽²⁰⁾.

Dans une optique explicative, l'intégration d'un PGS dans un modèle économétrique est problématique. Éliminer les facteurs de confusion génétiques et « purifier » l'effet de la variable explicative principale à l'aide d'un PGS implique de savoir précisément ce qui est « contrôlé » par le PGS. Or ce n'est pas le cas, en particulier du fait du biais de stratification de la population, qui fait que, lorsqu'on contrôle par une prédisposition génétique, on contrôle aussi en partie des facteurs sociodémographiques, sans que l'on puisse connaître leur nature ni l'ampleur du problème.

Par ailleurs, d'un point de vue méthodologique, les techniques économétriques employées présentent de nombreuses limites, d'autant plus lorsqu'elles sont appliquées à des données génomiques. C'est par exemple le cas de la plus répandue d'entre elles, la randomisation mendélienne, qui adapte l'approche économétrique par variable instrumentale en utilisant une variable génétique comme instrument, et peine à respecter les hypothèses statistiques du modèle⁽²¹⁾ (Davey Smith et Ebrahim, 2003 ; Nitsch *et al.*, 2006 ; Mills et Tropf, 2020).

De plus, ces usages se placent uniquement dans le cadre étroit de la « réalité unilinéaire générale » décrite et critiquée par Andrew Abbott (1988), et d'un « positivisme instrumental »⁽²²⁾ selon lequel il existerait un effet « réel » d'un

(20) La catégorie socioprofessionnelle n'épuise pas ce que représente l'origine sociale dans un modèle théorique par exemple.

(21) Comme les hypothèses de restrictions d'exclusion ou d'absence de confusion dans les associations entre génotype, trait étudié et phénotype intermédiaire.

(22) Selon Christopher Bryant, « le “positivisme instrumental” [...] est “instrumental” dans la mesure où ce sont les instruments de recherche disponibles qui délimitent l'objet de la recherche, et “positiviste” en ce que cette autocontrainte des sociologues traduit leur volonté de se soumettre à une rigueur d'analyse comparable à celle qu'ils attribuent aux sciences naturelles » (Bryant, 1989).

facteur, dont la modélisation permettrait de s'approcher en le séparant des effets de tous les autres facteurs pour aboutir à sa mesure précise (voire exacte). Or les déterminants, loin de s'additionner, se combinent et s'articulent, formant des réseaux de facteurs qu'il est vain de vouloir isoler, dans le monde social (Bourdieu, 1979 ; Ragin, 2006) comme dans le domaine biologique (Moore, 2013). De plus, il n'existe aucun moyen de s'assurer de la pureté d'un effet, autrement dit du fait que l'on a bien contrôlé l'ensemble des facteurs pertinents.

On notera enfin qu'ici l'articulation entre génétique et sociologie/démographie se limite à l'instrumentalisation des données génétiques à des fins de mesure d'effets sociodémographiques.

3. Interactions gènes-environnement (GxE)

Pour articuler biologie et sociologie, les sociogénomistes soulignent très souvent l'intérêt majeur d'étudier les interactions entre gènes et environnement (*GxE effects*), afin d'analyser comment les effets sociaux sont médiés par les prédispositions génétiques (Guo *et al.*, 2008 ; Conley, 2016 ; Freese, 2018). Certains sociogénomistes ont d'ailleurs fait un effort de théorisation des interactions entre gènes et environnement en conceptualisant différents idéaux-types d'interactions (Shanahan et Hofer, 2005 ; Boardman *et al.*, 2013). Ainsi, le modèle d'activation contextuelle (*diathesis-stress* ou *contextual triggering*) suppose une prédisposition génétique qui reste à l'état latent jusqu'à son activation par un facteur environnemental, souvent négatif. Le modèle de susceptibilité différentielle suppose que certains individus sont plus sensibles que d'autres à certains facteurs environnementaux, positifs comme négatifs. Le modèle bioécologique (dit aussi de compensation sociale ou de distinction sociale) suppose que les influences génétiques sont maximales dans des environnements stables, souvent ceux des classes supérieures, permettant aux individus d'atteindre leur potentiel génétique. Le modèle de contrôle social (ou d'impulsion sociale) suppose que les influences génétiques sont filtrées ou médiées par certains facteurs environnementaux comme les normes sociales ou les contraintes structurelles.

Cependant, on peut se demander ce qu'une reprise du terme d'« environnement » sans vigilance épistémologique produit sur la construction des objets sociologiques. Dans le modèle polygénique additif, l'environnement est tout ce qui n'est pas génétique. Dans le contexte de l'épidémiologie, l'environnement fait plutôt référence à des facteurs de risque (alimentation, pollution, etc.). Chez les sociogénomistes, l'environnement semble synonyme du « social ». La reprise du terme sans travail conceptuel conduit donc à regrouper sous un même vocable des processus sociaux d'une grande diversité.

Par ailleurs, en pratique, les sociogénomistes étudient plus souvent comment l'environnement modifie l'effet des gènes (i.e. le social pour mieux

comprendre les mécanismes génétiques)⁽²³⁾ que le contraire. *Stricto sensu*, il ne s'agit donc pas de mettre la génétique au service d'une meilleure compréhension des phénomènes sociaux.

Enfin, l'objectivation des interactions entre gènes et environnement est généralement très simpliste :

- « Les gènes » sont soit un gène unique (gène candidat) soit un agrégat de gènes (PGS).
- L'environnement est représenté par une (parfois quelques) variable(s), supposée(s) résumer une réalité complexe (environnement à risque, milieu familial, etc.).
- Pratiquement, l'interaction est définie par un terme d'interaction dans un modèle de régression.

Or, d'un point de vue strictement statistique, la mise en évidence d'interactions GxE souffre de nombreuses limites, comme le manque de puissance statistique ou la sensibilité aux échelles de mesure de l'environnement et aux facteurs de confusion (Domingue *et al.*, 2020 ; Mills et Tropf, 2020). Mais surtout, on ne modélise jamais des interactions complexes, d'ordres élevés, dynamiques, rétroactives, etc. La présence ou l'absence d'interactions statistiques entre gènes et environnements ne nous apprennent rien sur les interactions réelles, « physiques », entre les gènes et leurs contextes (Moore, 2018), qui sont omniprésentes et de plus en plus étudiées, notamment en épigénétique (Gottlieb, 2003). Les approches par « partitionnement de la variation » et par « élucidation des mécanismes » sont totalement distinctes, ce qui est à l'origine de bien des débats et incompréhensions depuis les origines de la génétique quantitative (Tabery, 2015).

Finalement, alors même que c'est là que se situe le cœur de l'articulation entre le biologique et le social chez les sociogénomistes, leurs travaux réduisent les interactions à l'abstraction de leur définition économétrique, ce qui, au bout du compte, conduit à reproduire un dualisme simpliste entre nature et culture.

4. Aide à la décision et politiques publiques

On notera, pour finir, que la possibilité pour l'information génétique de fournir une aide à la décision pour les politiques publiques (ciblage des populations à risque, etc.), sur le modèle de la médecine dite « personnalisée », est parfois évoquée par les sociogénomistes (Conley *et al.*, 2014 ; Baier et Van Winkle, 2020)⁽²⁴⁾.

Conley et Fletcher (2017) relaient ainsi les théories selon lesquelles l'hérédité constituerait, dans une société donnée, un indicateur de justice sociale.

(23) Comme c'est le cas avec l'exemple précédemment cité, dans lequel le lien entre la présence du gène GABRA2 et le risque de dépendance alcoolique peut se trouver modifié par le contexte social (soutien, déprivation...) (Pescosolido *et al.*, 2008).

(24) De manière cependant moins appuyée que dans d'autres spécialités de recherche, comme la génétique comportementale, notamment en ce qui concerne les politiques éducatives (Asbury et Plomin, 2013 ; Plomin, 2018).

Par exemple, une faible héritabilité de la réussite scolaire indiquerait un contexte où le potentiel génétique de la population ne se réaliserait pas (ou mal), car des facteurs sociaux seraient responsables des différences de réussite. Or on a vu les problèmes insurmontables que présente l'utilisation de l'héritabilité en génétique humaine. On peut ajouter que les sciences sociales ne manquent pas de mesures d'inégalités, qui ne nécessitent pas de recourir à des données génétiques.

Mais surtout, Conley et Fletcher (2017) défendent l'idée que l'étude des interactions entre gènes et environnements permet d'identifier des différences d'impact des politiques publiques selon les prédispositions génétiques des bénéficiaires, ce qui permettrait de mieux cibler ces politiques. Les progrès de la médecine personnalisée sont bien moins rapides que prévu, du fait notamment de la complexité de l'architecture génétique de la plupart des maladies. Les PGS sont parfois utilisés pour essayer de cibler des personnes à risque, mais en complément d'autres informations⁽²⁵⁾. Certains auteurs recommandent une grande prudence dans l'interprétation et l'utilisation des résultats⁽²⁶⁾, ainsi que la mise en balance des risques et des bénéfices dans le cas des prédictions cliniques (Rosenberg *et al.*, 2018 ; Barton *et al.*, 2019 ; Baverstock, 2019). L'apport des GWAS, surtout au regard des investissements massifs dont elles font l'objet, est très discuté, et parfois considéré comme à peu près nul (Paneth et Vermund, 2018)⁽²⁷⁾. On imagine donc sans mal que les retombées des GWAS ne soient pas plus positives pour l'étude des comportements sociaux, dont le lien avec les gènes est sans doute plus distendu et complexe. Et, de fait, le souhait de voir les PGS guider les politiques éducatives en identifiant des populations-cibles est à l'heure actuelle totalement irréaliste, en particulier parce que les performances prédictives des PGS en matière d'éducation sont bien trop faibles (Angers *et al.*, 2019).

V. Les conditions d'un développement à distance

Les limites et problèmes posés par l'articulation de la génétique et de la sociologie-démographie sont nombreux et profonds. On peut dès lors se demander quelles conséquences en tirent les sociogénomistes.

Certains travaux de sociogénomique pointent les fortes limites qui pèsent sur leurs résultats et les voies sur lesquelles ils placent leurs espoirs de dépassement de ces limites. Le pouvoir prédictif d'un PGS est faible ? On ne tardera

(25) Fachal et Dunning (2015) et Mavaddat *et al.* (2019) défendent ainsi l'usage des PGS pour améliorer le ciblage des personnes à risque de cancers et leur intégration aux programmes de prévention. Un grand programme de recherche international (<https://www.fondation-arc.org/mypebs>) explore actuellement cette question pour le cas du dépistage du cancer du sein.

(26) Du fait des limites des GWAS et des PGS décrites dans cet article.

(27) Pour des états des lieux plutôt positifs quant à l'utilité des GWAS pour la médecine, voir par exemple Hirschhorn (2009), Visscher *et al.* (2012), Visscher *et al.* (2017). Pour des constats nettement plus critiques, voir aussi Goldstein (2009), Jordan (2010), Bourgain (2014).

pas à collecter de plus grandes bases de données, ce qui améliorera les prédictions. On ne sait pas où est passée l'héritabilité manquante ? On collectera bientôt les marqueurs génétiques rares, ce qui comblera une partie du manque. Il est difficile de conclure sur la causalité des effets génétiques ? De nouvelles méthodes statistiques viendront résoudre le problème. On est donc en présence d'un acte de foi dans le progrès de la sociogénomique par l'intermédiaire des progrès techniques, sans remise en cause du modèle biologique sur lequel tout repose.

L'origine de ce hiatus repose probablement sur la « distance sociale » (Collins, 2010, p. 8-12) entre les sociogénomistes et les biologistes, qui explique l'adhésion des premiers à un modèle génétique d'arrière-garde, une épistémologie dépassée au regard des débats contemporains en biologie (Meloni, 2014), quand les seconds se montrent « plus prudents dans leurs conclusions et moins certains de l'état du savoir » (Larrègue, 2018a, p. 297). On sait maintenant que les gènes sont des ressources cellulaires parmi d'autres, et non des « codes » pour le développement, voire que les gènes n'exercent aucun effet causal (Baverstock, 2019). Le déterminisme génétique « scientifique » est lié à l'idée que le génotype et l'environnement externe ont tous deux un effet sur le phénotype par le biais de diverses interactions, et ce par les processus de développement dans l'environnement interne d'un organisme ; il s'oppose à un déterminisme génétique « dur » qui considère que les gènes seuls déterminent certains ou de nombreux traits individuels des organismes, y compris les êtres humains, exagérant ainsi grossièrement leur rôle (Aivelo et Uitto, 2015). La sociogénomique, comme d'autres spécialités s'inscrivant dans le prolongement de la génétique comportementale, a souvent tendance à pencher du côté du déterminisme dur et, de ce fait, à gaspiller temps et ressources à « chasser des fantômes » (Charney, 2013).

Cette distance sociale entre sociogénomistes et biologistes est favorisée par plusieurs facteurs. De prime abord, le coup d'entrée dans la pratique de la sociogénomique peut sembler élevé. Il ne l'est cependant pas tant que cela. Les données et les méthodes statistiques des sociogénomistes peuvent être utilisées avec un niveau de compétences minimal en génétique et en biologie moléculaire⁽²⁸⁾. Le coût d'entrée est encore abaissé par la disponibilité de manuels (Mills *et al.*, 2020), de nombreux packages informatiques et statistiques et de données « clé en mains ».

Le développement de la sociogénomique bénéficie en effet de celui des sources de données. Les bases de données génomiques les plus utilisées, comme la UK Biobank, intègrent des indicateurs sociaux (réussite scolaire, fécondité...), renvoyant le sociologue et le démographe à un rôle de consommateur de données (Larrègue, 2018a) obligé de « bricoler » (Larrègue, 2017, p. 176). Dans le même temps, quelques données d'enquêtes de sciences sociales incorporent

(28) Les compétences statistiques nécessaires sont, elles, relativement élevées, mais dans le même temps assez communes dans le contexte des sciences sociales anglo-saxonnes, où elles font même figure de droit d'entrée au pôle dominant du champ.

maintenant des informations génomiques⁽²⁹⁾, voire des scores de risque polygénique prêts à l'emploi. Trois enquêtes états-uniennes sont particulièrement utilisées par les sociogénomistes : The Wisconsin Longitudinal Study (WLS), The University of Michigan Health and Retirement Study (HRS) et, surtout, The National Longitudinal Study of Adolescent to Adult Health (Add Health) qui possède des données sur des jumeaux et des données génomiques⁽³⁰⁾.

En outre, la sociogénomique s'appuie sur une démarche déjà bien implantée dans des spécialités voisines, comme la génétique comportementale mais aussi les sciences politiques, l'économie ou la criminologie. Empiriquement, elle mobilise des approches qui se coulent facilement dans le moule de la « réalité linéaire générale » (Abbott, 1988) : les données génomiques peuvent se résumer en une variable, un score de risque polygénique, qui vient compléter les variables plus classiques dans les modèles économétriques. Les habitudes de recherche et de construction d'objet ne sont donc guère bouleversées.

Un dernier facteur favorise le développement de la sociogénomique « à distance » de la biologie. Lorsque ces travaux sont soumis dans les revues de sciences sociales, le processus d'évaluation par les pairs facilite la procédure de validation dans la mesure où la technicité des méthodes employées limite les évaluateurs potentiels au cercle des généticiens du comportement et des sociogénomistes (Larrègue, 2019).

Conclusion

La recherche avance souvent en s'appuyant sur des approximations de la réalité. Mais jusqu'à quel point un modèle peut-il être une approximation de la réalité sans devenir tout simplement faux et contre-productif ? Cette question très générale évoque d'autres débats traversant les sciences sociales « quantitatives ». Est-il pertinent de continuer à inférer des causes à partir de modèles de régression, alors même que l'on sait que c'est rigoureusement impossible (Freedman, 1991, 1997 ; Clogg et Haritou, 1997 ; Berk, 2004) ? Doit-on conserver une place centrale aux tests de signification et à leurs *p-values*, alors que les problèmes qu'ils posent sont documentés depuis près de 80 ans (Berkson, 1942 ; Poitevineau, 2004), régulièrement rappelés par les associations savantes de statistique et que des alternatives existent (Wasserstein et Lazar, 2016) ?

Dans le domaine des sciences de la nature, le modèle polygénique additif a sans aucun doute permis de faire progresser la génétique pendant un certain temps. Mais on sait maintenant depuis plusieurs décennies qu'il ne peut pas rendre compte de la complexité des phénomènes étudiés, *a fortiori* lorsqu'il

(29) Du fait de la baisse des coûts de collecte de ces informations, qui est devenue moins chère que la collecte des questionnaires.

(30) La concentration des travaux de sociogénomique sur un petit nombre de sources de données n'est pas sans conséquences sur la qualité des résultats (problèmes d'échantillonnage, indicateurs standardisés, etc.).

s'agit de phénomènes sociaux. L'essor des données génomiques a pu momentanément donner à croire que de nouveaux outils permettraient d'en corriger les problèmes, mais ces derniers ne se réduisent pas à des questions techniques. Or d'autres voies sont possibles, qui passent par de nouveaux modèles théoriques et la combinaison des approches et des outils (Bourgain *et al.*, 2007 ; Génin et Clerget-Darpoux, 2015a), et donc impliquent de dépasser le simple ajout d'une variable de prédisposition génétique dans un modèle de régression.

En proposant de dépasser l'opposition entre nature et culture, la sociogénomique s'inscrit en rupture avec les principaux paradigmes sociologiques⁽³¹⁾. L'impasse (peut-être temporaire) qu'elle constitue pour la sociologie et la démographie pourrait au moins permettre de rappeler un principe de sagesse scientifique élémentaire : dans un état donné des connaissances, des théories et des outils, il est tout simplement impossible de répondre à certaines questions (Liebersohn, 1987). Toutefois, cette impasse n'en est peut-être pas une pour la génétique et la biologie : les sociologues et démographes étant les spécialistes de la mesure des contextes et des institutions, ils sont sans doute les plus à même d'aider les sciences de la nature à modéliser l'environnement (Mills *in* BSA, 2017).

(31) Même si, à ce degré de généralité, la plupart des sociologues adhèreraient à cette proposition et que l'histoire de la sociologie – de Durkheim à Bourdieu en passant par Elias – est marquée par la réflexion sur l'articulation entre nature et culture.



ANNEXE

GLOSSAIRE

ALLÈLE : Une des versions possibles d'un même gène.

DOMINANCE : Interaction entre allèles.

ÉPIGÉNÉTIQUE : Étude des mécanismes moléculaires qui modifient l'expression des gènes sans en changer la séquence nucléotidique (ADN). Ces modifications sont réversibles, transmissibles (lors des divisions cellulaires) et adaptatives (elles varient selon les contextes).

ÉPISTASIE : Interaction entre gènes.

ÉTUDES D'ASSOCIATIONS PANGÉNOMIQUES (GENOME-WIDE ASSOCIATION STUDIES OU GWAS) : Études qui cartographient, pour chaque individu étudié, un nombre très élevé de polymorphismes nucléotidiques.

LD-PRUNING : Méthode de filtrage des polymorphismes nucléotidiques retenus dans le calcul d'un score de risque polygénique.

LOCUS (LOCI AU PLURIEL) : Localisation précise d'un gène sur un chromosome.

MARQUEUR GÉNÉTIQUE : Segment d'ADN dont l'emplacement physique (locus) est identifiable sur un chromosome.

MÉTHYLATION D'ADN : Mécanisme épigénétique utilisé par des cellules pour régler l'expression du gène.

MÉTHODE GENOME-BASED RESTRICTED MAXIMUM LIKELIHOOD (GREML) : Méthode d'estimation de l'héritabilité à partir des polymorphismes nucléotidiques recensés dans les études d'associations pangénomiques.

MOSAÏCISME (SOMATIC MOSAICISM) : Coexistence de deux ou plusieurs populations cellulaires de génotypes différents chez un même individu.

PHÉNOTYPE : En génétique, un phénotype désigne le plus souvent un caractère observable d'un organisme, tel que la taille, la couleur des yeux, etc.

POLYMORPHISME NUCLÉOTIDIQUE (SINGLE NUCLEOTIDE POLYMORPHISM OU SNP) : Forme la plus fréquente de variations génétiques (i.e. de différences entre individus) dans le génome humain (ils représentent environ 90 % de l'ensemble des variations génétiques humaines). Ils sont un type de variation de l'ADN dans lequel deux chromosomes diffèrent sur un segment donné par une seule paire de bases.

SCORE DE RISQUE POLYGÉNIQUE (POLYGENIC RISK SCORE OU PGS) : Variable quantitative qui résume la prédisposition génétique d'un individu pour un trait donné, calculée à partir des résultats d'une étude d'association pangénomique.

SÉGRÉGATION DES GÈNES : Séparation de chromosomes homologues, ceux d'origine paternelle et maternelle, au moment de la méiose, i.e. de la division cellulaire qui aboutit à la production des cellules sexuelles, pour la reproduction.

SÉQUENÇAGE ET GÉNOTYPAGE À HAUT DÉBIT : Le séquençage est une technique permettant de déterminer la succession des nucléotides qui composent l'ADN. Le génotypage vise à déterminer l'identité d'une variation génétique, à une position spécifique sur tout ou partie du génome, pour un individu ou un groupe d'individus donné. Les techniques à haut débit permettent l'analyse rapide de milliers, voire de millions de molécules d'ADN simultanément.

SOUVIS À L'EMPREINTE PARENTALE : Un gène est dit « soumis à l'empreinte parentale » quand la copie héritée de la mère et celle héritée du père ne sont pas exprimées de la même manière.

TRAIT : En génétique, un trait désigne une caractéristique spécifique d'un individu, dont on cherche généralement à étudier les prédispositions génétiques éventuelles. Il est aussi couramment utilisé en psychologie pour désigner un aspect durable de la personnalité.



RÉFÉRENCES

- ABBOTT A., 1988, Transcending general linear reality, *Sociological Theory*, 6(2), 169-186. <https://doi.org/10.2307/202114>
- ABDELLAOUI A., HUGH-JONES D., YENGO L., KEMPER K. E., NIVARD M. G. *et al.*, 2019, Genetic correlates of social stratification in Great Britain, *Nature Human Behaviour*, 3, 1332-1342. <https://doi.org/10.1038/s41562-019-0757-5>
- AIVELO T. J., UITTO A. E., 2015, Genetic determinism in the Finnish upper secondary school biology textbooks, *Nordic Studies in Science Education*, 11(22), 139-152. <https://doi.org/10.5617/nordina.2042>
- ANGERS A., KAGCLI D.-M., KOELLINGER P. D., PETRILLO M., QUERCI M., RAFFAEL B., VEN DEN EEDE G., 2019, Genome-wide association studies, polygenic scores and social science genetics: Overview and policy implications, *JRC Technical Reports*, European Commission. doi:10.2760/948992
- ASBURY K., PLOMIN R., 2013, *G is for genes: The impact of genetics on education and achievement*, John Wiley & Sons.
- BAIER T., WINKLE Z. VAN, 2020, Does parental separation lower genetic influences on children's school performance?, *Journal of Marriage and Family*, 83(3), 898-917. <https://doi.org/10.1111/jomf.12730>
- BARBAN N., JANSEN R., DE VLAMING R., VAEZ A., MANDEMAKERS J. J. *et al.*, 2016, Genome-wide analysis identifies 12 loci influencing human reproductive behavior, *Nature Genetics*, 48(12), 1462-1472. <https://doi.org/10.1038/ng.3698>
- BARTON N., HERMISSON J., NORDBORG M., 2019, Population genetics: Why structure matters, *Elife*, 8, e45380. <https://doi.org/10.7554/eLife.45380>
- BAVERSTOCK K., 2019, Polygenic scores: Are they a public health hazard?, *Progress in Biophysics and Molecular Biology*, 149, 4-8. <https://doi.org/10.1016/j.pbiomolbio.2019.08.004>
- BEARMAN P., 2008, Introduction: Exploring genetics and social structure, *American Journal of Sociology*, 114(S1), v-x. <https://doi.org/10.1086/596596>
- BENJAMIN D. J., CESARINI D., CHABRIS C. F., GLAESER E. L., LAIBSON D. I. *et al.*, 2012, The promises and pitfalls of genoconomics, *Annual Review of Economics*, 4(1), 627-662. <https://doi.org/10.1146/annurev-economics-080511-110939>
- BERG J. J., HARPAK A., SINNOTT-ARMSTRONG N., JOERGENSEN A., MOSTAFAVI H. *et al.*, 2019, Reduced signal for polygenic adaptation of height in UK Biobank, *ELife*, 8, p. e39725. <https://doi.org/10.7554/eLife.39725>
- BERK R. A., 2004, *Regression analysis: A constructive critique*, Sage. <https://dx.doi.org/10.4135/9781483348834>
- BERKSON J., 1942, Tests of significance considered as evidence, *Journal of the American Statistical Association*, 37(219), 325-335. <https://doi.org/10.1080/01621459.1942.10501760>
- BLISS C., 2018, *Social by nature: The promise and peril of sociogenomics*, Stanford University Press.

- BOARDMAN J. D., DAW J., FREESE J., 2013, Defining the environment in gene–environment research: Lessons from social epidemiology, *American Journal of Public Health*, 103(S1), S64–S72. <https://doi.org/10.2105/AJPH.2013.301355>
- BOURDIEU P., 1979, *La Distinction : Critique sociale du jugement*, Paris, Minuit.
- BOURGIN C., 2014, Génomique, de la recherche à la clinique. Questions autour d'un changement de perspective, *Médecine/Sciences*, 30, 14–17. <https://doi.org/10.1051/medsci/201430s203>
- BOURGIN C., GÉNIN E., COX N., CLERGET-DARPOUX F., 2007, Are genome-wide association studies all that we need to dissect the genetic component of complex human diseases?, *European Journal of Human Genetics*, 15(3), 260–263. <https://doi.org/10.1038/sj.ejhg.5201753>
- BOYLE E. A., LI Y. I., PRITCHARD J. K., 2017, An expanded view of complex traits: From polygenic to omnigenic, *Cell*, 169(7), 1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>
- BRILEY D. A., TROPF F. C., MILLS M. C., 2017, What explains the heritability of completed fertility? Evidence from two large twin studies, *Behavior Genetics*, 47(1), 36–51. <https://doi.org/10.1007/s10519-016-9805-3>
- BROWNING S. R., BROWNING B. L., 2011, Population structure can inflate SNP-based heritability estimates, *The American Journal of Human Genetics*, 89(1), 191–193. <http://dx.doi.org/10.1016/j.ajhg.2011.05.025>
- BROWNING S. R., BROWNING B. L., 2013, Identity-by-descent-based heritability analysis in the Northern Finland Birth Cohort, *Human Genetics*, 132(2), 129–138. <https://doi.org/10.1007/s00439-012-1230-y>
- BRYANT C., 1989, Le positivisme instrumental dans la sociologie américaine, *Actes de la recherche en sciences sociales*, 78(1), 64–74. <https://doi.org/10.3406/arss.1989.2894>
- BSA, 2017, Genetics research: Not in our nature?, *BSA Network, Magazine of the British Sociological Association*, 126, 14–21.
- BURT C. H., SIMONS R. L., 2014, Pulling back the curtain on heritability studies: Biosocial criminology in the postgenomic era, *Criminology*, 52(2), 223–262. <https://doi.org/10.1111/1745-9125.12036>
- CHABRIS C. F., HEBERT B. M., BENJAMIN D. J., BEAUCHAMP J., CESARINI D. *et al.*, 2012, Most reported genetic associations with general intelligence are probably false positives, *Psychological Science*, 23(11), 1314–1323. <https://doi.org/10.1177%2F0956797611435528>
- CHARNEY E., 2012, Behavior genetics and postgenomics, *Behavioral and Brain Sciences*, 35(5), 331–358. <https://doi.org/10.1017/S0140525X11002226>
- CHARNEY E., 2013, Still chasing ghosts: A new genetic methodology will not find the “missing heritability”, *Independent Science News*.
- CHARNEY E., 2017, Genes, behavior, and behavior genetics, *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(1-2), e1405. <https://doi.org/10.1002/wcs.1405>
- CHARNEY E., ENGLISH W., 2012, Candidate genes and political behavior, *American Political Science Review*, 106(1), 1–34. <https://doi.org/10.1017/S0003055411000554>
- CLERGET-DARPOUX F., BABRON M. C., PRUM B., LATHROP G. M., DESCHAMPS I., HORS J., 1988, A new method to test genetic models in HLA associated diseases: The MASC method, *Annals of Human Genetics*, 52(3), 247–258. <https://doi.org/10.1111/j.1469-1809.1988.tb01102.x>
- CLOGG C. C., HARITOU A., 1997, The regression method of causal inference and a dilemma confronting this method, in McKim V. R., Turner S. P. (eds.), *Causality in crisis?: Statistical methods and the search for causal knowledge in the social sciences*, University of Notre Dame Press.

- COLLINS H., 2010, *Gravity's shadow. The search for gravitational waves*, Chicago, University of Chicago Press.
- CONLEY D., 2016, Socio-genomic research using genome-wide molecular data, *Annual Review of Sociology*, 42, 275-299. <https://doi.org/10.1146/annurev-soc-081715-074316>
- CONLEY D., DOMINGUE B. W., CESARINI D., DAWES C., RIETVELD C. A., BOARDMAN J. D., 2015, Is the effect of parental education on offspring biased or moderated by genotype?, *Sociological Science*, 2, 82. DOI: 10.15195/v2.a6
- CONLEY D., FLETCHER J., 2017, *The genome factor*, Princeton University Press.
- CONLEY D., FLETCHER J., DAWES C., 2014, The emergence of socio-genomics, *Contemporary Sociology*, 43(4), 458-467. <https://doi.org/10.1177%2F0094306114539640>
- CONLEY D., RAUSCHER E., DAWES C., MAGNUSSON P., SIEGAL M. L., 2013, Heritability and the equal environments assumption: Evidence from multiple samples of misclassified twins, *Behavior Genetics*, 43(5), 415-426. <https://doi.org/10.1007/s10519-013-9602-1>
- COURGEAU D., 2017, La génétique du comportement peut-elle améliorer la démographie ?, *Revue d'études des populations*, 2, 17.
- CURTIS D., 2018, Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia, *Psychiatric Genetics*, 28(5), 85-89. doi: 10.1097/YPG.0000000000000206
- DANDINE-ROULLAND C., BELLENGUEZ C., DEBETTE S., AMOUYEL P., GÉNIN E., PERDRY H., 2016, Accuracy of heritability estimations in presence of hidden population stratification, *Scientific Reports*, 6(26471). <https://doi.org/10.1038/srep26471>
- DAVEY SMITH G., EBRAHIM S., 2003, 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease?, *International Journal of Epidemiology*, 32(1), 1-22. <https://doi.org/10.1093/ije/dyg070>
- DE VIENNE D., 2019, L'héritabilité : un concept utile, souvent dévoyé. <https://sfg.igh.cnrs.fr/DOC/MAP/Heritabilite.pdf>
- DOMINGUE B. W., BELSKY D. W., CONLEY D., HARRIS K. M., BOARDMAN J. D., 2015, Polygenic influence on educational attainment: New evidence from the National Longitudinal Study of Adolescent to Adult Health, *AERA open*, 1(3). <https://doi.org/10.1177%2F2332858415599972>
- DOMINGUE B. W., TREJO S., ARMSTRONG-CARTER E., TUCKER-DROB E. M., 2020, Interactions between polygenic scores and environments: Methodological and conceptual challenges, *Sociological Science*, 7, 465-486. <http://dx.doi.org/10.15195/v7.a19>
- FACHAL L., DUNNING A. M., 2015, From candidate gene studies to GWAS and post-GWAS analyses in breast cancer, *Current Opinion in Genetics & Development*, 30, 32-41. <https://doi.org/10.1016/j.gde.2015.01.004>
- FALCONER D. S., 1960, *Introduction to quantitative genetics*, Pearson Education India.
- FISHER R. A., 1918, The correlation between relatives on the supposition of Mendelian inheritance, *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, 52(2), 399-433.
- FREEDMAN D., 1991, Statistical models and shoe leather, *Sociological methodology*, 21, 291-313. <https://doi.org/10.2307/270939>
- FREEDMAN D., 1997, From association to causation via regression, *Advances in Applied Mathematics*, 18(1), 59-110. <https://doi.org/10.1006/aama.1996.0501>
- FREESE J., 2018, The arrival of social science genomics, *Contemporary Sociology*, 47(5), 524-536. <https://doi.org/10.1177%2F0094306118792214a>
- FREESE J., DOMINGUE B., TREJO S., SICINSKI K., HERD P., 2019, Problems with a causal interpretation of polygenic score differences between Jewish and non-Jewish respondents in the Wisconsin Longitudinal Study, SocArXiv. <https://doi.org/10.31235/osf.io/eh9tq>

- GALTON F., 1877, Typical laws of heredity, *Nature*, 15(388), 492-495.
- GÉNIN E., CLERGET-DARPOUX F., 2015a, Revisiting the polygenic additive liability model through the example of diabetes mellitus, *Human Heredity*, 80(4), 171-177. <https://doi.org/10.1159/000447683>
- GÉNIN E., CLERGET-DARPOUX F., 2015b, The missing heritability paradigm: A dramatic resurgence of the GIGO syndrome in genetics, *Human Heredity*, 79(1), 1-4. DOI:10.1159/000370327
- GOLDSTEIN D. B., 2009, Common genetic variation and human traits, *New England Journal of Medicine*, 360(17), 1696-1698. <https://doi.org/10.1056/nejmp0806284>
- GOTTLIEB G., 2001, Genetics and development, in Smelser N.J., Baltes P.B. (eds.), *International Encyclopedia of the Social and Behavioral Sciences*, Genetics, Behavior, and Society, New-York, Elsevier, 6121-6127.
- GOTTLIEB G., 2003, On making behavioral genetics truly developmental, *Human Development*, 46(6), 337-355. <https://doi.org/10.1159/000073306>
- GUO G., TONG Y., CAI T., 2008, Gene by social context interactions for number of sexual partners among white male youths: Genetics-informed sociology, *American Journal of Sociology*, 114(S1), S36-S66. <https://doi.org/10.1086/592207>
- GUO S.-W., 1999, The behaviors of some heritability estimators in the complete absence of genetic factors, *Human Heredity*, 49(4), 215-228. <https://doi.org/10.1159/000022878>
- HAWORTH S., MITCHELL R., CORBIN L., WADE K. H., DUDDING T. *et al.*, 2019, Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis, *Nature Communications*, 10(333). <https://doi.org/10.1038/s41467-018-08219-1>
- HERRNSTEIN R. J., MURRAY C. A., 1994, *The bell curve: Intelligence and class structure in American life*, New York, NY, US, Free Press.
- HIRSCHHORN J. N., 2009, Genomewide association studies—illuminating biologic pathways, *New England Journal of Medicine*, 360(17), 1699-1701. DOI:10.1056/NEJMp0808934
- JACQUARD A., 1978, L'inné et l'acquis : l'homme à la merci de l'homme, *Journal de la société française de statistique*, 119(3), 234-251.
- JANSS L., DE LOS CAMPOS G., SHEEHAN N., SORESENSEN D., 2012, Inferences from genomic models in stratified populations, *Genetics*, 192(2), 693-704. <https://doi.org/10.1534/genetics.112.141143>
- JENSEN A., 1969, How much can we boost IQ and scholastic achievement, *Harvard Educational Review*, 39(1), 1-123.
- JORDAN B., 2010, À la recherche de l'héritabilité perdue, *Médecine/Sciences*, 26(5), 541-543. <https://doi.org/10.1051/medsci/2010265541>
- JOSEPH J., 2013, The use of the classical twin method in the social and behavioral sciences: The fallacy continues, *The Journal of Mind and Behavior*, 34(1), 1-39.
- KELLER E. F., 2010, *The mirage of a space between nature and nurture*, Duke University Press.
- KEMPTHORNE O., 1978, A biometrics invited paper: Logical, epistemological and statistical aspects of nature-nurture data interpretation, *Biometrics*, 34(1), 1-23. <https://doi.org/10.2307/2529584>
- KERMINEN S., MARTIN A. R., KOSKELA J., RUOTSALAINEN S. E., HAVULINA A. S. *et al.*, 2019, Geographic variation and bias in the polygenic scores of complex diseases and traits in Finland, *The American Journal of Human Genetics*, 104(6), 1169-1181. <https://doi.org/10.1016/j.ajhg.2019.05.001>
- KUMAR S. K., FELDMAN M. W., REHKOPF D. H., TULJAPURKAR S., 2016a, Limitations of GCTA as a solution to the missing heritability problem, *Proceedings of the National Academy of Sciences*, 113(1), E61-E70. <https://doi.org/10.1073/pnas.1520109113>

- KUMAR S. K., FELDMAN M. W., REHKOPF D. H., TULJAPURKAR S., 2016b, Response to “Commentary on ‘Limitations of GCTA as a solution to the missing heritability problem’”, *bioRxiv*, 039594. <https://doi.org/10.1101/039594>
- LANDER E. S., SCHORK N. J., 1994, Genetic dissection of complex traits, *Science*, 265(5181), 2037-2048. <https://doi.org/10.1126/science.8091226>
- LARREGUE J., 2016, Sociologie d’une spécialité scientifique. Les désaccords entre les chercheurs ‘pro-génétique’ et ‘pro-environnement’ dans la criminologie biosociale états-unienne, *Champ pénal/ Penal field*, 13. <https://doi.org/10.4000/champpenal.9440>
- LARREGUE J., 2017, La criminologie biosociale à l’aune de la théorie du champ. Ressources et stratégies d’un courant dominé de la criminologie états-unienne, *Déviance et société*, 41(2), 167-201. <https://doi.org/10.3917/ds.412.0167>
- LARREGUE J., 2018a, « C’est génétique » : ce que les twin studies font dire aux sciences sociales, *Sociologie*, 9(3), 285-304. <https://doi.org/10.3917/socio.093.0285>
- LARREGUE J., 2018b, « Une bombe dans la discipline » : l’émergence du mouvement génopolitique en science politique, *Social Science Information*, 57(2), 159-195. <https://doi.org/10.1177%2F0539018418763131>
- LARRÈGUE J., 2019, Usages et enjeux en science sociales, Un siècle de Fisher. Atelier pluridisciplinaire pour une analyse critique d’un texte fondateur, Paris, 13 septembre 2019.
- LEWONTIN R. C., 1974, Annotation: The analysis of variance and the analysis of causes, *American Journal of Human Genetics*, 26(3), 400-411.
- LEWONTIN R. C., ROSE S., KAMIN L. J., 1984, *Not in our genes*, New York, Pantheon Books.
- LIEBERSON S., 1987, *Making it count: The improvement of social research and theory*, University of California Press.
- LUSH J. L., 1937, *Animal breeding plans*, Collegiate Press, Inc.
- MANOLIO T. A., COLLINS F. S., COX N. J., GOLDSTEIN D. B., HINDORFF L. A. *et al.*, 2009, Finding the missing heritability of complex diseases, *Nature*, 461(7265), 747-753. <https://doi.org/10.1038/nature08494>
- MARINI M. M., 1984, Women’s educational attainment and the timing of entry into parenthood, *American Sociological Review*, 49(4), 491-511. <https://doi.org/10.2307/2095464>
- MARTIN A. R., KANAI M., KAMATANI Y., OKADA Y., NEALE B. M., DALY M. J., 2019, Clinical use of current polygenic risk scores may exacerbate health disparities, *Nature Genetics*, 51(4), 584-591. <https://doi.org/10.1038/s41588-019-0379-x>
- MAVADDAT N., MICHAILIDOU K., DENNIS J., LUSH M., FACHAL L. *et al.*, 2019, Polygenic risk scores for prediction of breast cancer and breast cancer subtypes, *The American Journal of Human Genetics*, 104(1), 21-34. <https://doi.org/10.1016/j.ajhg.2018.11.002>
- MELONI M., 2014, Biology without biologism: Social theory in a postgenomic age, *Sociology*, 48(4), 731-746. <https://doi.org/10.1177%2F0038038513501944>
- MILLS M. C., RAHAL C., 2019, A scientometric review of genome-wide association studies, *Communications Biology*, 2(1), 1-11. <https://doi.org/10.1038/s42003-018-0261-x>
- MILLS M. C., TROPP F. C., 2020, Sociology, genetics, and the coming of age of sociogenomics, *Annual Review of Sociology*, 46(1), 553-581. <https://doi.org/10.1146/annurev-soc-121919-054756>
- MILLS M. C., BARBAN N., TROPP F. C., 2018, The sociogenomics of polygenic scores of reproductive behavior and their relationship to other fertility traits, *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 4(4), 122-136. <https://doi.org/10.7758/RSF.2018.4.4.07>

- MOORE D. S., 2006, A very little bit of knowledge: Re-evaluating the meaning of the heritability of IQ, *Human Development*, 49(6), 347-353. <https://doi.org/10.1159/000096534>
- MOORE D. S., 2013, Current thinking about nature and nurture, in Kampourakis K. (ed.), *The Philosophy of biology: A companion for educators*, Dordrecht, Springer Netherlands, History, Philosophy and Theory of the Life Sciences, 629-652.
- MOORE D. S., 2018, Gene $\times$ Environment interaction: What exactly are we talking about?, *Research in Developmental Disabilities*, 82, 3-9. <https://doi.org/10.1016/j.ridd.2018.04.012>
- MOORE D. S., SHENK D., 2017, The heritability fallacy, *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(1-2), e1400. <https://doi.org/10.1002/wcs.1400>
- MOSTAFAVI H., HARPAK A., AGARWAL I., CONLEY D., PRITCHARD J. K., PRZEWORSKI M., 2020, Variable prediction accuracy of polygenic scores within an ancestry group, in Loos R., Eisen M. B., O'Reilly P. (eds.), *eLife*, 9, e48376. doi: 10.7554/eLife.48376
- NELSON R. M., PETTERSSON M. E., CARLBORG Ö., 2013, A century after Fisher: Time for a new paradigm in quantitative genetics, *Trends in Genetics*, 29(12), 669-676. <https://doi.org/10.1016/j.tig.2013.09.006>
- NITSCH D., MOLOKHIA M., SMEETH L., DESTAVOLA B. L., WHITTAKER J. C., LEON D. A., 2006, Limits to causal inference based on Mendelian randomization: A comparison with randomized controlled trials, *American Journal of Epidemiology*, 163(5), 397-403. <https://doi.org/10.1093/aje/kwj062>
- NORTHCOTT R., 2008, Can ANOVA measure causal strength?, *The Quarterly Review of Biology*, 83(1), 47-55. <https://doi.org/10.1086/529562>
- OCDE, 2019, *L'ascenseur social en panne ? Comment promouvoir la mobilité sociale*, Paris, Éditions de l'OCDE. <https://doi.org/10.1787/bc38f798-fr>
- PANETH N., VERMUND S. H., 2018, Human molecular genetics has not yet contributed to measurable public health advances, *Perspectives in Biology and Medicine*, 61(4), 537-549. <https://doi.org/10.1353/pbm.2018.0063>
- PANOFSKY A., 2014, *Misbehaving science: Controversy and the development of behavior genetics*, University of Chicago Press.
- PANOFSKY A., 2011, Field analysis and interdisciplinary science: Scientific capital exchange in behavior genetics, *Minerva*, 49(3), 295. <https://doi.org/10.1007/s11024-011-9175-1>
- PEARSON K., 1898, Mathematical contributions to the theory of evolution, On the law of ancestral heredity, *Proceedings of the Royal Society of London*, 62, 379-387. <https://doi.org/10.1098/rspl.1897.0128>
- PESCOSOLIDO B. A., PERRY B. L., LONG J. S., MARTIN J. K., NURNBERGER J. I., HESSELBROCK V., 2008, Under the influence of genetics: how transdisciplinarity leads us to rethink social pathways to illness, *American Journal of Sociology*, 114(S1), S171-S201. <https://doi.org/10.1086/592209>
- PLOMIN R., 2018, *Blueprint: How DNA makes us who we are*, MIT Press.
- POITEVINEAU J., 2004, L'usage des tests statistiques par les chercheurs en psychologie : aspects normatif, descriptif et prescriptif, *Mathématiques et sciences humaines / Mathematics and Social Sciences*, 167. <https://doi.org/10.4000/msh.2900>
- POLDERMAN T. J. C., BENYAMIN B., DE LEEUW C. A., SULLIVAN P. F., VAN BOCHOVEN A. et al., 2015, Meta-analysis of the heritability of human traits based on fifty years of twin studies, *Nature Genetics*, 47(7), 702-709. <https://doi.org/10.1038/ng.3285>
- RAGIN C., 2006, The limitations of net-effects thinking, *Innovative comparative methods for policy analysis*, Springer, p. 13-41. DOI: 10.1007/0-387-28829-5_2

- RICHARDSON K., 2017, GWAS and cognitive abilities: Why correlations are inevitable and meaningless: GWA studies to identify genetic factors for educational achievements largely ignore underlying social structures and dynamics, *EMBO reports*, 18(8), 1279-1283. <https://doi.org/10.15252/embr.201744140>
- RICHARDSON K., JONES M. C., 2019, Why genome-wide associations with cognitive ability measures are probably spurious, *New Ideas in Psychology*, 55, 35-41. <https://doi.org/10.1016/j.newideapsych.2019.04.005>
- RISCH N., 2016, Presidential address: All in the family, or “Gee, Our Old LaSalle Ran Great”, *The American Journal of Human Genetics*, 98(3), 407-416. <http://dx.doi.org/10.1016%2Fj.ajhg.2016.02.009>
- ROBINSON G. E., GROZINGER C. M., WHITFIELD C. W., 2005, Sociogenomics: social life in molecular terms, *Nature Reviews Genetics*, 6(4), 257-270. <https://doi.org/10.1038/nrg1575>
- ROSENBERG N. A., EDGE M. D., PRITCHARD J. K., FELDMAN M. W., 2018, Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences, *Evolution, Medicine, and Public Health*, 2019(1), 26-34. <https://doi.org/10.1093/emph/eoy036>
- RUTTER M., 2002, Nature, nurture, and development: From evangelism through science toward policy and practice, *Child Development*, 73(1), 1-21. <https://doi.org/10.1111/1467-8624.00388>
- SCHMITZ L., CONLEY D., 2016, The long-term consequences of Vietnam-era conscription and genotype on smoking behavior and health, *Behavior Genetics*, 46(1), 43-58. <https://doi.org/10.1007/s10519-015-9739-1>
- SFGH, 2010, Quelle valeur accorder aux prédictions de risques pour les maladies multifactorielles ?
- SFGH, 2018, De quoi l'hérabilité est-elle l'habillage? *Manifeste*.
- SHANAHAN M. J., HOFER S. M., 2005, Social context in gene–environment interactions: Retrospect and prospect, *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 60(Special_Issue_1), 65-76. https://doi.org/10.1093/geronb/60.Special_Issue_1.65
- SHOSTAK S., BECKFIELD J., 2015, Making a case for genetics: Interdisciplinary visions and practices in the contemporary social sciences, *Genetics, Health and Society*, Emerald Group Publishing Limited, 95-125. <https://doi.org/10.1108/S1057-629020150000016004>
- SIEMENS D., 1924, *Zwillingspathologie: Ihre Bedeutung; ihre Methodik, ihre bisherigen Ergebnisse* [Twin pathology: Its meaning; its method; results so far], Berlin, Germany, Springer Verlag.
- SMITH J., 2019, Genome-wide association studies of intelligence: A review of the literature, *Authorea Preprints*. <https://doi.org/10.22541/au.156043260.01203836>
- SOHAIL M., MAIER R. M., GANNA A., BLOEMENDAL A., MARTIN A. R. *et al.*, 2019, Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies, *ELife*, 8, e39702. <https://doi.org/10.7554/eLife.39702>
- SPEED D., HOLMES J., BALDING D. J., 2020, Evaluating and improving heritability models using summary statistics, *Nature Genetics*, 52(4), 458-462. <https://doi.org/10.1038/s41588-020-0600-y>
- TABERY J., 2015, Debating interaction: The history, and an explanation, *International Journal of Epidemiology*, 44(4), 1117-1123. <https://doi.org/10.1093/ije/dyv053>
- TENESA A., HALEY C. S., 2013, The heritability of human disease: Estimation, uses and abuses, *Nature Reviews Genetics*, 14(2), 139-149. <https://doi.org/10.1038/nrg3377>

- TROPF F. C., BARBAN N., MILLS M. C., SNIEDER H., MANDEMAKERS J. J., 2015a, Genetic influence on age at first birth of female twins born in the UK, 1919–68, *Population Studies*, 69(2), 129-145. <https://doi.org/10.1080/00324728.2015.1056823>
- TROPF F. C., STULP G., BARBAN N., VISSCHER P. M., YANG J. *et al.*, 2015b, Human fertility, molecular genetics, and natural selection in modern societies, *PLoS ONE*, 10(6), e0126821. <https://doi.org/10.1371/journal.pone.0126821>
- TURKHEIMER E., PETTERSSON E., HORN E. E., 2014, A phenotypic null hypothesis for the genetics of personality, *Annual Review of Psychology*, 65(1), 515-540. <https://doi.org/10.1146/annurev-psych-113011-143752>
- UDRY J. R., 2000, Biological limits of gender construction, *American Sociological Review*, 65(3), 443-457. <https://doi.org/10.2307/2657466>
- VETTA A., 1976, Correction to Fisher's correlations between relatives and environmental effects, *Nature*, 263(5575), 316-317. <https://doi.org/10.1038/263316a0>
- VETTA A., COURGEAU D., 2003, Comportements démographiques et génétique du comportement, *Population*, 58(4), 457-488. <https://doi.org/10.3917/popu.304.0457>
- VISSCHER P. M., BROWN M. A., MCCARTHY M. I., YANG J., 2012, Five years of GWAS discovery, *The American Journal of Human Genetics*, 90(1), 7-24. <https://doi.org/10.1016/j.ajhg.2011.11.029>
- VISSCHER P. M., WRAY N. R., ZHANG Q., SKLAR P., MCCARTHY M. I. *et al.*, 2017, 10 years of GWAS discovery: Biology, function, and translation, *The American Journal of Human Genetics*, 101(1), 5-22. <https://doi.org/10.1016/j.ajhg.2017.06.005>
- WARE E. B., SCHMITZ L. L., FAUL J., GARD A., MITCHELL C., SMITH J. A., ZHAO W., WEIR D., KARDIA S. L., 2017, Heterogeneity in polygenic scores for common human traits, *bioRxiv*, 106062. <https://doi.org/10.1101/106062>
- WASSERSTEIN R. L., LAZAR N. A., 2016, The ASA's statement on p-Values: Context, process, and purpose, *The American Statistician*, 70(2), 129-133. <https://doi.org/10.1080/00031305.2016.1154108>
- YANG J., LEE S. H., GODDARD M. E., VISSCHER P. M., 2011, GCTA: A tool for genome-wide complex trait analysis, *The American Journal of Human Genetics*, 88(1), 76-82. <https://doi.org/10.1016/j.ajhg.2010.11.011>

Nicolas ROBETTE • LES IMPASSES DE LA SOCIOGÉNOMIQUE

À partir des années 2000, les technologies de séquençage et de génotypage à haut débit se développent très rapidement. On peut dès lors étudier simultanément un grand nombre de marqueurs génétiques chez un grand nombre de sujets, ce qui permet l'apparition des « études d'associations pangénomiques » et des « scores de risques polygéniques ». C'est dans ce contexte de progrès technologiques et statistiques que la « sociogénomique » – entendue comme la combinaison de la sociologie et de la génétique – apparaît et se diffuse dans le champ des sciences sociales. Or les méthodes utilisées par les sociogénomistes reposent sur un certain nombre de présupposés conceptuels et statistiques, dont la validité pose problème. Indépendamment des limites des outils utilisés, il apparaît que, à l'heure actuelle, les travaux de sociogénomique n'apportent qu'une faible contribution à la connaissance sociologique et démographique. On est le plus souvent en présence d'un acte de foi dans le progrès de la sociogénomique par l'intermédiaire des progrès techniques, sans remise en cause du modèle biologique sur lequel tout repose. De ce point de vue, l'écho des divers appels à la prudence des sociétés savantes en génétique humaine ne semble pas (encore) avoir porté jusqu'à ces « entrepreneurs de génétique » en sciences sociales.

Nicolas ROBETTE • THE DEAD ENDS OF SOCIOGENOMICS

Beginning in the 2000s, high-throughput sequencing and genotyping technologies developed rapidly. This made it possible to study many genetic markers at once in many subjects, leading to the development of genome-wide association studies and polygenic risk scores. It is in this context of technological and statistical progress that sociogenomics, understood as the combination of sociology and genetics, appeared and spread within the social sciences. But the methods used by sociogenomicists are based on several conceptual and statistical presuppositions of doubtful validity. Independently of the limitations of the tools used, the results of sociogenomics research have contributed little to sociological and demographic knowledge. Studies in sociogenomics most often express a faith in the progress of the field through technical advancements, without questioning the biological model on which the entire enterprise is based. From this perspective, the sound of multiple calls for caution from learned societies in human genetics does not seem (yet) to have reached these 'genetic entrepreneurs' in the social sciences.

Nicolas ROBETTE • LOS CALLEJONES SIN SALIDA DE LA SOCIO-GENÓMICA

A partir de los años 2000, las tecnologías de secuenciación y genotipado de alta velocidad se desarrollan muy rápidamente. Por tanto, es posible estudiar simultáneamente un gran número de marcadores genéticos en un gran número de sujetos, lo que permite la aparición de los « estudios de asociación pangénómica » y de las « puntuaciones de riesgo poligénicas ». Es en este contexto de avances tecnológicos y estadísticos que la « socio-genómica » - entendida como la combinación de la sociología y la genética - aparece y se difunde en el campo de las ciencias sociales. Ahora bien, los métodos utilizados por los investigadores en socio-genómica se basan en una serie de presupuestos conceptuales y estadísticos cuya validez plantea problemas. Independientemente de los límites de los instrumentos utilizados, resulta que en la actualidad, los trabajos de socio-genómica contribuyen escasamente al conocimiento sociológico y demográfico. En la mayoría de los casos se trata de un acto de fe en el progreso de la socio-genómica a través del progreso técnico, sin cuestionar el modelo biológico en el que todo se basa. Desde este punto de vista, el eco de las diversas llamadas a la prudencia de las sociedades científicas en genética humana no parece (todavía) haber llegado hasta estos « empresarios de genética » en ciencias sociales.

Mots-clés : sociogénomique, génétique humaine, génétique comportementale, hérédité, hérédité, études d'associations pangénomiques, nature/culture, épistémologie

Keywords: sociogenomics, human genetics, behavioural genetics, heritability, heredity, genome-wide association studies, nature/nurture, epistemology

