

Les scores polygéniques pangénomiques comme nouvelle forme de mesure de l'humain

Sonia Dheur¹ et Sven J. Saupe²

¹UMR5319 Passages CNRS – Université Bordeaux Montaigne
Maison des Suds, 12 Esplanade des Antilles, 33607 Pessac, France
sonia.dheur@cnsr.fr

²Reconnaissance du non-soi chez les champignons
Institut de Biochimie et Génétique Cellulaires
UMR5095 CNRS – Université de Bordeaux
1 rue Camille Saint Saëns, 33077 Bordeaux, France
sven.saupe@ibgc.cnsr.fr

Résumé

Chez l'homme, les études d'association pangénomiques basées sur les programmes génomiques institutionnels et commerciaux permettent aujourd'hui d'identifier des polymorphismes génétiques (SNPs) associés à des traits complexes de nature psychosociale et déterminés par des centaines de gènes différents (le niveau d'instruction, l'intelligence, des traits psychotiques, névrotiques et émotionnels...). L'enjeu immédiat n'est plus l'identification du lien causal et mécanistique qui explique la contribution d'un SNP à un trait mais la construction d'outils métrologiques nouveaux désignés *scores polygéniques pangénomiques*. Ces scores proposent d'exprimer en une valeur numérique unique la capacité d'un génome donné à produire un trait complexe. Ces scores deviennent des outils analytiques et prédictifs censés traduire la potentialité génétique d'une personne par rapport à ces traits psychosociaux et conduisent ainsi à une biologisation statistique de « l'exception humaine ».

Summary

In humans, genome-wide associations studies connected with institutional and commercial genomic programs can identify genetic polymorphisms (SNPs) that contribute to defining psychosocial complex traits determined by hundreds of genes (such as educational attainment, intelligence or

neuroticism). The goal of such studies is not so much to establish a causal link between a given genetic polymorphism and a complex trait but rather to construct novel metrological tools designated polygenic scores. These scores are proposed to be able to encapsulate in a single numerical value the genetic potential of an individual genome to contribute to a specific complex trait. These scores now represent analytical and predictive tools expressing the genetic potential of an individual towards a specific psychosocial trait and lead to a novel form of statistical biologisation of the human exception.

Auteurs

Sven Saupe est généticien. Il est directeur de recherche CNRS à l'UMR5095 IBGC. En biologie, il s'intéresse notamment à la question de la reconnaissance du non-soi chez les champignons.

Sonia Dheur est biologiste moléculaire et chercheuse au CNRS. Au sein de l'UMR5319 Passages, elle mène des analyses critiques et réflexives sur des développements récents de la recherche en biologie.

Dans un texte publié en 2000, intitulé « la vexation par les machines »¹, le philosophe Peter Sloterdijk décrit, le déroulement historique de l'*Aufklärung*, (la déclinaison germanique des Lumières), comme une série de vexations infligées à une forme collective de narcissisme anthropologique. Les premières étapes de vexation (que le philosophe reprend de Sigmund Freud) sont attribuées successivement à Copernic, qui prive l'homme « du privilège immense d'être le centre de l'univers »² (dans les termes de Goethe) ; puis au darwinisme, qui efface la limite entre l'Homme et le monde animal ; et enfin à la théorie psychanalytique freudienne, qui fait d'ego le théâtre de pulsions inconscientes. Ces étapes, cosmologique, biologique et psychologique, s'accompagnent d'une intériorisation croissante du processus de vexation narcissique anthropologique, qui devient progressivement plus intime. Sloterdijk propose aussi qu'il existe une économie de ce processus, par lequel les désagréments liés à ce dévoilement progressif sont compensés par la fierté qui découle de sa publication et de sa diffusion : « celui qui anticipe et publie les vexations peut sortir en vainqueur relatif du processus global de dégrisement – car du désavantage subi par celui qui découvre qu'il est celui qu'il est, celui qui rend le désenchantement public tire tout de même l'avantage de précéder les autres et de le leur dire en face »³. Sloterdijk estime que la phase dure des « vexations scientifiques du narcissisme anthropologique »⁴ ne commence qu'après Freud. Il propose, à la suite du biologiste de la cognition Gerhard Vollmer, plusieurs autres étapes majeures de ce processus historique de dégrisement, comme celle proposée par la sociobiologie, qui place les faits sociaux dans un cadre biologique et évolutionniste (et d'autres sur lesquels porte principalement son essai, ayant traité à une tendance techno-sociale que l'on désignerait à l'heure actuelle de transhumaniste).

Nous aimerions ici évoquer certains développements de la génétique humaine, qui n'étaient pas réalisés au moment de la rédaction du texte de Sloterdijk, mais dont nous proposons qu'ils s'inscrivent pleinement dans cette tendance historique par laquelle, comme l'écrit le philosophe, « les manifestations autonomes les plus intimes de l'existence humaine comme la créativité, l'amour et le libre arbitre disparaissent dans un marais de technologies réflexives »⁵. On décrira les développements, dans les deux à trois dernières années, des études d'association pangénomiques (ou GWAS pour *genome-wide association studies*) pour l'analyse de traits complexes en insistant en particulier sur la construction de « scores polygéniques » associés à ces traits. Ces développements signent à l'heure actuelle une intrusion nouvelle de la génétique dans la sphère sociale et politique. Si du point de vue de la génétique, on peut estimer que l'ère de l'ADN débute avec le travail d'Oswald Avery en 1944 sur le principe transformant, ces toutes dernières années marquent, nous semble-t-il, un nouveau commencement du rapport culturel de l'homme à son héritage biologique défini par son ADN.

¹ P. Sloterdijk, « La vexation par les machines. Remarques philosophiques sur la position psycho-historique de la technologie médicale avancée », in *L'heure du crime et le temps de l'œuvre d'art*, Paris, Calmann-Lévy, 2000.

² P. Sloterdijk, *op. cit.*, p. 58.

³ P. Sloterdijk, *op. cit.*, p. 47.

⁴ P. Sloterdijk, *op. cit.*, p. 48.

⁵ P. Sloterdijk, *op. cit.*, p. 52.

Richard Lewontin, biologiste dialectique marxiste et farouche opposant au réductionnisme génétique et à ses conséquences politiques sur les questions raciales et sociales, écrivait de manière optimiste, il y a une quinzaine d'années, au moment de l'achèvement du séquençage du génome humain, que maintenant que les génomes étaient disponibles, il n'y avait plus personne pour prétendre qu'on peut « calculer » l'organisme à partir de son ADN⁶. Cependant, l'évolution du paysage scientifique dans ce domaine semble à présent lui donner tort car c'est bel et bien une proposition de ce type qu'avancent certains acteurs de la génomique humaine contemporaine⁷.

Conceptuellement, les études d'association s'inscrivent dans la continuité historique de la génétique quantitative, qui se posait comme concurrente du mendélisme avant que Ronald Fisher ne montre que les deux perspectives sont en fait compatibles⁸. La génétique quantitative humaine, ou génétique des traits quantitatifs, n'est ainsi pas nouvelle et on peut la faire remonter jusqu'à Francis Galton et à son obsession métrologique. Galton, cousin de Darwin, figure polémique mais majeure de la science de l'époque victorienne, portera un intérêt constant à la question de l'héritabilité des traits humains⁹. Il introduit l'opposition « *nature* » et « *nurture* » dans une étude sur les jumeaux. Non seulement les questions fondamentales mais aussi certains des outils méthodologiques et théoriques mis en place par Galton sont encore à l'œuvre dans les études d'association contemporaines (comme le concept de corrélation ou l'utilisation de questionnaires). Dans un article de 1886, « *Regression towards mediocrity in hereditary stature* »¹⁰, il décrit une loi mathématique qui lie la taille des enfants à la taille de leurs deux parents qui statistiquement « régresse » vers la moyenne. Galton dérive une « loi » qui gouverne la transmission héréditaire des traits quantitatifs (dans ses termes : “*a [...] law that governs the hereditary transmission of, I believe, every one of those simple qualities which all possess, though in unequal degrees*”¹¹). Il présente même les plans d'une machine mécanique « *forecaster of stature* » visant à prédire la taille des enfants à partir de celle des deux parents. Galton prend en compte la complexité de la transmission des traits quantitatifs, mais marque aussi son accord avec « la doctrine générale » selon laquelle les « qualités » sont héréditaires¹².

Les sciences biologiques contemporaines embrassent ce paradigme, cette même doctrine générale (même après une prise en compte renouvelée de l'épigénétique et du rôle du microbiote).

⁶ R. C. Lewontin, *La triple hélice*, Paris, Seuil, collection Science ouverte, 2003.

⁷ R. Plomin, Robert. *Blueprint: How DNA makes us who we are*, Cambridge, Mit Press, 2019.

⁸ M. Morrison “Modelling Populations: Pearson and Fisher on Mendelism and Biometry”, *The British Journal for the Philosophy of Science*, 2002, 53(1), 39-68.

⁹ N. W. Gillham, “Sir Francis Galton and the Birth of Eugenics”, *Annual Review of Genetics*, 35(1), 2001, 83-101.

¹⁰ F. Galton, “Regression towards mediocrity in hereditary stature”, *The Journal of the Anthropological Institute of Great Britain and Ireland*, 1886, 15, 246-263.

¹¹ F. Galton, *op. cit.*, p. 246.

¹² “*This law tells heavily against the full hereditary transmission of any gift, as only a few of many children would resemble their mid- parentage. The more exceptional the amount of the gift, the more exceptional will be the good fortune of a parent who has a son who equals, and still more if he has a son who overpasses him in that respect. The law is even-handed; it levies the same heavy succession-tax on the transmission of badness as well as of goodness. If it discourages the extravagant expectations of gifted parents that their children will inherit all their powers, it no less discourages extravagant fears that they will inherit all their weaknesses and diseases. [...] Let it not be supposed for a moment that any of these statements invalidate the general doctrine that the children of a gifted pair are much more likely to be gifted than the children of a mediocre pair*” (F. Galton, *op. cit.*, p. 253-254).

Depuis Galton, le débat des rôles respectifs de la génétique et de l'environnement, *nature* et *nurture*, dans la transmission des traits complexes est largement fondé sur la comparaison de traits dans des paires de jumeaux mono- et dizygotiques. La différence de corrélation entre traits chez les paires de jumeaux monozygotiques et dizygotiques est attribuée à la composante génétique de l'hérédité. Une méta-analyse portant sur 2748 publications, analysant 17 804 traits, correspondant à 14 millions de paires de jumeaux issus de 39 pays différents, pratiquement l'ensemble du corpus d'études dites « *twin studies* » publiées entre 1958 et 2012, a été réalisée¹³. On peut noter ici au passage, une forme de mise en abyme par laquelle les polémiques associées aux approches *big data* sont elles-mêmes tranchées par une approche *big data* de type méta-analyse. Les distributions des corrélations, pour plusieurs milliers de traits, chez les jumeaux mono- et dizygotiques, martèlent ce qui est par ailleurs une évidence phénoménologique : les vrais jumeaux se ressemblent plus que les faux jumeaux. Cette étude conclut que l'héritabilité des traits humains est de 49 %. Et pour 69 % des traits étudiés, les résultats sont compatibles avec un modèle simple, où la ressemblance entre jumeaux n'est due qu'à des variations génétiques additives (la corrélation chez les monozygotes est égale au double de la corrélation chez les dizygotiques), un modèle qui minimise l'influence de l'environnement partagé. Les auteurs concluent simplement en ces termes : « nos résultats fournissent des preuves irréfutables que tous les traits humains sont héréditaires »¹⁴.

Il existe une vaste littérature critique sur les limites des *twin studies*¹⁵, mais il reste que le paradigme scientifique actuel est que de très nombreux traits complexes sont héréditaires et dus à des effets génétiques additifs. A l'ère post-génomique, on peut donc se proposer de faire plus que de poser ce constat. Ici, entrent en scène les études d'association pangénomiques (ou GWAS pour *genome-wide association studies*) qui, à l'origine, pouvaient avoir pour objectif d'établir une chaîne *causale* entre une variation génique et une variation du trait, mais qui, comme nous le verrons, ajourne (au moins temporairement) cet objectif pour en viser d'autres à plus court terme, s'inscrivant dans un régime prédictif et normatif. De manière importante, les GWAS sont par ailleurs de nature à mesurer, de façon indépendante des *twin studies*, l'influence des gènes sur les traits complexes.

Les études d'association modernes sont fondées sur l'établissement de corrélations entre des variations de la séquence du génome et la valeur du trait quantitatif d'intérêt. Dans une population, les individus diffèrent génétiquement – notamment – par des variations d'une seule paire de bases (ou SNP pour *single nucleotide polymorphism*). On en a répertorié des centaines de millions chez l'homme. Il s'agit de réunir une cohorte d'individus, de les « mesurer » pour le trait considéré, d'établir leur génotype pour plusieurs centaines de milliers, voire de millions, de SNPs, et de déterminer s'il existe une corrélation entre un SNP donné et le trait considéré, c'est-à-dire s'il est statistiquement surreprésenté dans le groupe d'individus dans lequel le trait quantitatif a une valeur

¹³ T. J. Polderman, B. Benyamin, C. A. De Leeuw, P. F. Sullivan, A. Van Bochoven, P. M. Visscher and D. Posthuma, "Meta-analysis of the heritability of human traits based on fifty years of twin studies", *Nature genetics*, 2015, 47(7), 702-709.

¹⁴ T. J. Polderman *et al.*, *op. cit.*, p. 709 (notre traduction).

¹⁵ R. C. Lewontin, S. Rose and L. J. Kamin, *Not in our genes: Biology, ideology and human nature*, London, Penguin, 1984.

élevée. Initialement, cette méthode a été mise en œuvre pour l'identification de variations génétiques prédisposant à différentes maladies somatiques et les premières études complètes datent de 2007.

Durant les premières années, une partie de la communauté scientifique s'est montrée sceptique sur la validité de l'approche, car souvent les identifications des SNPs n'étaient pas répliquées dans des populations contrôles. Il est actuellement généralement admis que ces premières études étaient sous-dimensionnées¹⁶. Cette période de crise du domaine semble aujourd'hui surmontée, car les études récentes s'appuient sur de très grandes cohortes, liées soit à des programmes génomiques institutionnels comme la *UK Biobank* britannique, soit privés, de génomique récréative comme *23andme*. Ces cohortes atteignent aujourd'hui parfois plus d'un million d'individus¹⁷. Nous voudrions ici discuter les applications de cette technique, pas tant à l'étude de maladies somatiques mais à celle, en cours de développement, de traits psychosociaux normaux ou pathologiques, tant il nous paraît que cette démarche corresponde à une nouvelle métrologie de dimensions très intimement humaines, comme la pensée, l'affectivité, les choix et les trajectoires de vies.

Les études d'association pangénomiques portant sur le niveau d'instruction sont l'exemple paradigmatique dans le domaine, mais ces dernières années ont vu la parution d'une quarantaine d'études semblables qui portent sur des traits aussi divers que l'intelligence, le comportement antisocial, les comportements à risques, les tendances suicidaires, le bien-être subjectif, le neuroticisme, des traits psychotiques ou très récemment l'orientation sexuelle¹⁸.

Une étude d'association pangénomique du niveau d'instruction publiée en août 2018 dans la revue *Nature Genetics*¹⁹ fait suite à deux précédentes études, en 2013²⁰ et 2016²¹. Chaque étude augmente la taille de la cohorte étudiée et, de façon corrélative, le nombre de SNPs identifiés. Alors que l'étude de 2013 conduite sur 126 000 individus identifie 3 SNPs et celle de 2016 menée sur 293 000 personnes en identifie 74, l'étude de 2018 trouve 1271 SNPs associés au niveau d'instruction en partant d'une cohorte de plus d'un million d'individus. De façon importante, il s'agit de saisir que l'épistémologie génétique classique, qui vise à l'identification d'un gène individuel pour mettre à jour les mécanismes sous-jacents qui régissent le lien génotype-phénotype, est ici évacuée. Il n'y a pas, dans ces études, de recherche individuelle du rôle causal des gènes associés aux SNPs. Il y en a simplement trop, qui contribuent, chacun de façon minime, à la définition du trait.

Les données issues de ces études sont exploitées de deux manières qui relèvent toutes deux d'une exploitation collective, en-masse. Le premier point concerne une recherche « statistique »,

¹⁶ P. Sullivan, "Don't give up on GWAS", *Molecular psychiatry*, 2012, 17(1), p. 2.

¹⁷ J. J. Lee, R. Wedow, A. Okbay, E. Kong, O. Maghziyan, M. Zacher, T. A. Nguyen-Viet *et al.*, "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals", *Nature genetics*, 2018, 50(8), 1112-1121.

¹⁸ A. Ganna, K. J. H. Verweij, M. G. Nivard, R. Maier, R. Wedow, A. S. Busch, A. Abdellaoui, *et al.*, "Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior", *Science*, 2019, 365, 7693-7700.

¹⁹ J. J. Lee *et al.*, *op. cit.*

²⁰ C. A. Rietveld, S. E. Medland, J. Derringer, J. Yang, T. Esko, N. W. Martin, H. J. Westra *et al.*, "GWAS of 126,559 individuals identifies genetic variants associated with educational attainment", *Science*, 2013, 340(6139), 1467-1471.

²¹ A. Okbay, J. P. Beauchamp, M. A. Fontana, J. J. Lee, T. H. Pers, C. A. Rietveld, P. Turley *et al.*, "Genome-wide association study identifies 74 loci associated with educational attainment", *Nature*, 2016, 533(7604), 539-542.

globale des mécanismes sous-jacents établissant la relation génotype-phénotype. Il s'agit de déduire quels gènes pourraient être affectés par les SNPs identifiés et de déterminer dans quel type tissulaire les gènes candidats sont préférentiellement exprimés. Cette démarche ne nécessite pas d'expérience *wetlab* supplémentaire, elle est réalisée purement *in silico* puisqu'il s'agit simplement de croiser deux types de données : des données préexistantes d'expression tissulaire avec les données de GWAS. L'étude GWAS de 2018 montre ainsi que les gènes identifiés, sur la base de leur liaison avec des SNPs corrélés avec le niveau d'instruction, sont spécifiquement enrichis dans différents types cellulaires du système nerveux à l'exclusion de la plupart des autres tissus²². Dans l'étude en question, cette approche est utilisée plutôt comme mode de validation de l'approche GWAS que dans une démarche heuristique visant à révéler des chaînes causales. S'il est difficilement contestable que les approches d'association s'inscrivent dans un régime prédictif, il est probablement excessif de dire que ces études abandonnent totalement l'approche de recherche causale. Par exemple, une étude GWAS sur les défauts optiques comme la myopie révèle de façon surprenante qu'un grand nombre de gènes prédits comme impliqués sont exprimés de manière préférentielle dans les systèmes de traitement neuronal du signal, plutôt que dans les systèmes purement optiques de l'œil. Ainsi dans ce cas au moins, l'étude présente une valeur heuristique potentielle²³.

Le second apport de ce type d'études et sur lequel nous nous attarderons principalement ici, est la construction et l'utilisation de scores polygéniques. Il s'agit d'agrèger en une seule valeur numérique l'ensemble de corrélations SNP-traités détectés dans les approches pangénomiques. On additionne les contributions de chaque SNP (homo- ou hétérozygote) en les pondérant de leur niveau de corrélation avec le trait mesuré²⁴. Dans les études pangénomiques, ces scores agrègent ainsi la contribution de milliers, voire de millions de SNPs individuels (même ceux dont le niveau de corrélation avec le trait est statistiquement très faible). L'étude de 2018 montre ainsi par exemple que pour des scores polygéniques croissants, la fraction de personnes ayant reçu un diplôme universitaire augmente, et la fraction ayant subi un redoublement dans l'éducation secondaire diminue, et ceci dans deux cohortes différentes²⁵. Robert Plomin, personnage central de l'étude pangénomique de traits psychosociaux, promeut l'utilisation de l'acronyme GPS (*genome-wide polygenic score*) pour désigner ces scores²⁶. Par cette ruse sémantique orwellienne, il fait bien de cette donnée génétique un *Global Positioning System* de chaque individu sur la carte ou dans l'espace social. De façon importante, ce score est établi sur l'agrégation de variations génétiques dont on ne connaît pas le mécanisme d'action fonctionnel, le rôle causal.

²² J. J. Lee *et al.*, *op. cit.*, p. 1115.

²³ M. S. Tedja, R. Wojciechowski, P. G. Hysi, N. Eriksson, N. A. Furlotte, V. Verhoeven, A. I. Iglesias *et al.*, "Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error", *Nature genetics*, 2018, 50(6), 834-848.

²⁴ R. Plomin and S. von Stumm, "The new genetics of intelligence", *Nature reviews. Genetics*, 2018, 19(3), 148-159.

²⁵ J. J. Lee *et al.*, *op. cit.*, p. 1116.

²⁶ R. Plomin and S. von Stumm, *op. cit.*, p. 150.

Ces scores deviennent alors des outils nouveaux qui peuvent être déployés indépendamment des immenses cohortes qui ont servi à les construire. Ils peuvent servir à mettre en évidence et à analyser l'implication d'une composante génétique dans différentes formes de stratification sociale même à partir d'échantillons de populations relativement modestes (d'environ 2000 individus). Dans un exemple très chargé du point de vue politique, la corrélation entre le score polygénique du niveau d'éducation, niveau d'éducation effectif et statut socio-professionnel a été comparée dans l'Estonie communiste et après l'indépendance ; l'étude conclut que dans un système méritocratique libéral, le score polygénique serait un meilleur prédicteur du statut social effectif que dans le système communiste²⁷. Une autre étude parue en 2018 rapporte que le score polygénique du niveau d'éducation (*EduYears*) est anti-corrélé avec la présence d'un casier judiciaire²⁸. Dans deux cohortes (comprenant cette fois moins de 1000 individus), le GPS est plus faible dans le groupe présentant un casier judiciaire que dans le groupe témoin. Ainsi ce score en vient à être présenté comme ayant une certaine valeur prédictive du comportement antisocial. On voit que s'il est nécessaire de déployer des cohortes de millions d'individus pour construire des scores polygéniques, une fois établis ceux-ci peuvent être utilisés comme outils analytiques et prédictifs sur des groupes de taille beaucoup plus modeste.

Une autre façon d'utiliser les scores polygéniques et les données GWAS est d'étudier dans quelle mesure ces scores sont corrélés pour différents traits. Cette approche met en évidence, au-delà d'une corrélation phénotypique entre les traits, une corrélation génétique qui indique donc que des traits distincts s'enracinent dans des mécanismes génétiques sous-jacents communs. Cette tendance peut être illustrée à partir de l'exemple d'une étude GWAS du comportement à risque et de la tolérance au risque²⁹. De façon absolument remarquable, le trait étudié dans cette étude génétique correspond à la réponse à une simple question d'auto-évaluation des sondés, l'une issue de l'étude *UK Biobank* et l'autre de la cohorte *23andme*. Cette approche a identifié plusieurs centaines de SNPs et met en évidence par la suite une corrélation génétique entre le score et différents traits comportementaux ou sociaux, comme la conduite en excès de vitesse ou le nombre de partenaires sexuels. Une autre étude, hollandaise, mesure la corrélation de différents scores polygéniques pour des traits de personnalité ou psychotiques avec le sentiment de solitude³⁰. Insistons ici sur le fait qu'il ne s'agit pas d'une corrélation phénotypique, ce qui est mis en évidence est une corrélation génétique.

²⁷ K. Rimfeld, E. Krapohl, M. Trzaskowski, J. Coleman, S. Selzam, P. S. Dale, T. Esko, A. Metspalu and R. Plomin, "Genetic influence on social outcomes during and after the Soviet era in Estonia", *Nature human behaviour*, 2018, 2(4), 269-275.

²⁸ J. Wertz, A. Caspi, D. W. Belsky, A. L. Beckley, L. Arseneault, J. C. Barnes., D. L. Corcoran *et al.*, "Genetics and crime: integrating new genomic discoveries into psychological research about antisocial behavior", *Psychological science*, 2018, 29(5), 791-803.

²⁹ R. K. Linner, P. Biroli, E. Kong, F. W. Meddens, F. W. R. Wedow, M. A. Fontana, M. A. Lebreton *et al.* "Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences", *Nature Genetics*, 2019, 51(2), 245-257.

³⁰ A. Abdellaoui, M. G. Nivard, J. J. Hottenga, I. Fedko, K. J. Verweij, B. M. Baselmans, E. A. Ehli *et al.*, "Predicting loneliness with polygenic scores of social, psychological and psychiatric traits", *Genes, Brain and Behavior*, 2018, 17(6), e12472.

Ces exemples illustrent comment, à travers l'usage des scores polygéniques et des études pangénomiques, on peut passer, grâce à l'étude de la base génétique d'un trait quantitatif donné, étudié sur une très grande cohorte, à une « génétisation » d'un autre trait plus spécifique à partir d'une cohorte bien plus limitée (et qui n'a pas – encore – été étudié par une approche de type GWAS ou *twin studies*). Dans nos exemples, la conduite en excès de vitesse, le sentiment de solitude, le casier judiciaire, etc... acquièrent par ce biais une composante génétique. Notons que la façon dont les phénotypes initiaux sont définis dans les études GWAS à grande cohorte est absolument critique, dans la mesure où les GPS issus de ceux-ci sont ensuite impliqués dans d'autres études. Comme mentionné plus haut, ce qu'on appelle GPS du comportement à risque est formellement le GPS d'une réponse à une unique question d'auto-évaluation du sondé de son rapport à la prise de risque.

Ce paradigme de « génétisation » des traits psychosociaux doit être replacé dans une perspective évolutive de succession des générations. Si ces traits ont une base génétique, ils sont soumis aux principes de l'évolution darwinienne. Une étude islandaise portant sur plus de 100 000 personnes, soit une fraction très significative de la population totale, établit qu'un score polygénique élevé du niveau d'instruction est corrélé avec un âge reproductif plus avancé et un nombre plus faible d'enfants³¹. Cet effet reste significatif après correction pour le niveau d'éducation effectif. Il existerait ainsi une base génétique commune entre la disposition génétique à l'instruction et les choix reproductifs. En termes de génétique des populations, la valeur sélective des individus avec un GPS élevé est plus faible que celui des individus avec un GPS faible, et on observe en effet une diminution progressive du GPS en fonction de l'année de naissance dans la population islandaise. En d'autres termes, il y a une sélection négative sur les gènes associés à un niveau d'instruction élevé. Les mêmes auteurs islandais décrivent également une notion émergente complexe, celle de « *genetic nurture* », l'environnement génétique, basé sur le calcul d'un score polygénique du niveau d'instruction construit à partir de la part génétique non-transmise des parents aux enfants³². Ce score polygénique est aussi prédictif du niveau d'instruction des enfants. Le score polygénique basé sur les gènes non-transmis a un effet sur le niveau d'instruction du sondé qui correspond à environ 30 % de l'effet du score polygénique basé sur les gènes transmis. Dit autrement, même les gènes que les parents ne transmettent pas à leurs enfants influencent leur niveau d'instruction.

Comme toute étude génétique, les GWAS mettent en relation phénotypes et génotypes. Au-delà de l'acquisition des données de génotype qui, comme nous l'avons vu, se généralise à travers des programmes médicaux institutionnels ou l'usage privé de plateformes de génomique récréative, l'élément clef est la collecte d'informations personnalisées sur les traits et les comportements psychosociaux, les phénotypes. Cette collecte devient paradoxalement le goulot d'étranglement de ce

³¹ A. Kong M. L. Frigge, G. Thorleifsson, H. Stefansson, A. I. Young, F. Zink, G. A. Jonsdottir *et al.* "Selection against variants in the genome associated with educational attainment", *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(5), E727-E732.

³² A. Kong, G. Thorleifsson, M. L. Frigge, B. J. Vilhjalmsón, A. I. Young, T. E. Thorgeirsson, S. Benonisdottir *et al.* "The nature of nurture: Effects of parental genotypes", *Science*, 2018, 359(6374), 424-428.

type d'étude. On se rappellera que, déjà confronté au même problème de collecte de données, Galton installera, dans l'*International Health Exhibition* de 1884, son laboratoire d'anthropologie où les visiteurs après avoir payé une contribution de trois pence, se voient « mesurés sous toutes les coutures »³³. On retrouve dans ce projet de Galton des aspects communs avec certaines pratiques génomiques actuelles : l'argument sanitaire, la contribution financière du sujet à sa métrologie (il paye pour se connaître lui-même) et sa participation à la constitution d'un savoir scientifiquement institué.

On peut citer deux initiatives actuelles de collecte de phénotypes. *Genes for good*³⁴ est un programme porté par l'Université de Michigan, qui développe des questionnaires sous la forme d'une application de téléphone portable, renseignant sur les traits et les comportements les plus divers des participants qui, s'ils répondent à un nombre suffisant de questions, se voient envoyer un kit de génotypage à partir d'un échantillon salivaire, et on leur donne accès à leur génotype ainsi déterminé. Autrement dit, en échange de leurs données personnelles (phénotypiques et génétiques), le service de génotypage classiquement facturé par d'autres plateformes environ 50 à 100 \$, devient « gratuit ». Bien plus ambitieuse est l'approche de Yaniv Erlich³⁵, un des grands animateurs de la génomique récréative qui propose de dépasser la pénibilité et la lenteur de l'approche par questionnaire numérique, par un croisement direct des données génomiques et numériques. Il est le créateur d'une plateforme nommée *DNA.Land* sur laquelle les utilisateurs qui ont payé pour leur génotypage auprès d'autres opérateurs déposent leurs données génétiques³⁶. Il espère que ces mêmes utilisateurs partagent également leurs données numériques, notamment leurs données de navigation internet, pour permettre une nouvelle mise en relation entre génétique et comportement ; par exemple, en dérivant des traits de personnalité dans le modèle *big five* à partir de « likes » Facebook. Dans un article publié dans *Nature Genetics* en 2018, il écrit : « nous projetons de réduire pour nos participants la charge associée à la récolte de leurs données phénotypiques. Les procédures actuelles sont pénibles, ne sont pas faciles à implémenter, car elles nécessitent des visites répétées sur notre site web. Les dernières années ont mis en lumière l'avènement de phénotypes numériques, un terme se rapportant à la quantification de phénotypes issus de l'interaction des humains avec les technologies numériques. Des études récentes montrent que différents traits peuvent être mesurés à partir de données d'activités sur le réseau »³⁷. Pour finalement conclure : « Notre but ultime est de créer une biobanque digitale qui intègre des flots de données génomiques, généalogiques et issues des média sociaux. Notre société riche en données offre des opportunités croissantes de recueillir les données de ressources existantes et nous pensons que la valeur et la portée de telles approches intégratives continuera de croître »^{38,39}.

³³ F. Galton, "On the Anthropometric Laboratory at the late International Health Exhibition", *Journal of the Anthropological Institute*, 1885, 14, 205-218.

³⁴ <https://genesforgood.sph.umich.edu/> (consulté le 1^{er} août 2019).

³⁵ www.myheritage.fr/management/yaniv_erlich (consulté le 1^{er} août 2019).

³⁶ <https://dna.land/> (consulté le 1^{er} août 2019).

³⁷ J. Yuan, A. Gordon, D. Speyer, R. Aufrichtig, D. Zielinski, J. Pickrell and Y. Erlich, "DNA.Land is a framework to collect genomes and phenomes in the era of abundant genetic information", *Nature Genetics*, 2018, 50(2), 160-165, p. 165 (notre traduction).

³⁸ J. Yuan *et al.*, *op. cit.*, p. 165 (notre traduction).

On conclura qu'à l'heure actuelle c'est sans doute dans le domaine de la psychologie et de la psychiatrie que les scores polygéniques font l'intrusion la plus marquée^{40,41}. Elles établissent l'existence de bases génétiques communes à divers désordres psychiatriques majeurs, mais aussi entre traits de personnalité (dans le cadre théorique des *big five*) et désordres psychiatriques. Il est possible que dans l'avenir, les scores polygéniques deviennent des outils diagnostiques « objectifs » qui viendront modifier ou compléter la définition des tableaux cliniques pour les désordres psychiatriques majeurs. Dans le domaine de la psychologie, les études GWAS signent bien une nouvelle attaque contre l'édifice psychanalytique. Dans la bouche de certains acteurs comme Robert Plomin, les succès de la sociogénomique actuelle sonnent comme une revanche contre un contexte politique des années 60-70 condamnant fondamentalement la notion de déterminisme génétique : « les articles de génétique – je veux dire, vraiment, dans les années 70, c'était *verboten*. Tout était environnemental. Même la schizophrénie était causée par ce que votre mère avait fait dans les premières années de la vie. Ça semble ridicule maintenant, mais c'était l'orthodoxie à l'époque. Et mentionner la génétique, ça dépassait simplement les bornes »⁴².

Après les dérives eugénistes de la Seconde Guerre Mondiale, on avait fait le choix éthique de poser un voile sur les études génétiques des comportements visant des groupes humains. A la place, on a préféré se concentrer sur la santé humaine, les applications médicales venant dès lors justifier les études génomiques et les GWAS, avec une approche à l'échelle individuelle. Des préoccupations médicales, en passant par les maladies mentales, on en est venu progressivement aux préoccupations socio-éducatives, aux intérêts pour les modes de vie, les préférences tous domaines confondus. Il est notable qu'un grand nombre de ces études n'émanent à l'heure actuelle pas de départements universitaires de biologie, ou de génétique, mais relèvent directement de départements des sciences sociales⁴³. Les travaux en questions sont potentiellement hautement performatifs du point de vue social. Certains auteurs appellent directement à l'usage de scores polygéniques dans le domaine éducatif notamment, ou à la prise en compte de l'héritabilité du niveau d'instruction dans les

³⁹ Ces lignes résonnent avec le texte maintenant vieux de 40 ans de Jean-François Lyotard, *La Condition postmoderne*, dans lequel Lyotard propose que la science post-moderne délégitimée par le naufrage des grands métarécits se re-légitime par sa performativité. Il écrit au sujet du savoir dans les sociétés informatisées : « l'ancien principe que l'acquisition du savoir est indissociable de la formation (*Bildung*) de l'esprit, et même de la personne, tombe et tombera davantage en désuétude. Ce rapport des fournisseurs et des usagers de la connaissance avec celle-ci tend et tendra à revêtir la forme que les producteurs et les consommateurs de marchandises ont avec ces dernières, c'est-à-dire la forme valeur. Le savoir est et sera produit pour être vendu, et il est et sera consommé pour être valorisé dans une nouvelle production: dans les deux cas, pour être échangé. Il cesse d'être à lui-même sa propre fin, il perd sa « valeur d'usage »³⁹. Ou encore : « dans cette transformation générale, la nature du savoir ne reste pas intacte. Il ne peut passer dans les nouveaux canaux, et devenir opérationnel, que si la connaissance peut être traduite en quantités d'informations. On peut donc en tirer la prévision que tout ce qui dans le savoir constitué n'est pas ainsi traduisible sera délaissé, et que l'orientation des recherches nouvelles se subordonnera à la condition de traduisibilité des résultats éventuels en langage de machine » (J. F. Lyotard, *op. cit.*, p. 13).

⁴⁰ J. M. Fullerton and J. I. Nurnberger, "Polygenic risk scores in psychiatry: Will they be useful for clinicians?", *F1000Research*, 31, 8.

⁴¹ J. S. Anderson, J. Shade, E. DiBlasi, A. A. Shabalin, A. R. and Docherty, "Polygenic risk scoring and prediction of mental health outcomes", *Current Opinions in Psychology*, 2019, 27, 77-81.

⁴² <https://www.theguardian.com/science/2018/sep/29/so-is-it-nature-not-nurture-after-all-genetics-robert-plomin-polygenic-testing> (consulté le 1er août 2019, notre traduction).

⁴³ D. W. Belsky, B. W. Domingue, R. Wedow, L. Arseneault, J. D. Boardman, A. Caspi, D. Conley *et al.*, "Genetic analysis of social-class mobility in five longitudinal studies", *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(31), E7275-E7284.

politiques visant à favoriser la mobilité sociale. Un champ modifié s'offre à la philosophie politique pour réfléchir l'articulation de ce paradigme avec les attentes sociales contemporaines. Historiquement, le fait de considérer que de nombreux traits psychosociaux ont une composante génétique se rattache généralement à des courants libéraux, de droite et parfois racistes et fascistes. Certains proposent cependant qu'à présent ces données scientifiques doivent être prises en compte pour adapter les modalités de l'action publique à certains choix politiques visant à promouvoir la mobilité sociale⁴⁴.

Ce cours aperçu doit permettre d'apprécier le contexte épistémologique dans lequel s'inscrivent ces études, la visée prédictive y domine au détriment du mode explicatif. Ces études sont ainsi au diapason de notre contemporanéité, où les mots « probabilité », « prédisposition », « risque », « prédiction »... ont pris la place d'« explication », « cause » ou « origine ». À un régime d'historicité moderne tourné vers un progrès à venir promettant un savoir complet fait place un régime présentiste où compte seule la connaissance dont on a un usage immédiat pour appréhender, maîtriser – dans le présent – ce qui se passe et ce qui risque d'arriver⁴⁵. Voilà ce qu'un tel fera avec un tel génome. Le paradoxe, cependant, tient à ce que ces prédictions ont une base génétique, naturelle, innée et qu'elles pointent donc douloureusement une forme d'impuissance de l'homme. Celui-ci ne peut se résoudre à y réagir simplement avec humilité. Ces études génomiques sont dans cette mesure inscrites dans la série de vexations narcissiques anthropologiques et dans leur économie, évoquées par Sloterdijk. La promotion de l'usage de ce type de métrologie génétique pourrait être, si on suit la perspective du philosophe allemand, une façon pour les diffuseurs de ces concepts d'échapper partiellement à la vexation qu'ils véhiculent. La vexation génomique porte sur des éléments centraux de notre être-humain et, à travers le déploiement des scores polygéniques, elle s'insinue jusque dans des détails spécifiques de notre comportement. Sloterdijk parle de vexation par les machines et pense qu'*in fine* toutes ces vexations sont fondées sur l'équivalence entre l'homme et la machine. Et en effet, la perspective génomique conduit à une description de l'homme comme machine organique programmée génétiquement. A l'ère de l'intelligence artificielle, une part conséquente des études de GWAS se focalise sur l'intelligence humaine. Le fait que les machines informatiques simulent aujourd'hui spontanéité, originalité et créativité rend, selon Sloterdijk, l'équation homme = machine plus facile à endosser. Peut-être est-ce pour cette raison que l'homme accepte de se pencher de cette façon sur les aspects machiniques de son esprit. La question du « propre de l'homme » telle que soulevée par les études de génomique aujourd'hui se pose peut-être moins au regard de l'animal, que par rapport à la machine.

Certains aspects de la génomique humaine mettent en avant la personnalisation, on pense en particulier à la fameuse promesse de la médecine personnalisée ou à l'argument publicitaire du « connais-toi toi-même » des entreprises de génomique récréative. Mais la triangulation des

⁴⁴ D. W. Belsky *et al.*, *op. cit.*

⁴⁵ F. Hartog, *Régimes d'historicité. Présentisme et expérience du temps*, Paris, Le Seuil, 2003.

dimensions humaines intimes par une poignée de scores polygéniques définissant, ici la santé mentale, là l'intelligence ou la sociabilité, constitue un gigantesque processus de réduction de la complexité qui d'une certaine façon est peut-être l'horizon paradoxal de toutes les approches *big data*. Le sentiment trouble, la vexation pour reprendre Sloterdijk, qui accompagne l'émergence de ces scores qui sont censés nous décrire, est peut-être véhiculé de façon juste par cette phrase du récit *Acqua alta* du prix Nobel de littérature Joseph Brodsky qui écrit, en parlant des miroirs des chambres d'hôtels à Venise : « ce qu'ils vous renvoient n'est pas votre identité mais votre anonymat »⁴⁶.

⁴⁶ J. Brodsky, *Acqua alta*, Paris, Gallimard, collection Arcades, 1992, p. 23.