



Commentary

Harnessing genomics to identify environmental determinants of heritable disease

Carole Lyn Yauk^{a,*}, J. Lucas Argueso^b, Scott S. Auerbach^c, Philip Awadalla^d, Sean R. Davis^e, David M. DeMarini^f, George R. Douglas^a, Yuri E. Dubrova^g, Rosalie K. Elespuru^h, Thomas W. Gloverⁱ, Barbara F. Hales^j, Matthew E. Hurler^k, Catherine B. Klein^l, James R. Lupski^m, David K. Manchesterⁿ, Francesco Marchetti^a, Alexandre Montpetit^j, John J. Mulvihill^o, Bernard Robaire^j, Wendie A. Robbins^p, Guy A. Rouleau^d, Daniel T. Shaughnessy^c, Christopher M. Somers^q, James G. Taylor VI^r, Jacquetta Trasler^j, Michael D. Waters^s, Thomas E. Wilsonⁱ, Kristine L. Witt^c, Jack B. Bishop^c

^a Health Canada, Canada^b Colorado State University, United States^c National Institute of Environmental Health Sciences, United States^d Université de Montréal, Canada^e National Institutes of Health, United States^f United States Environmental Protection Agency, United States^g University of Leicester, United Kingdom^h United States Food and Drug Administration, United Statesⁱ University of Michigan, United States^j McGill University, Canada^k Wellcome Trust Sanger Institute, United Kingdom^l New York University School of Medicine, United States^m Baylor College of Medicine, United Statesⁿ Children's Hospital Colorado, United States^o University of Oklahoma Health Sciences Center, United States^p University of California Los Angeles, United States^q University of Regina, Canada^r National Heart Lung and Blood Institute, United States^s Integrated Laboratory Systems, Inc, United States

ARTICLE INFO

Article history:

Received 12 July 2012

Received in revised form 17 August 2012

Accepted 19 August 2012

Available online 28 August 2012

Keywords:

Germ cell

Heritable mutation

Next generation sequencing

Copy number variants

ABSTRACT

Next-generation sequencing technologies can now be used to directly measure heritable *de novo* DNA sequence mutations in humans. However, these techniques have not been used to examine environmental factors that induce such mutations and their associated diseases. To address this issue, a working group on environmentally induced germline mutation analysis (ENIGMA) met in October 2011 to propose the necessary foundational studies, which include sequencing of parent–offspring trios from highly exposed human populations, and controlled dose–response experiments in animals. These studies will establish background levels of variability in germline mutation rates and identify environmental agents that influence these rates and heritable disease. Guidance for the types of exposures to examine come from rodent studies that have identified agents such as cancer chemotherapeutic drugs, ionizing radiation, cigarette smoke, and air pollution as germ-cell mutagens. Research is urgently needed to establish the health consequences of parental exposures on subsequent generations.

Published by Elsevier B.V.

1. Commentary

Over 12,000 genes are identified in the Online Mendelian Inheritance in Man database (OMIM: www.ncbi.nlm.nih.gov/omim/); sequence variants in these genes are associated with a

diverse array of genetic phenotypes. *De novo* mutations occur in each generation and are increasingly being recognized as contributing to a range of human disorders associated with a broad spectrum of these genes, including autism, schizophrenia, intellectual disability, and epilepsy [1–5]. Recent estimates suggest that a human genome on average contains approximately 100 loss-of-function variants, with as many as 20 exhibiting complete loss of gene function [6]. However, despite major advances in human

* Corresponding author.

E-mail address: carole.yauk@hc-sc.gc.ca (C.L. Yauk).

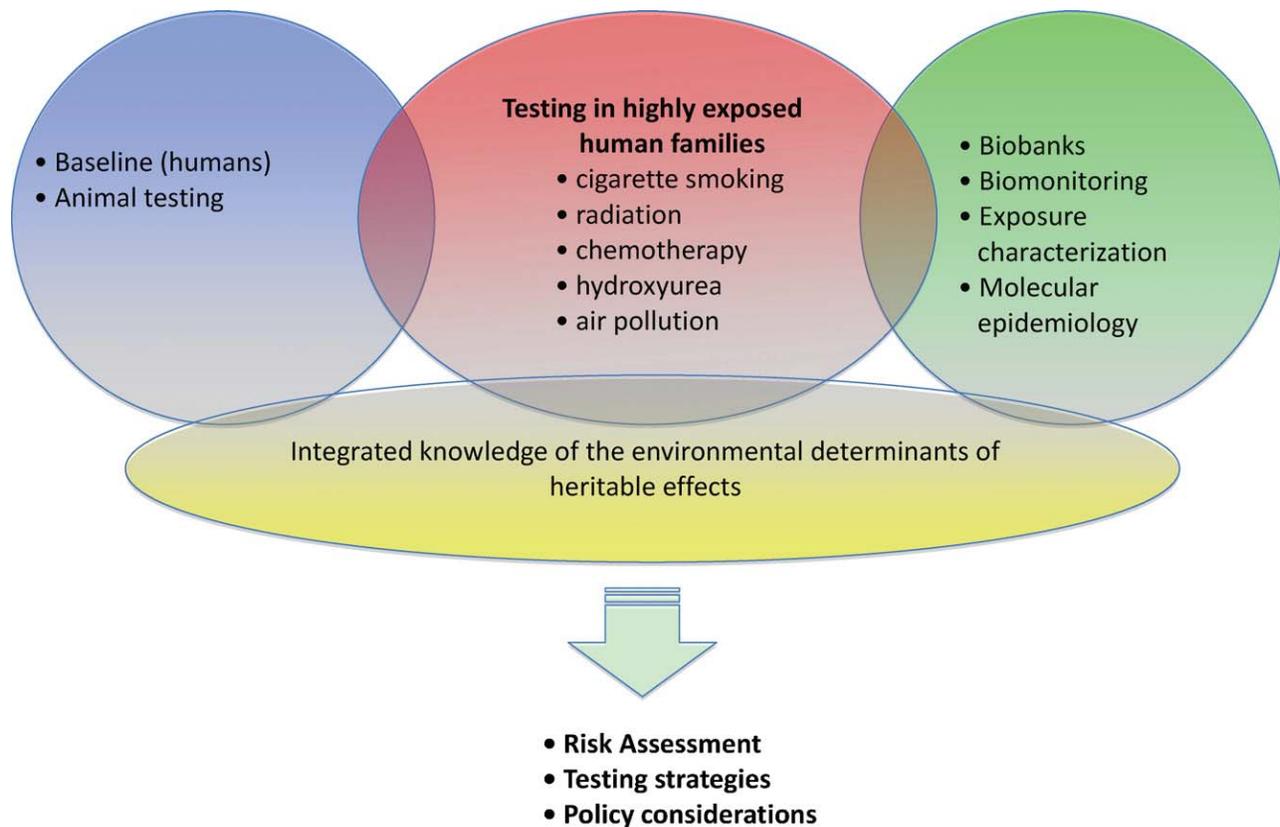


Fig. 1. Recommended approach for deciphering the environmental variables that cause germline mutation and heritable epigenetic changes in humans. Establishment of baseline genetic variance in parallel with controlled experiments in animals and on highly exposed humans will provide the foundational experiments to estimate power. Tissue repositories from well-characterized human cohorts will be developed for future studies aimed at investigating the more subtle environmental variables that influence heritable outcomes. The results of these experiments will be used to develop human health policy and testing guidelines to ensure the protection of human gametes and future generations.

genomics, the causative variables that influence mutational load have remained an enigma. Although animal models have demonstrated that exposure to mutagens in drugs, food, and the environment can cause germ-cell mutations, no agent has been shown definitively to produce heritable mutations in humans despite the similarity in gametogenic processes in humans and rodents. For example, of the 47 agents tested to date for germ cell mutagenesis, 33 chemicals, ionizing radiation, and 3 complex mixtures (urban air, and main- and side-stream cigarette smoke) cause germ-cell mutations in rodents [7–13]. It is likely that some of these agents induce germ-cell mutations in sufficiently exposed humans [14].

Identifying environmentally induced human germline mutations has been stymied by a lack of technologies that can efficiently and economically provide a high-resolution map of a large number of individual genomes [15–17]. Thus, the ENIGMA working group met recently¹ to assess the current state of knowledge in environmental mutagenesis and the potential role of environmental agents in causing germline mutations and sporadic disease. The goal was to address how genomic technologies might be deployed to identify variables that impact germline mutation rates and heritable epigenetic changes in humans. The group developed a vision for future research that focuses on studies designed to identify environmental factors that may influence *de novo* mutation rates and sporadic disease traits in individuals and populations.

¹ The Working Committee to Develop a Protocol for Detecting Environmentally Induced Genomic Damage in the Human Germline; Hilton Montréal Bonaventure, Montréal, Québec, Canada; 14 October 2011.

2. Opportunities and challenges

Two recent studies of human families directly characterized and quantified *de novo* DNA mutations derived from both male and female gametes [18,19]. Sequence and structural differences were identified by comparing whole genome sequences or copy number variants between offspring and their parents. These findings show that direct estimates of *de novo* mutation rates are comparable to indirect estimates from population studies [20,21] and suggest an average mutation rate of 1×10^{-8} per nucleotide per gamete per generation for single nucleotide variants, and 3×10^{-2} per genome per generation for copy number variants greater than 500 bp [5]. Variation in the numbers of copies of a particular region of DNA has generated great excitement not only because of its appreciable rate of occurrence, but also because of its likely impact on human health [22]. In addition to sequencing parents and their offspring, emerging technologies may soon permit quantification of DNA mutation frequencies directly in sperm [23], which will open up new avenues of research in this field. The power of such pedigree and sperm analyses seems clear, presenting tremendous new opportunities for research on the factors influencing mutation rates, such as age, sex, and social and environmental factors. In particular, the study of environmental factors contributing to variation in mutation frequency is a key opportunity. Human cohorts exposed to xenobiotic agents were identified as priority study subjects because extensive animal data, and some human data, support the contention that these exposures should induce heritable DNA changes. Examples include individuals receiving radio- or chemotherapy before conceiving children [12,24], inhabitants of radioactively contaminated areas (e.g., near

Table 1

Description of the current challenges associated with applying next-generation sequencing to human families, and how inbred animals can resolve some of these issues.

Humans	Inbred mice
High inter-individual variability will require large sample size to find an exposure effect	Decreased inter-animal variability will yield increased statistical power
Heterogeneous environmental exposures.	Controlled dose–response experiments can be used to derive an estimate of the environmental levels required for an effect to be observed
Confounding lifestyle variables	Controlled diet and environment
Limited tissue availability	Can sample stable non-proliferating tissues (e.g., brain); lower probability of somatic mutations
High error rates for sequencing technologies that are difficult to decipher from existing heterozygosities	Errors may be more readily identified in inbred animals; homozygotes

Chernobyl) [25], cigarette smokers [8,26,27], sickle cell anemia patients treated with hydroxyurea [28], and populations exposed to elevated levels of air pollution [11,29].

A critical first step to advance human germ-cell biology will be to precisely establish the baseline of human germline genetic and epigenetic variability. The range of variation in mutation frequency among gametes and individuals is quite large [18] and represents the background against which any induced mutations must be assessed. Consequently, it is difficult to estimate the needed power of studies over a range of anticipated effect sizes.

An additional challenge is minimizing the error rates associated with next-generation sequencing that can lead to an excess of false positives [17]. However, adequate genomic coverage and sample size, in combination with statistical modeling and the increasing ease with which mutations can be validated greatly diminish the number of false positives [19]. In addition, new technologies are emerging that may overcome this limitation by substantially reducing sequencing error rates [30]. Thus, technological advances make quantifying rates and spectra of inherited mutations in the offspring of exposed individuals an achievable goal.

3. Vision and recommendations

The technologies necessary to discover the complex variables that influence germline genome stability are largely available [17]. Given this premise, the ENIGMA working group recommends a tiered plan comprised of a combination of foundational studies to accurately characterize background variability, and definitive studies designed to determine the environmental impacts on germline mutation and epigenetic variation (Fig. 1). Foundational experiments must directly determine background genomic variability to identify differences between offspring and their parents in humans, and use controlled animal experiments to estimate anticipated effects sizes. These foundational studies are necessary to determine the required sample sizes and statistical approaches for pilot and definitive studies in humans. Pilot experiments should investigate highly exposed human cohorts to fully characterize dose responses. Animal models, which allow better control of genetic and environmental variables compared with humans, should be employed to strengthen causal inference in human health effects research (Table 1). Definitive studies of suspect germline mutagens must first establish biological sample repositories from human families across a broad range of potential mutagenic exposures. These studies should be prospective with well-documented epidemiological data collection, including standard epidemiological metrics, family medical history, exposure history, and samples for analyses of relevant exposure biomarkers. Definitive studies should also aim to include epigenomics, creating an integrated understanding of the epigenetic, genetic, and environmental factors explaining heritable human disease.

The time is right to test the hypothesis that environmental exposures induce *de novo* heritable mutations in humans. This information is critical to determining the validity of extrapolations from animal models, which have already identified a number of

germ-cell mutagens. These studies will determine if environmental agents that are well known to impact health in exposed individuals also cause germline mutations that affect both individual and population health in unexposed, future generations.

Author contributions

All authors listed are members of the ENIGMA working group, attended the ENIGMA workshop in October 2011, and participated in the round-table discussions. JBB was chair of the workshop. JJM, CLY, DMD, KLW and FM were on the organizing committee for the workshop. CLY drafted the manuscript, and all authors edited the manuscript. All authors approved the final manuscript.

Conflict of interest statement

There are no conflicts of interest for any of the authors.

Acknowledgements

We thank NIH/NIEHS, NIH/NCI and Health Canada for support for this workshop and for the logistical help of the Environmental Mutagen Society, under whose auspices the workshop was organized. This manuscript was reviewed by Health Canada, the National Health and Environmental Effects Research Laboratory, the United States Environmental Protection Agency, the United States Food and Drug Administration, the National Institute of Environmental Health Sciences, National Institutes of Health, and was approved for publication. Approval does not signify that the contents reflect the views of any agency or of the United States government, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

References

- [1] L.E. Vissers, J. de Ligt, C. Gilissen, I. Janssen, M. Stehouwer, P. de Vries, B. van Lier, P. Arts, N. Wieskamp, M. del Rosario, B.W. van Bon, A. Hoischen, B.B. de Vries, H.G. Brunner, J.A. Veltman, A *de novo* paradigm for mental retardation, *Nat. Genet.* 42 (2010) 1109–1112.
- [2] E.M. Morrow, Genomic copy number variation in disorders of cognitive development, *J. Am. Acad. Child Adolesc. Psychiatry* 49 (2010) 1091–1104.
- [3] J.R. Lupski, New mutations and intellectual function, *Nat. Genet.* 42 (2010) 1036–1038.
- [4] S. Girirajan, C.D. Campbell, E.E. Eichler, Human copy number variation and complex genetic disease, *Ann. Rev. Genet.* (2010).
- [5] D.F. Conrad, D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T.D. Andrews, C. Barnes, P. Campbell, T. Fitzgerald, M. Hu, C.H. Ihm, K. Kristiansson, D.G. MacArthur, J.R. Macdonald, I. Onyiah, A.W. Pang, S. Robson, K. Stirrups, A. Valsesia, K. Walter, J. Wei, C. Tyler-Smith, N.P. Carter, C. Lee, S.W. Scherer, M.E. Hurles, Origins and functional impact of copy number variation in the human genome, *Nature* 464 (2010) 704–712.
- [6] D.G. MacArthur, S. Balasubramanian, A. Frankish, N. Huang, J. Morris, K. Walter, L. Jostins, L. Habegger, J.K. Pickrell, S.B. Montgomery, C.A. Albers, Z.D. Zhang, D.F. Conrad, G. Lunter, H. Zheng, Q. Ayub, M.A. DePristo, E. Banks, M. Hu, R.E. Handsaker, J.A. Rosenfeld, M. Fromer, M. Jin, X.J. Mu, E. Khurana, K. Ye, M. Kay, G.I. Saunders, M.M. Suter, T. Hunt, I.H. Barnes, C. Amid, D.R. Carvalho-Silva, A.H. Bignell, C. Snow, B. Yngvadottir, S. Bumpstead, D.N. Cooper, Y. Xue, I.G. Romero, J. Wang, Y. Li, R.A. Gibbs, S.A. McCarroll, E.T. Dermitzakis, J.K. Pritchard, J.C. Barrett, J. Harrow, M.E. Hurles, M.B. Gerstein, C. Tyler-Smith, A systematic survey of loss-of-function variants in human protein-coding genes, *Science* 335 (2012) 823–828.

- [7] C.D. Glen, A.G. Smith, Y.E. Dubrova, Single-molecule, PCR analysis of germ line mutation induction by anticancer drugs in mice, *Cancer Res.* 68 (2008) 3630–3636.
- [8] F. Marchetti, A. Rowan-Carroll, A. Williams, A. Polyzos, M.L. Berndt-Weis, C.L. Yauk, Sidestream tobacco smoke is a male germ cell mutagen, *Proc. Natl. Acad. Sci. U.S.A.* 108 (2011) 12811–12814.
- [9] F. Marchetti, A.J. Wyrobek, Mechanisms and consequences of paternally-transmitted chromosomal abnormalities, *Birth Defects Res. C Embryo Today Rev.* 75 (2005) 112–129.
- [10] L.B. Russell, Effects of male germ-cell stage on the frequency, nature, and spectrum of induced specific-locus mutations in the mouse, *Genetica* 122 (2004) 25–36.
- [11] C.M. Somers, Ambient air pollution exposure and damage to male gametes: human studies and in situ 'sentinel' animal experiments, *Syst. Biol. Reprod. Med.* 57 (2011) 63–71.
- [12] K.L. Witt, J.B. Bishop, Mutagenicity of anticancer drugs in mammalian germ cells, *Mutat. Res.* 355 (1996) 209–234.
- [13] C. Yauk, A. Polyzos, A. Rowan-Carroll, C.M. Somers, R.W. Godschalk, F.J. Van Schooten, M.L. Berndt, I.P. Pogribny, I. Koturbash, A. Williams, G.R. Douglas, O. Kovalchuk, Germ-line mutations, DNA damage, and global hypermethylation in mice exposed to particulate air pollution in an urban/industrial location, *Proc. Natl. Acad. Sci. U.S.A.* 105 (2008) 605–610.
- [14] D.M. DeMarini, Declaring the existence of human germ-cell mutagens, *Environ. Mol. Mutagen.* 53 (2012) 166–172.
- [15] A.J. Wyrobek, J.J. Mulvihill, J.S. Wassom, H.V. Mallng, M.D. Shelby, S.E. Lewis, K.L. Witt, R.J. Preston, S.D. Perreault, J.W. Allen, D.M. DeMarini, R.P. Woychik, J.B. Bishop, Assessing human germ-cell mutagenesis in the postgenome era: a celebration of the legacy of William Lawson (Bill) Russell, *Environ. Mol. Mutagen.* 48 (2007) 71–95.
- [16] T.M. Singer, C.L. Yauk, Germ cell mutagens: risk assessment challenges in the 21st century, *Environ. Mol. Mutagen.* 51 (2010) 919–928.
- [17] M.A. Beal, T.C. Glenn, C.M. Somers, Whole genome sequencing for quantifying germline mutation frequency in humans and model species: cautious optimism, *Mutat. Res. Rev.* 750 (2) (2012) 96–106.
- [18] J.C. Roach, G. Glusman, A.F. Smit, C.D. Huff, R. Hubley, P.T. Shannon, L. Rowen, K.P. Pant, N. Goodman, M. Bamshad, J. Shendure, R. Drmanac, L.B. Jorde, L. Hood, D.J. Galas, Analysis of genetic inheritance in a family quartet by whole-genome sequencing, *Science* 328 (2010) 636–639.
- [19] D.F. Conrad, J.E. Keebler, M.A. DePristo, S.J. Lindsay, Y. Zhang, F. Casals, Y. Idaghdour, C.L. Hartl, C. Torroja, K.V. Garimella, M. Zilversmit, R. Cartwright, G.A. Rouleau, M. Daly, E.A. Stone, M.E. Hurler, P. Awadalla, Variation in genome-wide mutation rates within and between human families, *Nat. Genet.* 43 (2011) 712–714.
- [20] M.W. Nachman, S.L. Crowell, Estimate of the mutation rate per nucleotide in humans, *Genetics* 156 (2000) 297–304.
- [21] M. Lynch, Rate, molecular spectrum, and consequences of human mutation, *Proc. Natl. Acad. Sci. U.S.A.* 107 (2010) 961–968.
- [22] J.R. Lupski, Genomic rearrangements and sporadic disease, *Nat. Genet.* 39 (2007) S43–S47.
- [23] J. Wang, H.C. Fan, B. Behr, S.R. Quake, Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm, *Cell* 150 (2012) 402–412.
- [24] J.F. Winther, J.D. Boice Jr., J.J. Mulvihill, M. Stovall, K. Frederiksen, E.J. Tawn, J.H. Olsen, Chromosomal abnormalities among offspring of childhood-cancer survivors in Denmark: a population-based study, *Am. J. Hum. Genet.* 74 (2004) 1282–1285.
- [25] Y.E. Dubrova, Long-term genetic effects of radiation exposure, *Mutat. Res.* 544 (2003) 433–439.
- [26] B. Secretan, K. Straif, R. Baan, Y. Grosse, F. El Ghissassi, V. Bouvard, L. Benbrahim-Tallaa, N. Guha, C. Freeman, L. Galichet, V. Coglianò, A review of human carcinogens – part E: tobacco, areca nut, alcohol, coal smoke, and salted fish, *Lancet Oncol.* 10 (2009) 1033–1034.
- [27] J. Laubenthal, O. Zlobinskaya, K. Poterłowicz, A. Baumgartner, M.R. Gdula, E. Fthenou, M. Keramarou, S.J. Hepworth, J.C. Kleinjans, F.J. van Schooten, G. Brunborg, R.W. Godschalk, T.E. Schmid, D. Anderson, Cigarette smoke-induced transgenerational alterations in genome stability in cord blood of human F1 offspring, *FASEB J.* (2012).
- [28] M.F. Arlt, A.C. Ozdemir, S.R. Birkeland, T.E. Wilson, T.W. Glover, Hydroxyurea induces de novo copy number variants in human cells, *Proc. Natl. Acad. Sci. U.S.A.* 108 (2011) 17360–17365.
- [29] C.M. Somers, D.N. Cooper, Air pollution and mutations in the germline: are humans at risk? *Hum. Genet.* 125 (2009) 119–130.
- [30] M.W. Schmitt, S.R. Kennedy, J.J. Salk, E.J. Fox, J.B. Hiatt, L.A. Loeb, Detection of ultra-rare mutations by next-generation sequencing, *Proc. Natl. Acad. Sci. U.S.A.* (2012).