Current Biology 24, R577–R581, June 16, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.05.021

Somatic Mosaicism and Disease

Minireview

Steven A. Frank

The large number of cell divisions required to make a human body inevitably leads to the accumulation of somatic mutations. Such mutations cause individuals to be somatic mosaics. Recent advances in genomic technology now allow measurement of somatic diversity. Initial studies confirmed the expected high levels of somatic mutations within individuals. Going forward, the big questions concern the degree to which those somatic mutations influence disease. Theory predicts that the frequency of mutant cells should vary greatly between individuals. Such somatic mutational variability between individuals could explain much of the diversity in the risk of disease. But how variable is mosaicism between individuals in reality? What is the relation between the fraction of cells carrying a predisposing mutation and the risk of disease? What kinds of heritable somatic change lead to disease besides classical DNA mutations? What molecular processes connect a predisposing somatic change to disease? We know that predisposing somatic mutations strongly influence the onset of cancer. Likewise, neurodegenerative diseases may often begin from somatically mutated cells. If so, both neurodegeneration and cancer may be diseases of later life for which much of the risk may be set by early life somatic mutations.

Introduction

For many years, human bodies were typically thought of as genetically uniform. Rare cellular mutations in cancer or other diseases were known but generally considered to be aberrations. As technology and genetic understanding slowly improved, the idea began to spread that individuals may often be genetic mosaics. Scattered observations, particularly of skin diseases (Figure 1), suggested that there may be much hidden variability within bodies. But the actual amount of mosaicism and its consequences remained unclear. Most investigators paid little attention to the subject.

Theory suggested that human bodies should, in fact, be mosaics of genetic mutations [1]. A body has about $N = 10^{14}$ cells, each of which arose from a cell division. If the somatic mutation rate for each gene per cell division is $u = 10^{-6}$, then each gene suffers approximately $Nu = 10^8$ somatic mutational events. The number of mutated cells may be higher, because a single mutational event early in development will be carried forward to all descendant cells. Although the initial theory predicted significant genetic mosaicism, it had little impact. It was hard to measure the actual level of mosaicism, and there was little evidence that widespread mosaicism might have important consequences.

Later theories then suggested, even more strongly, that somatic mosaicism could be widespread and be a very strong factor in determining the variability in disease risk between individuals [2–5]. In particular, some individuals will experience their first mutation early in development and carry many somatically mutated cells. Other individuals will have their first mutation later in development and have fewer somatically mutated cells. The inevitable variation in the degree of somatic mosaicism arising early in life could determine much of the variation in the risk of disease later in life [2–5]. If so, that would explain a lot about the unknown causes of disease. How can we know if it is true?

Only recently, it has become possible to analyze the genomes of single cells or small samples of cells. This has opened a new window onto the actual mutational diversity within individuals, and several recent studies demonstrate widespread somatic mosaicism [1,6–16]. However, the existence of somatic diversity tells us relatively little about disease. In this review, I consider what steps are needed to connect somatic diversity to disease progression and to the variation in disease risk between individuals.

What types of disease may be affected by somatic mosaicism? Certainly mosaicism may influence cancer, in which somatic mutations play a central role [17]. However, in the typical view of cancer, isolated somatic mutations occur in rare cells, and disease develops only when those isolated cells acquire multiple mutations. By contrast, a perspective of common and highly variable somatic mosaicism emphasizes that some individuals will inevitably have many mutated cells that strongly predispose to disease by increasing the number of cells that carry an initial predisposing somatic mutation [3]. The recent theory and observations on mosaicism are just starting to change the common view of the events that initiate cancer from rare isolated somatic mutations to the potentially important role of widespread somatic mosaicism.

It is also possible that somatic mosaicism influences neurodegenerative disease [1,4,6–9,12–16,18]. However, the importance of somatic mutation in neurodegeneration remains unclear. In cancer, a tumor is thought to arise from a somatically mutated cell that initiates a clonal expansion. In neurodegeneration, how would a small focus of somatically mutated cells initiate the spread of disease? Several possibilities have been discussed, but the problem remains open [4,8,18–21]. If neurodegeneration does tend to spread from a small initial focus of damaged cells, then somatic mosaicism may be a primary risk factor in neurodegenerative disease.

Direct Evidence of Mosaicism

Numerous recent genomic studies describe widespread somatic mosaicism [12,13], following early hints from patterns of skin disease (Figure 1). Those recent genomic studies provide the proof of existence for what seemed theoretically inevitable, setting the stage for future progress.

In a study of primary fibroblast cells obtained from seven different humans, Abyzov *et al.* [14] estimate "that approximately 30% of the fibroblast cells have CNVs (copy number variants) in their genomes." CNVs correspond to deletions or duplications of genomic regions that are at least a few kilobases (kb) in length. A rough theory predicts that the average frequency of cells with a somatic mutation should be approximately the number of cellular generations since the zygote stage multiplied by the mutation rate per cell division [2–4,22]. For 30 generations from an adult fibroblast back to the zygote, a 30% frequency of mutated cells is

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525, USA.
E-mail: safrank@uci.edu

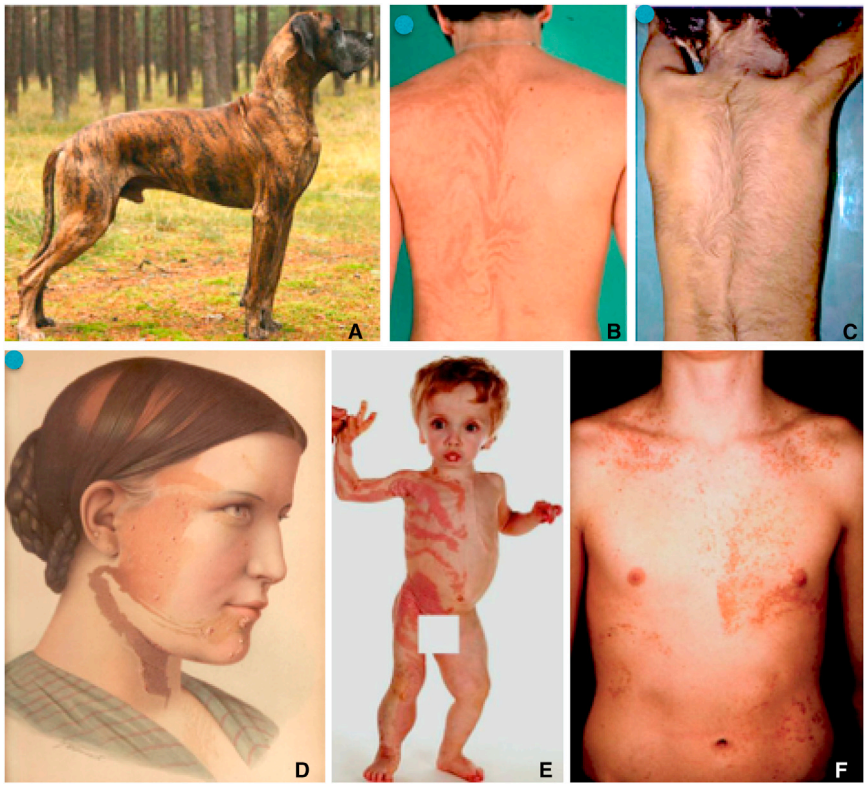

Figure 1. Examples of somatic mosaicism in skin.

These patterns illustrate the spatial distribution of altered cells associated with cellular changes that transmit to descendant cells. These extreme cases hint at the more frequent cases of mosaicism that go undetected because they occur in fewer cells or do not produce easily observed phenotypes known to arise from a mosaic cause. Several mosaic skin diseases are associated with known mutations. Other skin patterning may be caused by various epigenetic changes that transmit to daughter cells. All figures are presented and discussed in Happle [32]. (A) *Brindle* trait of dogs (reproduced from Harlis.jpg, Creative Commons). (B) Classic patterning of human skin mosaicism in which affected cells follow the lines of Blaschko. (C) Zimmermann-Laband syndrome showing typical pattern of pigment lines on the skin. (D) Systematized sebaceous nevus, which has been associated in some cases with somatic *HRAS* or *KRAS* mutations [33,34] (with kind permission from Springer Science and Business Media, from [29]). (E) Type 2 segmental PTEN hamartoma syndrome, apparently caused by a germline mutation that leads to heterozygosity at the *PTEN* locus, followed by somatic mutations that cause mosaic loss of heterozygosity in some cells [35,36] (reprinted with permission from John Wiley & Sons [36]). (F) Type 1 segmental Darier disease, associated with *de novo* somatic mutations in some cases [37,38] (with kind permission from Springer Science and Business Media, from [29]).

consistent with a mutation rate of 0.01 mutations per cell division. That rate seems reasonable or perhaps a bit conservative for the genome-wide CNV changes per cellular generation [23]. In another study, McConnell *et al.* [15] estimated that 13–41% of human frontal cortex neurons have at least one somatic CNV of a megabase or more, and O'Huallachain *et al.* [16] found widespread somatic CNVs in diverse human tissues.

These genome-wide studies of CNVs emphasize high frequency mutational events that lead to a high frequency of somatically mutated cells. When focusing on a single gene or on a class of mutations with lower frequency, we face the empirical difficulty that only a small fraction of cells are typically expected to carry a somatic mutation, even though the total number of mutant cells may be large. Using the theory mentioned above, the frequency of mutated cells would on average be approximately uG, where u is the mutation rate per cellular generation, and G is the number of cellular generations since the zygote stage. If we focus on a single gene with a mutation rate of 10^{-6} [24], and we assume that G is roughly 30, then the frequency of mutated cells is on the order of 10^{-5}. However, in a large cellular population that greatly exceeds 10^5, very many cells carry a somatic mutation. So we arrive at one of the essential empirical difficulties. It is often hard to detect low frequency mutations, but mutations in low frequency may be carried by very many cells and may be significant with regard to disease processes.

Two studies are interesting in light of this distinction between frequency and number [10,11]. These studies analyzed large chromosomal abnormalities of the sort often associated with leukemia. In cancer-free individuals under 50 years, the frequency of individuals with somatic mosaicism was estimated to be 0.23% [11] or under 0.5% [10]. The frequency of mosaic individuals rose with age, reaching 2–3% in the elderly. Among all cancer-free individuals, mosaicism increased the risk of subsequent progression to leukemia. Those studies could detect only large chromosomal abnormalities in individuals with at least 5% of cells carrying the abnormality. The vast majority of mosaicism is expected to occur at a frequency of less than 5%. Thus, many of the mosaic individuals may have had a clonal expansion of the mutant before detection.

Interpreting these studies in terms of variable mosaicism would require information about the origin of the first mutation. On the one hand, the mutation may have occurred during development, causing the mutant to be carried in early life by many descendant hematopoietic stem cells, although often in low or moderate overall frequency but typically below the 5% level required for detection. Subsequent mutations within this population of developmentally mutated cells may have led to a clonal expansion and a rise in frequency above the 5% detection level. In this case, the developmental mutation and early life mosaicism would have been the main factor in setting the risk of subsequent clonal expansion that led to a detectable level of mosaicism. On the other hand, the first chromosomal mutation may have happened after development, and the subsequent mutations that caused a clonal expansion must have occurred among the very rare mutated cells.

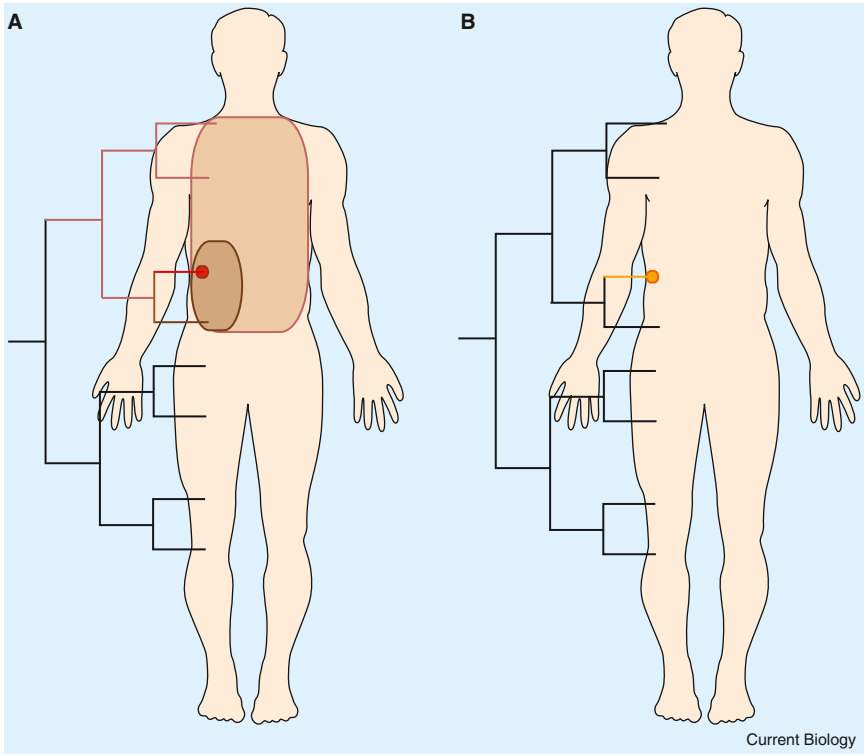

Figure 2. Variability in mosaicism between individuals.

Each figure shows a phylogenetic tree representation of the cellular lineage history within an individual, rooted at the zygote. The history represents only a particular, very small genomic region. The lighter-colored lines specify mutant cells and their descendants. (A) This individual had an early somatic mutation, which caused a high frequency of adult cells to carry the mutation, shown by the large colored area in the body. A second, later mutation carried forward to a smaller set of adult cells, and a third mutation happened in the last division before the current adult cell. The different stages of development at which the mutants arose cause a mixture of widespread and restricted mutant cell populations. (B) A different individual had only a single somatic mutation relatively late in development, causing limited mosaicism. The actual amount of mosaicism for a small genomic region must be vastly greater than illustrated. For N cell divisions and a mutation rate of u per cell division, Nu mutational events occur in the somatic history of an individual. For approximately 10^{16} cell divisions in a typical human lifetime and a mutation rate of roughly 10^{-6} for a small gene-sized genomic region, that region will experience on the order of 10^{10} mutational events over the lifetime of an individual. Most of those mutations will be in the last cell division, but some will occur early in development, causing widespread mosaicism and variability between individuals.

The classical explanation of how cancer starts usually invokes an initial mutation in relatively few stem cells. By contrast, the widespread mosaicism caused by developmental mutations is a more recently developed idea that has been increasingly invoked to explain the origin of disease [1,2,25–27]. Given the complexity of these two alternative scenarios, how can we assess the fraction of disease risk caused by developmental mutations and early life mosaicism?

Variability in the Risk of Disease

It will not be easy to connect variation in disease risk to developmental mutations that cause variation in early life mosaicism. But it is a topic worth pursuing, because much of the unexplained cause of disease in later life may trace back to early developmental mutations.

The following questions set the agenda: How variable is mosaicism between individuals? If mosaicism is common but at approximately the same level in each individual, then mosaicism cannot explain why some individuals get sick and others do not. However, theory suggests very high variability in mosaicism [2–5]. In essence, there should be a continuum between individuals for the frequency of adult cells that carry a particular mutation. Rare individuals carry the mutation in all cells by germline inheritance or by *de novo* somatic mutation in the zygote. More individuals carry a lower frequency of an inherited somatic mutation because, as the zygote divides, more cells become a target for a *de novo* somatic mutation, increasing the frequency of individuals in the population who suffer such mutations while simultaneously decreasing the associated fraction of mutated adult cells (Figure 2).

What is the relation between the fraction of cells carrying a predisposing mutation and the risk of disease? In cancer, certain germline mutations shift disease onset to earlier ages [3,17,28,29]. Often, somatic mutations of the same genes initiate the same disease, but typically cause a later age of onset. The difference in the age of onset between germline and somatic mutations matches closely the expected difference in the frequency of adult cells that are expected to carry the mutation [30]. However, the available data only allow a comparison between individuals who carry a germline mutation that is present in all adult cells and all other individuals who have a particular average frequency of cells carrying a somatic mutation. In contrast, theory suggests wide variability between individuals in the frequency of somatically mutated cells and thus a continuum of risk between individuals [2–5]. Germline mutations or initial somatic mutations in the zygote are carried forward to all adult cells and should be associated with the earliest age of onset. Increasingly later mutations in development are in a decreasing fraction of an individual's cells, which should associate with a continuous shift to later ages of disease onset.

The big questions are whether an increase in the fraction of somatically mutated cells shifts disease to an earlier age, and what fraction of disease is caused by early developmental mutations. It seems almost certain that, for cancer, the frequency of cells that carry a somatic mutation will strongly influence the age of onset. But the strength of the association will likely vary between tissues. For example, in tissues that divide relatively rarely later in life, mosaicism caused by early-life mutations may dominate risk. By contrast, in renewing epithelial tissues that divide frequently

throughout life, there may be greater opportunity for initiating mutations to occur later in life and thus a lower fraction of the total late-life disease risk may be set by early-life mutations.

For neurodegeneration or other diseases, the association between the frequency of somatically mutated cells and disease remains unknown. Neurodegeneration is a heterogeneous collection of diseases, so one must be careful about generalizing from one example to the broader problem. At present, some analogies exist between how the ages of onset for cancer and neurodegeneration may respond to inherited versus somatically acquired mutations [4,6–9,18]. In one extreme example, an early somatic mutation in the human prion protein gene occurred in nearly all adult cells [6]. That individual suffered the same early-life onset (age 46) and symptoms of Creutzfeldt-Jakob disease that are often found in people who inherit a germline mutation in the same gene. The disease similarity between cases with either a germline or somatic mutation in nearly all cells is not surprising. The interesting problem concerns the change in risk profile for individuals who carry a somatic mutation in a decreasing fraction of somatic cells.

To understand how risk changes with the frequency of mutant cells, we must understand which molecular processes connect a predisposing mutation to disease? For cancer, there is enough evidence to understand how molecular processes may connect the frequency of mutated cells to tumor initiation [3,17]. By contrast, it is not clear at present whether an association might exist between the frequency of mutated cells and neurodegeneration and, if so, what sort of molecular processes would cause that association. Many possibilities have been raised [4,6–9,18–21]. Misfolded proteins may spread by prion-like processes. The initial seeds for misfolding may first arise in a small focus of somatically mutated cells. Alternatively, changed RNAs may first arise in a few mutated cells, and then move extracellularly to alter other cells [31]. In principle, any factor that can transform a naïve cell into a producer of that same factor may potentially spread in an infectious way. If a mutated cell is at higher risk for becoming transformed to seed initiation of the infectious process, then there will be an association between the fraction of mutated cells and the risk of disease. In addition, a higher fraction of mutated cells will be associated with an earlier age of onset.

How does the lineage history of cells during development affect the relation between mutational events and mutational frequency in different tissues? In the simplest model, a whole tissue would derive from a single ancestral cell, and descendants from that ancestral cell would not contribute to other tissues. Such partitioning would lead to a smooth continuum between the number of cellular generations that a mutation occurred after the tissue-specific progenitor and the frequency of mutant cells in that tissue. However, tissues inevitably derive from mixing between cellular lineages. More mixing may reduce the variance between individuals in the frequency of mutated cells. Advancing genomic technology combined with the computational methods of lineage (phylogenetic) reconstruction will eventually provide a sense of the developmental map and the degree of lineage mixing in different tissues. Studies on whole cadavers with cell-level genomic resolution will be particularly valuable for learning about cellular lineage histories within individuals and somatic mutational diversity between individuals.

What kinds of heritable somatic change lead to disease besides classical DNA mutations? I have emphasized mutations to the primary DNA sequence. However, any somatic cellular change that transmits heritably to descendant cells will have consequences similar to a change in the DNA, including various non-genetic (epigenetic) changes. The key processes concern the rate at which changes arise in cells and the rate at which such changes may decay or revert back to the original state in descendants. I focused on DNA mutations because they provide a simple way to compare germline and somatic changes, and because most data are about DNA changes. As the technology to measure other types of change improves, we will obtain better estimates for the relative importance of different kinds of heritable somatic change in relation to disease.

Conclusion

In summary, a human body contains about 10^{14} cells and produces about 10^{16} cells over the course of a lifetime. That huge population size probably exceeds the total number of primates that have ever existed. The scope for diversity within bodies is great. Until recently, technology limited our ability to see into individuals and measure cellular diversity. New genomic methods have opened a window onto that diversity. We are now poised to understand how somatic diversity affects the risk of disease. The most important problem concerns how much the inevitable variability between individuals translates into differences in disease.

References

1. Hall, J.G. (1988). Somatic mosaicism: observations related to clinical genetics. Am. J. Hum. Genet. *43*, 355–363.
2. Frank, S.A., and Nowak, M.A. (2003). Developmental predisposition to cancer. Nature *422*, 494.
3. Frank, S.A. (2007). Dynamics of Cancer: Incidence, Inheritance, and Evolution (Princeton, NJ: Princeton University Press).
4. Frank, S.A. (2010). Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. Proc. Natl. Acad. Sci. USA *1* (107 Suppl), 1725–1730.
5. Meza, R., Luebeck, E.G., and Moolgavkar, S.H. (2005). Gestational mutations and carcinogenesis. Math. Biosci. *197*, 188–210.
6. Alzualde, A., Moreno, F., Martinez-Lage, P., Ferrer, I., Gorostidi, A., *et al*. (2010). Somatic mosaicism in a case of apparently sporadic Creutzfeldt-Jakob disease carrying a de novo D178N mutation in the PRNP gene. Am. J Med. Genet. B. Neuropsychiatr. Genet. *153B*, 1283–1291.
7. Poduri, A., Evrony, G.D., Cai, X., and Walsh, C.A. (2013). Somatic mutation, genomic variation, and neurological disease. Science *341*, 1237758.
8. Proukakis, C., Houlden, H., and Schapira, A.H. (2013). Somatic alpha-synuclein mutations in Parkinson's disease: hypothesis and preliminary data. Mov. Disord. *28*, 705–712.
9. Engeholm, M., and Gasser, T. (2013). Parkinson's disease: is it all in the genes? Mov. Disord. *28*, 1027–1029.
10. Laurie, C.C., Laurie, C.A., Rice, K., Doheny, K.F., Zelnick, L.R., *et al*. (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat. Genet. *44*, 642–650.
11. Jacobs, K.B., Yeager, M., Zhou, W., Wacholder, S., Wang, Z., *et al*. (2012). Detectable clonal mosaicism and its relationship to aging and cancer. Nat. Genet. *44*, 651–658.
12. Macosko, E.Z., and McCarroll, S.A. (2012). Exploring the variation within. Nat. Genet. *44*, 614–616.
13. Lupski, J.R. (2013). Genetics. Genome mosaicism–one human, multiple genomes. Science *341*, 358–359.
14. Abyzov, A., Mariani, J., Palejev, D., Zhang, Y., Haney, M.S., *et al*. (2012). Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature *492*, 438–442.
15. McConnell, M.J., Lindberg, M.R., Brennand, K.J., Piper, J.C., Voet, T., *et al*. (2013). Mosaic copy number variation in human neurons. Science *342*, 632–637.
16. O'Huallachain, M., Karczewski, K.J., Weissman, S.M., Urban, A.E., and Snyder, M.P. (2012). Extensive genetic variation in somatic human tissues. Proc. Natl. Acad. Sci. USA *109*, 18018–18023.
17. Weinberg, R.A. (2007). The Biology of Cancer (New York: Garland Science).

18. Armon, C. (2005). Acquired nucleic acid changes may trigger sporadic amyotrophic lateral sclerosis. Muscle Nerve *32*, 373–377.
19. Jucker, M., and Walker, L.C. (2013). Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. Nature *501*, 45–51.
20. Prusiner, S.B. (2012). Cell biology. A unifying role for prions in neurodegenerative diseases. Science *336*, 1511–1513.
21. Watts, J.C., Giles, K., Oehler, A., Middleton, L., Dexter, D.T., *et al*. (2013). Transmission of multiple system atrophy prions to transgenic mice. Proc. Natl. Acad. Sci. USA *110*, 19555–19560.
22. Frank, S.A. (2003). Somatic mosaicism and cancer: inference based on a conditional Luria-Delbruck distribution. J. Theor. Biol. *223*, 405–412.
23. Campbell, C.D., and Eichler, E.E. (2013). Properties and rates of germline mutations in humans. Trends. Genet. *29*, 575–584.
24. Araten, D.J., Golde, D.W., Zhang, R.H., Thaler, H.T., Gargiulo, L., *et al*. (2005). A quantitative measurement of the human somatic mutation rate. Cancer Res. *65*, 8111–8117.
25. Youssoufian, H., and Pyeritz, R.E. (2002). Mechanisms and consequences of somatic mosaicism in humans. Nat. Rev. Genet. *3*, 748–758.
26. Greaves, M. (2005). In utero origins of childhood leukaemia. Early Hum. Dev. *81*, 123–129.
27. Gottlieb, B., Beitel, L.K., Alvarado, C., and Trifiro, M.A. (2010). Selection and mutation in the "new" genetics: an emerging hypothesis. Hum. Genet. *127*, 491–501.
28. Ashley, D.J. (1969). Colonic cancer arising in polyposis coli. J. Med. Genet. *6*, 376–378.
29. Knudson, A.G. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA *68*, 820–823.
30. Frank, S.A. (2005). Age-specific incidence of inherited versus sporadic cancers: a test of the multistage theory of carcinogenesis. Proc. Natl. Acad. Sci. USA *102*, 1071–1075.
31. Dinger, M.E., Mercer, T.R., and Mattick, J.S. (2008). RNAs as extracellular signaling molecules. J. Mol. Endocrinol. *40*, 151–159.
32. Happle, R. (2014). Mosaicism in Human Skin (New York: Springer).
33. Groesser, L., Herschberger, E., Ruetten, A., Ruivenkamp, C., Lopriore, E., *et al*. (2012). Postzygotic HRAS and KRAS mutations cause nevus sebaceous and Schimmelpenning syndrome. Nat. Genet. *44*, 783–787.
34. Esmarch, F., and Kulenkampff, D. (1885). Die elephantiastischen Formen: eine umfassende Darstellung der angeborenen und erworbenen Elephantiasis sowie aller verwandten Leiden (Hamburg: J. F. Richter).
35. Happle, R. (2007). Type 2 segmental Cowden disease vs. Proteus syndrome. Br. J. Dermatol. *156*, 1089–1090.
36. Loffeld, A., McLellan, N.J., Cole, T., Payne, S.J., Fricker, D., *et al*. (2006). Epidermal naevus in Proteus syndrome showing loss of heterozygosity for an inherited PTEN mutation. Br. J. Dermatol. *154*, 1194–1198.
37. Sakuntabhai, A., Dhitavat, J., Burge, S., and Hovnanian, A. (2000). Mosaicism for ATP2A2 mutations causes segmental Darier's disease. J. Invest. Dermatol. *115*, 1144–1147.
38. Wada, T., Shirakata, Y., Takahashi, H., Murakami, S., Iizuka, H., *et al*. (2003). A Japanese case of segmental Darier's disease caused by mosaicism for the ATP2A2 mutation. Br. J. Dermatol. *149*, 185–188.