Symbiosis

www.symbiosisonline.org
www.symbiosisonlinepublishing.com

Review Article | International Journal of Structural and Computational Biology | Open Access

Are Genes and Their Mutations Responsible for Disease?

Jung C Lee*

BioMolecular Engineering Program and Department of Physics and Chemistry, Milwaukee School of Engineering, Milwaukee, Wisconsin 53202

Received: July 25th, 2017; Accepted: August 14th, 2017; Published: August 22nd, 2017

***Corresponding author:** Jung C. Lee, Assistant Professor, BioMolecular Engineering Program and Department of Physics and Chemistry, Milwaukee School of Engineering, 1025 N Broadway, Milwaukee, Wisconsin 53202, USA . Tel: +414-277-7316, Fax: +414-277-2878;E-mail: lee@msoe.edu*

Word count

main text (2804 words) and abstract (127 words)

Keywords: Bioinformatics; Human reference genome; Comparative biosequence analysis; Biological robustness, Protein-coding gene; Genetic variations/mutations, Whole-exome sequencing; Repetitive elements

Abstract

For more than 45 years since the War on Cancer in 1971, the ruling conceptual framework on fighting disease has been, and still is, on hunting for genes and their mutations, churning out countless "promising" breakthroughs, none of which really panned out into treatment. As the human genome was declared "completed" in 2003, gene-obsessed scientists turned to whole-exome sequencing, in the guise of Personalized Medicine, striving to comparatively identify all the deleterious mutations likely to cause "personalized" disease. Nonetheless, the genome-guided Personalized Medicine is not that guaranteed either. This review is not merely to discuss the inherent problems around the common wisdom on disease, but to document a growing number of multidimensional evidences, all coming together to point to one thing: genetic mutation is not behind disease.

Introduction

At the June 26, 2000 Human Genome Announcement Speech at the White House, President Clinton praised the completion of the initial working draft sequence of the entire 3.2-Gb haploid human genome, stating: "It will revolutionize the diagnosis, prevention and treatment of most, if not all, human diseases." The announcement was universally hailed as a historic achievement for some hopes for galloping advances in the diagnosis and treatment of genetic disease. Two years later in April 2003, the Human Genome Project (HGP) was officially declared finished. The finished human genome map was published in October 2004, with a substantial improvement over the initial draft sequence [1], seemingly heralding the soon-to-be advent of Personalized Medicine. Bioinformatic analysis predicted that the human genome contains only ~20,000 protein-coding genes, a shock compared to ~100,000 as estimated at the turn of the new millennium. It turned out that not only the human genome map marked a beginning of new research to crack the code of disease, but it also was a launchpad to validate the ruling conceptual framework on fighting disease. Armed with the finished human genome, bioinformaticians have teamed up with geneticists and clinicians to address many confronting challenges, including: how to determine the exact total number of protein-coding genes embedded in the human genome; how to search for all the disease-causing deleterious mutations; how to accurately predict alternative splicing patterns of protein-coding genes; how to identify the functional and/or regulatory roles of the introns and the ubiquitous repetitive elements (REs) in gene expression and regulation; how to create a complete human genome reference sequence assembly which could be comparatively analyzed against anyone's genome. Nonetheless, three big problems around the human genome persist, making it hard to address these challenges.

Is the finished human genome finished?

The "finished" human genome has three inherent problems. First, it was haploid (3.2 Gb), not diploid (6.4 Gb), not able to make it possible not only to identify the accurate number of protein-coding genes embedded in the human genome, but also to provide the full spectrum of human diseases and disorders.

Second, it has never been completely sequenced, even with many substantial improvements over the initial draft, despite continued advances in genome sequencing and bioinformatics algorithms. Still, ~7% of the entire 3.2-Gb haploid human genome are missing from GRCh38, the human genome reference assembly build 38 (Figure: 1). The missing (or unsequenced) gaps mostly fall on

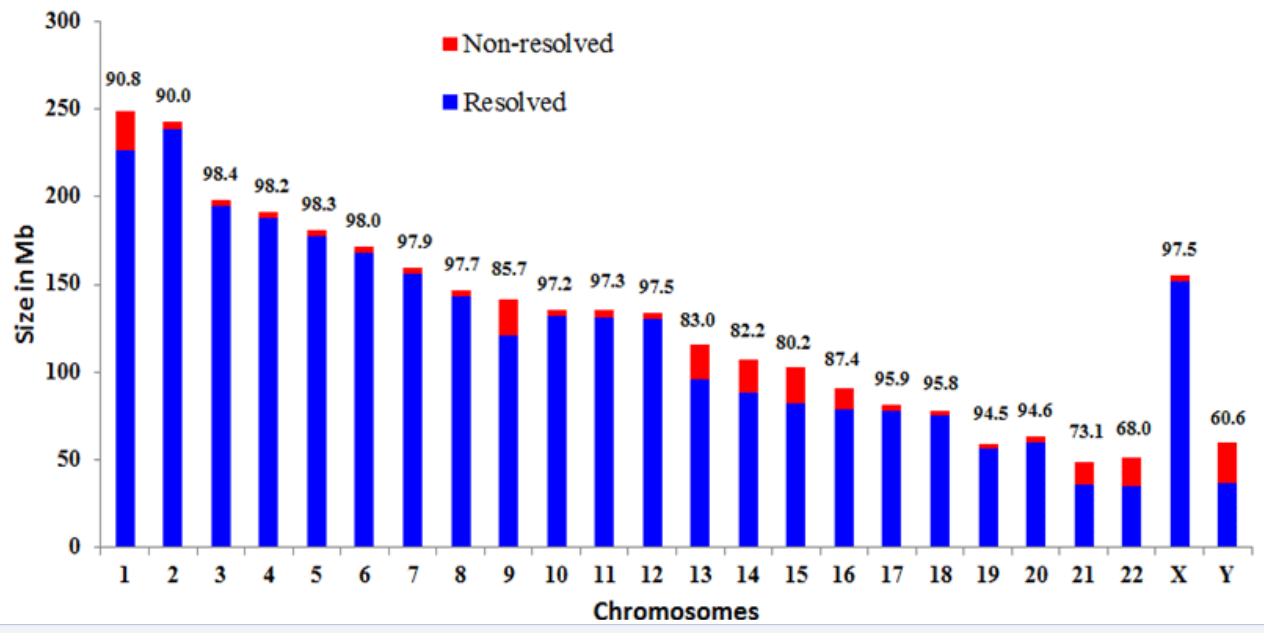

Figure 1: %Completion of the individual human chromosomes. Each of the 23 individual human chromosomes is not completely resolved yet, with overall %completion of 93%.

Corresponding author:lee@msoe.edu

"non-genic" REs, found clustered in the centromeres and telomeres of the human genome. Notwithstanding the gaps, leading "gene-obsessed" scientists never bothered with saying that the unsequenced should not play any role in disease. Subsequently, they have turned to whole-exome sequencing, aspiring to target only the protein-coding regions in the genome. GRCh38 is, in fact, a composite genome of multiple individuals; it is vastly updated compared to its poorly assembled predecessor GRCh37 [2], with new annotations for the centromeres and a highly expanded collection of alternate loci to better represent diverse human populations. It was recently suggested that a de novo genome assembly procedure, called 3D genome assembly originally developed to sequence the 1.2 Gb genome of the Zika virus-carrying *Aedes aegypti* mosquito, may help build a high-quality human genome reference sequence assembly with chromosome-length scaffolds, entirely from scratch [3].

Third, the compositional and structural organization of the human genome is extremely complex [4], containing many functional and regulatory elements (Figure: 2). Shockingly, however, only ~1% of the human genome codes for proteins, while up to 78% constitute REs including transposable elements (TEs) and short tandem repeats (STRs). In particular, the structural complexity of human protein-coding genes, composed of exons, introns, and untranslated regions (UTRs), requires alternative splicing for proper gene expression. Alternative splicing transcribes a single protein-coding gene into its several-to-many mRNA transcripts, each of which contains a different set of exons and is subsequently translated into its respective protein isoform. To make it worse, the patterns of alternative splicing are supremely variable and sporadically tissue-dependent, not predictable yet. Moreover, protein-coding genes, forward (+) and reverse (-), occur on both strands of the diploid human genome. Furthermore, supposedly noncoding introns can include additional protein-coding genes within themselves. Together, it is unlikely to develop a novel algorithm to unambiguously and accurately identify all protein-coding genes present in the human genome, raising questions about the rationale and validity of exome sequencing.

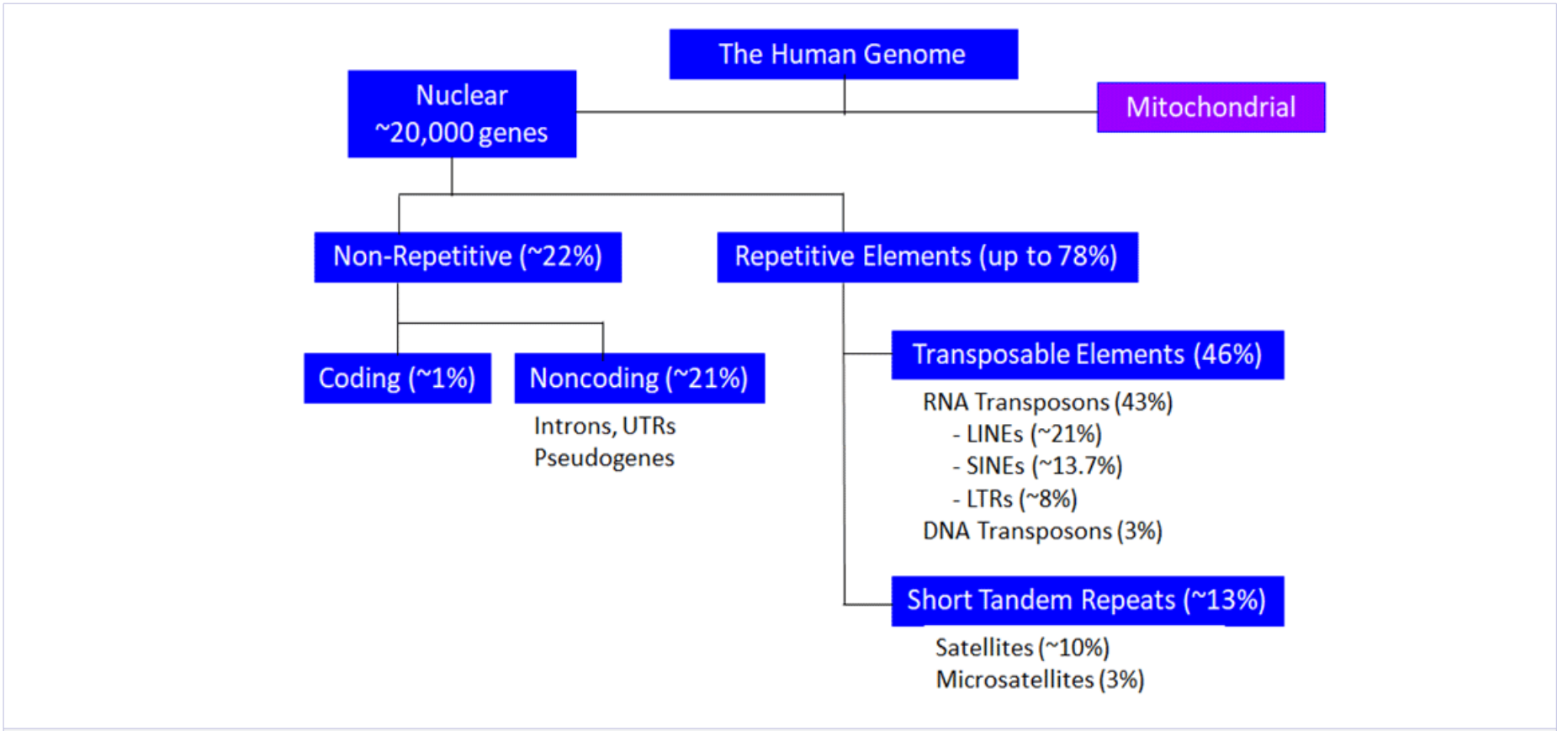

Figure 2: Compositional complexity of the human genome. The human genome contains many sequence components, including the ubiquitous non-genic repetitive elements (REs). Transposable elements (TEs) are comprised of long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and long terminal repeats (LTRs). ~19% of REs are still not annotated yet.

Are we on the right track to fight disease?

Since President Nixon's National Cancer Act of 1971, or the War on Cancer, there have been massive efforts to fight against cancer. Inspired largely by the 1976 discovery of retroviral oncogenes by the two 1989 Nobel laureates Michael Bishop and Harold Varmus, viruses had long been suspected as the cause of cancer, and billions of dollars had been poured into cancer virus research until the suspicion was removed in 2013. In parallel, "gene-obsessed" geneticists and oncologists have blamed genetic mutations as the bona-fide cause of cancer, having alleged that medical conditions and personal traits are all attributed to genetic mutations, inherited or caused by environment. Writing in The New York Times in 2009, the 1962 Nobel laureate James Watson declared: "We shall soon know all the genetic changes that underlie the major cancers that plague us." Gene-obsessed scientists keep focused on chasing after cancer-causing genes and their variations, announcing a myriad of cancer "breakthroughs." Nonetheless, cancer defied every one of them, remaining still elusive. What an abysmal failure! The time is now ripe to take a careful look at where we are and to redefine the conceptual framework for the War on Cancer.

Are genetic variations supposed to be mutations?

"Nothing in biology makes sense except in the light of evolution," writes Theodosius Dobzhansky in 1973. Evolution is all about change (or variation), forward and positive!

The basic premise of comparative biosequence analysis is that biosequences with a homologous function adopt similar higher-order structures to maintain the homologous function throughout evolution, regardless of their sequence variations, consistent with biological robustness [5]. In other words, biosequences respond with changes to perturbations, internal or external, while rendering their 3D structure and, in particular, their function unchanged. For example, both types of RNase P RNA, despite their substantial sequence differences, fold into overall similar 3D structures and perform the exact same function of tRNA maturation (Figure: 3). This implies that biosequences change in accordance with the Mother Nature, optimizing an organism's adaptation to their ever-changing environments and, in turn, evolving in the direction of increasing the entropy of the organism. In this regard, sequence variations in protein-coding genes, or genetic variations, are supposed to not harm an organism's function. Nonetheless, comparatively derived genetic variations are taken for granted as disease-causing mutations, not only violating the very premise of comparative biosequence analysis, but also ignoring biological robustness. In particular, gene-obsessed scientists keep alleging that a genetic mutation is a "permanent" alteration that triggers harmful changes in the structure of its coded protein; the resulting protein goes awry and initiates disease. In their perspective, genetic mutation is the only 'must-have' tool to fight disease.

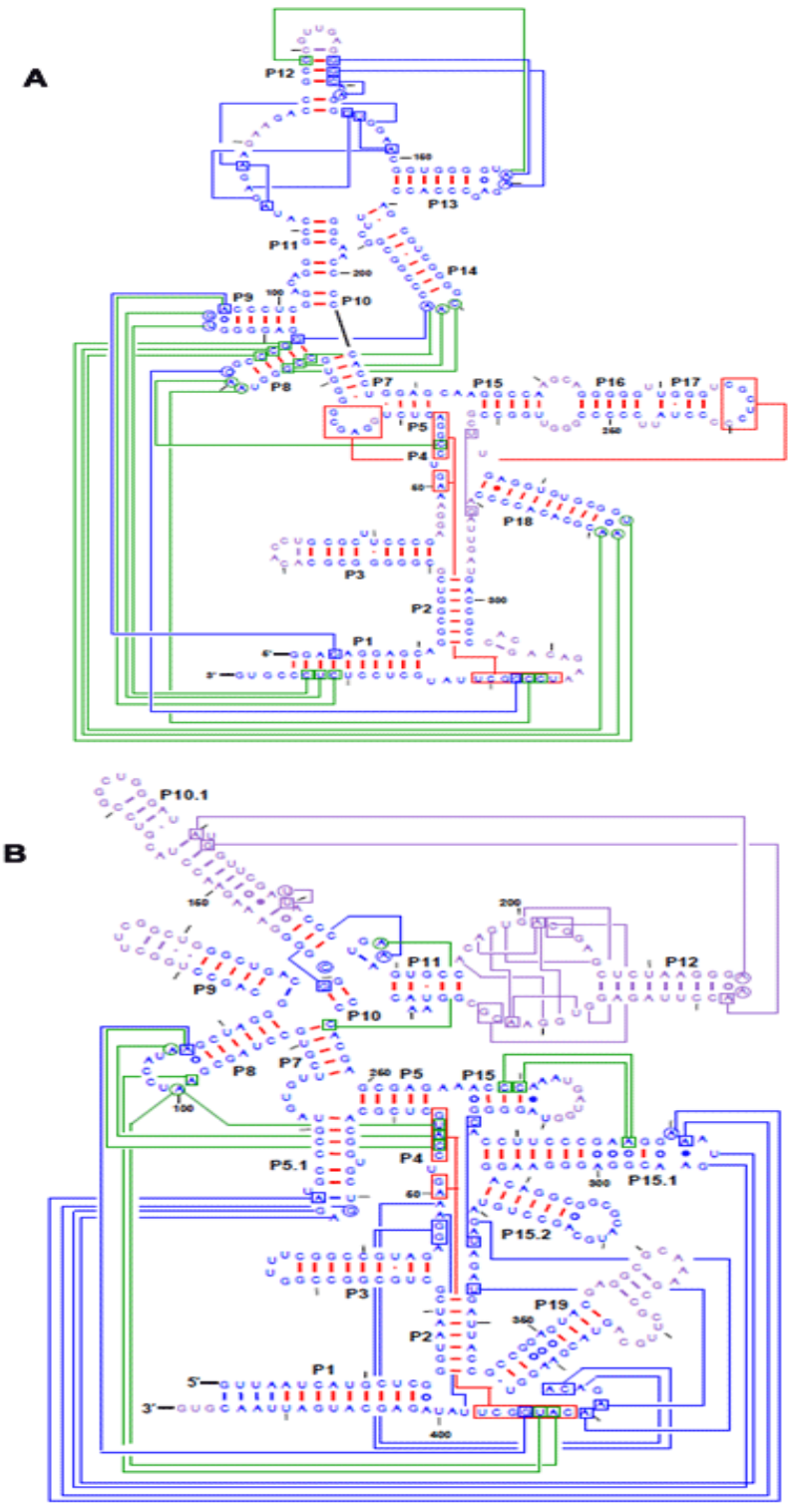

Figure 3: : RNase P RNA secondary structures with annotated tertiary interactions. (A) Type A from *Thermotoga maritima* (PDB ID 2A2E). (B) Type B from *Bacillus stearothermophilus* (PDB ID 2A64). Long-range tertiary interactions are colored differently; while red and blue lines are for base-base interactions, green lines are for base-backbone or backbone-backbone interactions.

Do genetic variations shed light on disease diagnosis and prevention?

Gene-obsessed scientists consider that every protein-coding gene we have exists for a reason. If something "mutates" the sequence of a gene, we would not survive or at least we would suffer from a disease. This suggests that genetic "mutations" be reflected in diseases or medical conditions. In this scenario, gene mutation databases, especially those based on whole-exome sequencing, should hold the key for disease diagnosis and treatment. However, they turned out not to be that reliable, either with too many false positives (up to 25%) or with too small number of data points for many rare diseases [6]. Most surprisingly, a recent bioinformatics analysis of >10 million genetic variations from 60,706 people's genomes revealed that only 15% (or 3,230) of the ~20,000 genes in the human genome are essential, with universally (or 100%) conserved among people; the remaining are highly variable from person to person and, in turn, can be removed with no impact on human health [7]. Similarly, only 18.7% (or 1,105) of the 5,916 genes in the *Saccharomyces cerevisiae* (yeast) genome are essential for growth [8]. According to the gene-centric view on disease, these apparently suggest that any change in any of the essential genes could lead to disease.

Interestingly, the blood disorder thalassemia is implicated by many simultaneous "mutations" occurring in both coding and noncoding regions of the β-hemoglobin gene [9]; it's so challenging to pinpoint which one, coding or noncoding, is responsible for the disease. More interestingly, primate-specific non-genic TEs called Alu elements, aka short interspersed elements (SINEs) accounting for ~13.7% of the human genome (Figure: 2), underlie numerous "genetic" diseases including breast cancer, colon cancer, Ewing's sarcoma, hemophilia, Leigh's syndrome, diabetes type II [10]. Besides, a recent genome sequencing study confirmed that a woman lacking a gene putatively "essential" for fertility had three children [11]. It was further confirmed that ependymomas, the third most leading childhood brain cancer, bears no genetic variations, but rather reveals a strong correlation with epigenetic changes [12]. Strikingly, a new cohort study of 1,007 Ashkenazi Jewish patients diagnosed with breast cancer showed that 903 (or 90%) carried none of the three conventional 'founder' mutations in the BRCA1 or BRCA2 genes [13]; 865 (or 96%) of 903 had no genetic changes, raising a big question about the practicality of gene-based breast cancer screening. Most shockingly, in January 2016, the Daily Mail Online reported a then 7-year-old "bionic" girl who doesn't eat, sleep, and feel pain, even with her entire chromosome 6 missing. All these studies clearly suggest not only that the "essential" genes be not that essential, but that genetic "mutations" be not listed as a dominant risk factor for disease.

A recent large-scale screening against 874 genes from 589,306 healthy individuals' genomes identified 13 adults harboring resilient mutations for 8 serious single-gene Mendelian rare diseases (aka orphan diseases), each with no positive childhood conditions for the orphan diseases [14]. Another recent study, the first randomized trial of whole-exome sequencing in primary care, revealed that 11 out of 50 volunteers, whose exomes were sequenced, had "mutations" in at least one position of the 4,631 genes thought to be linked to ~1,000 orphan diseases [15]. Surprisingly, however, only 2 of the 11 people with disease-causing mutations turned out to be positive, raising questions about the application of routine whole-exome sequencing for disease diagnosis in a primary care setting. Furthermore, most non-Mendelian diseases are simultaneously linked to many different genes. For example, more than 100 genes are implicated in autism spectrum disorder (ASD) [16], which makes it extremely difficult not merely to track down what gene causes ASD, but to hunt for its proper treatment.

Together, there is no compelling evidence to support that genetic "mutations" cause disease. It is thus crystal clear that genetic variations do not shed light on disease diagnosis and treatment. Consequently, neither whole-exome sequencing-based Personalized Medicine nor CRISPR/Cas9-based gene editing to treat genetic diseases will be that practical. Therefore, the misleading mutation-centric view on disease must be phased out.

New frontiers beyond genetic variation

The human genome sequences vary from person to person. On average, humans acquire ~100 de novo variations per genome per generation, each of which is currently suspected as a cause of early-onset genetic disorders [17]. Its variations including ubiquitous copy number variations (CNVs), however, should not be simply perceived as disease-causing mutations, but rather as evolutionary consequences of its responses to perturbations as a way of adaptation to its fluctuating environments. With such wide variations among individual genomes, small and large, it's not possible to build a human genome reference sequence assembly; the current human genome "reference" sequence assembly, GRCh38, is a hodgepodge of many people's genomes. Thus, it does not make any sense for the reference genome sequence to be comparatively aligned against an individual's genome to identify personalized disease-driving "mutations" and then diagnose disease in the individual, healthy or ill. This indicates that genome-guided Precision Medicine, launched by President Obama's 2015 Precision Medicine Initiative (PMI) with a bold ambition to precisely diagnose and treat disease in the ill, will go nowhere unless "mutations" are really behind disease. Consequently, the gene-centric view on disease will never work, prompting a next-generation of strategies to fight disease.

Studies have shown that cancer cells produce far more lactic acid than normal cells regardless of oxygen availability [18], implying a strong association between cancer and glucose metabolism. Besides, a plethora of cancer-causing bacteria have been identified for more than a century [19]; the first ones were discovered by William Russell in 1890. It is now accepted that long-term infection with the bacterium *Helicobacter pylori* is the very cause of gastric cancer [20]. It was also

shown that pancreatic cancer, one of the most deadly with only 3% survival rate, is directly tied to two periodontitis-causing bacteria, *Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans* [21]. Other than cancer, bacteria lead to other diseases as well. A most recent microbiome analysis uncovered a direct link between gut microbiome composition and Alzheimer's disease (AD), the most frequent form of senile dementia [22].

Inflammation, the immune response to infection, is another leading culprit in all cancers and other diseases; the connection between inflammation and cancer was first proposed by Rudolf Virchow in 1863. It is now obvious that inflammation, including inflammatory bowel disease, promotes colon cancer development [23]. It was demonstrated for the first time that cancer cells keep making prostaglandin E2 (PGE2), a pro-inflammatory mediator in the colon, which in turn triggers the production of cancer stem cells (CSCs) responsible for cancer progression, metastasis, and resistance to chemotherapy [24]. It was also shown that the endotoxin-releasing gut bacterium *Enterobacter cloaceae* B29 induced obesity and diabetes through endotoxin-mediated inflammation [25]. However, little is still understood about how exactly inflammation triggers disease, suggesting that inflammation be a promising target for disease prevention and treatment. Inspired by the fact that anaerobic bacteria thrive inside anaerobic cancer cells and die inside normal cells, a previous study showed that, if combined with chemotherapy, the engineered anaerobic soil bacterium Clostridium novyi killed living tumors in mice [26]. Another study demonstrated that the combination of an engineered self-destructible bacterium Escherichia coli and chemotherapy destroyed cancer cells in mice significantly [27]. The development of such programmable cancer-targeting bacterial strains would be a game changer in fighting cancer.

Furthermore, all cancer cells are characterized by aneuploidy, an abnormal variation in chromosome number, including monosomy (2n – 1) or trisomy (2n + 1), unlike normal cells with diploidy (2n) [28]. Aneuploidy causes many other disorders as well, including monosomic Turner syndrome and trisomic Down syndrome. Unbelievably, however, aneuploidy in cancer has been completely ignored by scientists and researchers for the past several decades. Consequently, little is still well understood about how aneuploidy induces cancer [29]. Thus, it's presently not clear whether aneuploidy could be targeted for cancer treatment, despite some treatment opportunities to target putatively aneuploidy-induced energy and proteotoxic stress [30]. Strikingly, it was very recently reported that aneuploidy is linked to the variations in the centromeric, functional satellite repetitive elements (or simply satellites), which constitute ~10% of the human genome but mostly still remain unsequenced even with the most advanced sequencing technology [31].

Conclusions and Discussion

Genetic variations have been taken for granted as disease-causing mutations. In particular, gene-obsessed scientists have been focused on the hunt for genetic mutations in the ill. As the haploid human genome "reference" sequence came out in 2003, they turned to whole-exome sequencing in the hopes of uncovering all the disease-leading mutations by comparatively analyzing the genomes of individuals, healthy and ill, seemingly opening the door to Personalized Medicine. The past decade, however, has witnessed multidimensional discoveries and findings, all coming together pointing to one thing: genetic mutation is not responsible for disease. This implies that the gene-focused conventional view on disease is misleading and deceiving; not only does it ignore the very premise of comparative analysis of biosequences, but also violates biological robustness. In addition, the current human genome "reference" assembly is still not complete, containing many unsequenced gaps. Despite continued advances in genome sequencing methods and bioinformatics algorithms, stitching together large numbers of short sequence reads from DNA sequencing machines to build a correct, complete human genome assembly is still very challenging.

Multidimensional evidences now clearly reveal that the mutation-focused view on disease has no ground to stand on any more; gene-obsessed scientists have been "barking up the wrong tree." This means that a next-generation of strategies beyond genetic mutation needs to be established to fight disease. More and more evidences indicate that human health is largely and directly influenced by the gut microbiome, the complex bacterial community in the gut, which is composed of >10 trillion bacteria. The gut bacteria help digestion, produce vitamins B and K, and drive cancer, obesity, and other metabolic diseases. The gut bacteria generate signals, through which they communicate regarding their needs and stress. For example, hunger ache is likely to be their "trick or treat" signal for nutrients. This means that we need to listen and respond to bacterial signals promptly and properly, without disrupting the gut microbiome itself, to stay healthy and live longer. Thus, the gut microbiome is not only a new, untapped territory to explore for fighting disease, but the "Blue Ocean" to establish for drug discovery and development.

Acknowledgements

The publication fee of this work was supported by Physics and Chemistry Department, Milwaukee School of Engineering.

References

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-945. doi: 10.1038/nature03001

2. Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, et al. Resolving the complexity of the human genome using single-molecule sequencing. Nature. 2015;517(7536):608-611. doi: 10.1038/nature13907

3. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC , et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356 (6333):92-95 doi: 10.1126/science.aal3327

4. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. 2011. 7(12):e1002384. doi:10.1371/journal.pgen.1002384

5. Lee JC . Does sequence dictate structure which dictates function? Int. *J. Struct. Comp.* Biol. 2016;1(1):1-3

6. Sankaran, VG and Gallagher PG. Applications of high-throughput DNA sequencing to benign hematology. Blood. 2013;122 (22): 3575-3582. doi:10.1182/blood-2013-07-460337

7. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T2, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature . 2016;536(7616):285-292. doi: 10.1038/nature19057

8. Giaever G1, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the *Saccharomyces cerevisiae genome. Nature.* 2002;418(6895):387-391. doi: 10.1038/nature00935

9. Baker M. The changes that count. Nature. 2012;482(7384):257-262. doi: 10.1038/482257a

10. Deininger P. Alu elements: know the SINEs. Genome Biol.2011;2(12):236. doi:10.1186/gb-2011-12-12-236

11. Narasimhan VM, Hunt KA, Mason D, Baker CL, Karczewski KJ, Barnes MR,et al. Health and population effects of rare gene knockouts in adult humans with related parents. Science. 2016;352(6284):474-477. doi: 10.1126/science.aac8624

12. Bayliss J, Mukherjee P, Lu C, Jain SU, Chung C, Martinez D. et al. Lowered H3K27me3 and DNA hypomethylation define poorly prognostic pediatric posterior fossa ependymomas. Sci. Transl. Med. 2016;8(366): 366ra161. doi: 10.1126/scitranslmed.aah6904

13. Walsh T, Mandell JB, Norquist BM, Casadei S, Gulsuner S, Lee MK, et al. Genetic predisposition to breast cancer due to mutations other than BRCA1 and BRCA2 founder alleles among ashkenazi Jewish women. *JAMA Oncol.* 2017. doi:10.1001/jamaoncol.2017.1996.

14. Chen R, Shi L, Hakenberg J, Naughton B, Sklar P, Zhang J, et al. Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. Nat. Biotechnol. 2016;34(5):531-538. doi: 10.1038/nbt.3514

15. Vassy JL, Christensen KD, Schonman EF, Blout CL, Robinson JO, Krier JB.The impact of whole-genome sequencing on the primary care and outcomes of healthy adult patients: a pilot randomized trial. *Ann. Intern. Med.* 2017. doi: 10.7326/M17-0188

16. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515(7526):216-221. doi: 10.1038/nature13908

17. Rocio Acuna-Hidalgo, Joris A Veltman, Alexander Hoischen. New insights into the generation and role of de novo mutations in health and disease. Genome Biol. 2016;17(1):241. doi.org/10.1186/s13059-016-1110-1

18. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science. 2009;324(5930):1029-1033. doi: 10.1126/science.1160809.

19. Cummins J, Tangney M. Bacteria and tumors: causative agents or opportunistic inhabitants? *Infect. Agent.* 2013. *Cancer* 8:11. doi.org/10.1186/1750-9378-8-11

20. Helicobacter and Cancer Collaborative Group. Gastric cancer and Helicobacter pylori: A combined analysis of 12 case control studies nested within prospective cohorts. *Gut.* 2001;49(3): 347–353.

21. Fan X, Alekseyenko AV, Wu J, Peters BA, Jacobs EJ, Gapstur SM. Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. Gut. 2016;doi:10.1136/gutjnl-2016-312580

22. Harach T, Marungruang N, Duthilleul N, Cheatham V, Mc Coy KD, Frisoni G, et al. Reduction of Aβ amyloid pathology in APPPS1 transgenic mice in the absence of gut microbiota. Sci. Rep. 2017;7: 41802. doi: 10.1038/srep46856

23. Terzić J, Grivennikov S, Karin E, Karin M. Inflammation and colon cancer. *Gastroenterol.* 2010.138(6):2101-2114. doi: 10.1053/j.gastro.2010.01.058

24. Wang D, Fu L, Sun H, Guo L, DuBois RN. Prostaglandin E2 promotes colorectal cancer stem cell expansion and metastasis in mice. *Gastroenterol.* 2015;149(7):1884-1895. doi: 10.1053/j.gastro.2015.07.064

25. Fei N, Zhao L . An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. *ISME J. 7*(4):880-884. doi: 10.1038/ismej.2012.153

26. Dang LH, Bettegowda C, Huso DL, Kinzler KW, Vogelstein B. Combination bacteriolytic therapy for the treatment of experimental tumors. *Proc. Natl. Acad. Sci.* 98(26):15155-15160. doi: 10.1073/pnas.251543698

27. Din MO, Danino T, Prindle A, Skalak M, Selimkhanov J, Allen K. Synchronized cycles of bacterial lysis for *in vivo* delivery. Nature. 2016;536(7614):81-85. doi: 10.1038/*nature*18930

28. Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, et al. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science.* 2011;333(6045):1039-1043. doi: 10.1126/science.1203619

29. Gordon DJ, Resio B, Pellman D. Causes and consequences of aneuploidy in cancer. *Nat. Rev. Genet.* 2012;13(3):189-203. doi: 10.1038/nrg3123

30. Manchado E and Malumbres M. Targeting aneuploidy for cancer therapy. *Cell.* 2011;144 (4): 499-512.doi:10.1016/j.cell.2011.01.037

31. Sullivan LL, Chew K, Sullivan BA.α Satellite DNA variation and function of the human centromere. *Nucleus*. 2017;1-9. doi:10.1080/19491034.2017.1308989