On Matters Of Nature and Science

L A Liggett

2019-04-11

Contents

Chapter 1

Introduction

Here is an unexceptional glimpse of the universe as it was during my moment on this ball.

Chapter 2

Genetics and Genomics

2.1 Introduction

A haplotype block is a set of closely linked alleles or markers on a chromosome that tend to be inherited together over evolutionary time.

Across Eukaryotes, the frequency of recombination is inversely proportional to overall genome size. The result is that yeast have a recombination rate a few orders of magnitude higher than that of humans (She and Jarosz, 2018).

2.1.1 Subpoint

This is some sub info

2.1.2 Second subpoint

A haplotype block is a set of closely linked alleles or markers on a chromosome that tend to be inherited together over evolutionary time.

Across Eukaryotes, the frequency of recombination is inversely proportional to overall genome size. The result is that yeast have a recombination rate a few orders of magnitude higher than that of humans (She and Jarosz, 2018).

2.2 DNA Replication

When the origin of replication(s) is removed from bacteria or eukaryotes, growth and division is restricted or entirely eliminated, but in some strains of archaea like H volcanii, deletion of the origin of replication accelerates cell growth rates. It turns out that this archaea can use a process that is similar to homologous recombination to create a replication fork and replicate its chromosome (Hawkins et al., 2013).

2.3 Cloning

A macaque was the first primate to be cloned by SCNT (Liu et al., 2018).

2.4 Mutation Rate

Using whole-genome sequencing or next-gen sequencing to determine mutation rates by base, it appears that C>T mutations at CpG sites mutate at a frequency of 10-7 changes/cell division, and all other sites are within the range of 10-8-10-9 base changes per division (Arnheim and Calabrese, 2009, 2016; Campbell and Eichler, 2013; Ségurel et al., 2014).

Providing an example of how human mutation rates can differ by geographic origination, europeans compared to african/asian populations have a 1.6 increased mutation rate of a particular mutation (Harris, 2015).

It appears that humans have the highest germline mutation rate of all analyzed species (Lynch, 2016)

2.5 Mutation Hotspots

There are a number of sporadic mutation hotspots associated with disease incidence, like achondroplasia which has a sporadic incidence rate of 4.5 x 10-5 per generation (Arnheim and Calabrese, 2016; Waller et al., 2008). This disease originates from a single mutation in the FGFR3 gene at a mutation rate ~450 times higher than what would ordinarily be expected at a CpG site (Bellus et al., 1995; Rousseau et al., 1994; Shiang et al., 1994).

2.6 Mutation Detection

Pyrophosphorolysis-activated polymerization is a mutation detection method that can detect a single mutant molecule of DNA within 25,000 genomes (Liu and Sommer, 2004; Qin et al., 2007).

2.7 Genetic Modifications

Caffeine was cloned to allow for caffeine-deficient coffee and teas without the decaffeination process (Kato et al., 2000).

When a DNA-associating protein from tardigrades was cloned into mammalian cells, they became about 40% more tolerant to radiation (Hashimoto et al., 2016).

2.8 Sequencing Methods

Using a new sequencing method called sci-RNA-seq, the transcriptome of every cell of 762 cells in C Elegans was sequenced to yield single-cell sequencing results and transcriptome profiling of every cell in the body. The way this is done is by methanol fixing nuclei and then incorporating a UMI when converting to cDNA, then mixing cells again and incorporating another UMI when synthesizing the other strand (Cao et al., 2017).

It appears that DAPI does not increase sequencing error rates by Illumina sequencing (Leung et al., 2016).

One group came up with a method that is essentially identical to mine in which they use barcoded probes to detect leukemia but they tracked the mutation manually and ignored background (Wong et al. 2015)

2.9 Diagnostics

In Li and Snyder Cell 2018, the EHR from hospitals is used to integrate with a machine learning algorithm trained on aneurysm detection. Patients are then whole genome sequenced, and the genome sequencing plus the lifestyle of the individual on EHR is then used to predict if the person has an aneurysm. They were able to achieve pretty robust detection results that could then be used in a prediction setting in the clinic.

2.10 Detecting Common Diseases

Much of the following information comes from this review: (Shendure et al., 2019).

Linkage disequilibrium studies were designed to detect Mendelian diseases GWAS designed back in 1996 to detect non-mendelian multigenic traits that have much less penetrant effects The promise that GWAS could risk stratify people for diseases has been challenging because most diseases seem to be driven by an extremely large number of variants with small effects that will likely require extremely large sample sizes There exists a problem of missing heritability, and it was often believed that common SNPs only held part of the puzzle, and more rare variants accounted for a great deal of heritability, but this does not yet seem to be the case, and SNPs seem to have a much greater effect size Another problem with GWAS is it is haplotype specific in that it can implicate a stretch of DNA inherited from one parent, but is blind to the individual effect sizes of each of the individual variants A challenge raised by Jonathan Pritchard is that gene regulatory networks are so interconnected that variants in one gene may actually cause changes in other genes and are therefore only peripherally relevant to a phenotype One continued promise of the utility of GWAS to identify the causal genetics behind diseases is that most of the strongest GWAS associations came from small studies of european populations that identified mutations of large effect sizes. By expanding studies to populations, especially those like african populations that have less linkage disequilibrium many more variants of large effect sizes could be identified and used to tease out relationships of smaller effect sizes in other populations. Methods are also improving fo linking regulatory elements to the genes they regulate like (Gasperini et al. 2018; Gasperini et al. 2019). Linking regulatory elements to their corresponding genes can be quite helpful, because this information can be incorporated into GWAS calculations to refine causal linkage probabilities. Polygenic risk scores have often been used to predict phenotypic variance in plants and animals, and have yet to really be applied to human genomics (Khera et al. 2018). Training of PRSs seem to not require fine-mapping, and their use has been aided by the UK Biobank (Bycroft et al. 2018).

2.11 Detecting Rare Diseases

There are some 7k mendelian monogenic disorders that impact about 0.5% of live births, but contribute to about 70% of pediatric hospital admissions An important surprise has been that de-novo mutations account for a substantial amount of intellectual disabilities and autism, where as many as 30-60% of ASD is caused by de-novo mutations Currently as many as half of acutely ill inpatient infants can be diagnosed from WGS. There are currently 59 genes designated by the American College of Medical Genetics as being sufficiently clinically actionable as to warrant sequencing and reporting in patients (Kalia et al. 2017).

2.12 Tissue Evolution

By sequencing 7,664 tumors spanning 29 different cancer types, it appears that unlike species evolution, the force of positive selection in developing tumors outweighs that of negative selection as evidenced by the loss of less than 1 coding nucleotide substitution per tumor (Martincorena et al., 2017). The number of mutations per cancer varied from 1 per thyroid and testical cancer to over 10 per endometrial and colorectal cancers. This information helps to answer how many mutations are needed to effectively create cancers and how this can vary with across tissue types. A number of groups have tried to answer these questions in the

past by mathematically estimating the number of rate limiting steps required in the process of oncogenesis (Armitage and Doll, 1954; Tomasetti and Vogelstein, 2015). There are two important problems with this approach, first that not all driver mutations need to be rate-limiting (Yates et al., 2015), and not all rate limiting steps in oncogensis need to be driver mutations (Martincorena et al., 2015). It has also been problematic to sequence tumors and count the number of high frequency mutations in oncogenic genes, but this has the added challenges of distinguishing passenger from driver mutations and is limited to current lists of oncogenes. Lists of genes involved in cancer have become increasingly detailed, but are still limited (Lawrence et al., 2014; Kandoth et al., 2013). The absence of negative selection in cancer may well indicate how dispensable the majority of genes are for somatic cells.

In their paper, Martincorena and Campbell show that the dN/dS ratio for somatic tissues and cancer tissue is 1 or greater showing that the effect of negative selection is minimal. In contrast, the dN/dS for germline species evolution is less than 0.5 showing a much greater effect of negative selection. Surprisingly, the non-synonomous mutations showed a dN/dS ration of 1 whether they existed in haploid or diploid regions suggesting the cells were not simply tolerating the mutations by having two copies.

Using sequencing of blood to find spontaneous mutations and go back and use population biology to calculate the number of HSCs, it was found that there were between 50k-200k HSCs contributing to hematopoiesis at any given time (Lee-Six et al., 2018).

Chapter 3

Neuroscience

3.1 Alzheimer's Disease

It seems that the bacteria Porphyromonas gingivalis may be partially responsible for exacerbating symptoms in people with AD. Observation of the bacteria in the brains of people with AD showed evidence that the bacteria secreted proteases called gingipains that increased A production and were also neurotoxic. Inhibition of the proteases reduced neuroinflammation and rescued neurons in the hippocampus (Dominy 2019).

3.2 Sleep

Sleep deprivation results in the global phosphorylation of the brain proteome, an effect which is reversed by sleep. A mouse mutant for SIK3 causes mice to sleep more, and may play an important role in the cause of sleep desire (Wang et al. 2018).

Chapter 4

Aging

4.1 Somatic Mutation Theory

In their study covering 7,664 tumors, Martincorena and Campbell illustrate an almost complete absence of negative selection against somatic mutations (Martincorena et al., 2017). This is lack of negative selection is quite different from the force at the population level where it acts quite strongly on the germline. This result may be an important finding for the somatic theory of aging (Morley, 1995), as it may be an indication that point mutations are largely of negligible effect size within somatic cells. As such, if point mutations do play a role in aging, it seems that this may be limited to those that are neutral or advantageous to a cell.

Bibliography

Armitage, P. and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *British journal of cancer*, 8(1):1.

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471):333.

Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S. B., Lander, E. S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, 505(7484):495.

Lee-Six, H., Øbro, N. F., Shepherd, M. S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R. J., Huntly, B. J., Martincorena, I., Anderson, E., et al. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature*, 561(7724):473.

Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., Davies, H., Stratton, M. R., and Campbell, P. J. (2017). Universal patterns of selection in cancer and somatic tissues. *Cell*, 171(5):1029–1041.

Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D. C., Fullam, A., Alexandrov, L. B., Tubio, J. M., et al. (2015). High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, 348(6237):880–886.

Morley, A. A. (1995). The somatic mutation theory of ageing. *Mutation Research/DNAging*, 338(1-6):19–23.

Shendure, J., Findlay, G. M., and Snyder, M. W. (2019). Genomic medicine–progress, pitfalls, and promise. *Cell*, 177(1):45–57.

Tomasetti, C. and Vogelstein, B. (2015). Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, 347(6217):78–81.

Yates, L. R., Gerstung, M., Knappskog, S., Desmedt, C., Gundem, G., Van Loo, P., Aas, T., Alexandrov, L. B., Larsimont, D., Davies, H., et al. (2015). Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nature medicine*, 21(7):751.