EEB 297: Population genomics of structural variants and transposable elements

Jesse Garcia

2019-01-14

Contents

Chapter 1

Description

This is my study guide for the class EEB 297 and its associated articles.

Chapter 2

Week 2 Biology of CNVs & CNVs in Drosophila

2.1 Hastings et al. Mechanisms of change in gene copy number (2009) Nat Rev Genet

2.1.1 Abstract

- Deletions and duplications underlie human phenotypes and form at rates much bigger than other kinds of mutations
- Repair of broken replication forks might promote CNV production

2.1.1.1 Introduction

- (Somatic/Meiotically Generated) Identical twins differ in CNV and different organs and tissues vary in copy number in the same individual (Woah)
- CNV is at LEAST as important in the differences between humans as SNPs
- Can change protein structure
- CNV variation is disadvantageous and incolved in cancer formation and progression. Contributes to cancer proneness
- *Question*: Wouldn't this be adaptive for the cancer cells and not necessarily deleterious?
- Although we are looking at other species, it is probably ok to extrapolate what we get from bacteria

2.1.1.2 Characteristics of copy number variants

- Change in copy number == change in chromosome structure
- Low Copy Repeats (LCRs): recurrent CNVs whose end-points are confined to few genomic positions
  - Probably come from homologous recombination between repeated sequences
- *Question* Paragraph 2: "Most non-recurrent CNVs occur at sites of very limited homology of 2 to 15 base pairs (bp)"
  - How do you measure homology within yourself?
- Recurrent CNVs
  - Short
  - Tend to be in regions where LCRs (recurrent CNVs) are

2.1.1.3 Mechanisms of structural change

- Mechansims of all structural changes are the same as those that cause CNV
- Homologous Recombination (HR) requires *extensive sequence identity* (to what? sister chromatid?)
  - Important in accurate DNA repair
  - HR causes CNVs not because the mechanism is inaccurate, but because genomes have tracts of low copy repeats or segmental dupplications
- Rad51: Strand exchange protein important in most homologous recombination
- Nonhomology: use only "microhomology" of a few bases or no homology
- NAHR: Non-allelic or ectopic homologous recombination
  - when a damaged sequence is repaired by a homologous sequence in different chromosomal positions

2.1.1.4 Homologous recombination mechanisms

- Homologous recombination underlies many DNA repair processes.
  - repair of dna breaks and gaps
  - DSB: double-strand break induced recombination -Spontaneous mitotic recombination is probably initionated by single strand DNA gaps

2.1.1.4.1 Double Holliday junction and synthesis-dependent strand annealing models of double-strand break repair

- Holliday junction Double strand break repair is a mechanism that can lead to gene converseion and crossing over,
- Synthesis-dependent strand-annealing (SDSA)
  - does not generate crossovers
  - mechanism for avoiding crossing-over, and loss of Heterozygosity (*Question*: Would LOH events contribute to ROH computations?)
  - can still produce CNVs
- If chromatids carrying the same allele segregate together at mitosis then you can get LOH
- Repeats can cause NAHR (non-allelic or ectopic homologous recombination)
- *Question*: What's a vegetative cell?
- Length of repeats might affect probability of Homologuous recombination which. Too short==Less HR because of physical constraints of loop
- Break Induced Replication (BIR) Homologous recombination is also used to repair collapsed/broken replication forks
  - Can induce LOH if uses homologu instead of sister chromatid
  - suggested major mechanism for change in copy number
  - Can cause small deletions

2.1.1.4.2 Correct choice of recombination partner prevents chromosomal structural change

- Just don't pick a nonallelic partner for repair and you might be ok
- MutS and MutL work together to undo base-paired DNA molecules that are imperfectly matched
- Homeologous Sequenced: Sequences that share less than about 95% identity.
- Cohesins are porteins that literaly bind two sister chromatids together
  - loss of cohesion may induce copy number change at other loci
- Proteins even hold ends of a single Double strand break together
- Homologous recombination is dope for repair but can cause CNVS.

2.1.1.5 Nonhomologous repair

- Mechanisms of DNA repair that use very limited of NO homology.
- Can cause CNVs
- Two type: replicative and non-replicative mechansims

2.1.1.6 Nonhomologous repair: non-replicative mechanisms

2.1.1.6.1 Nonhomologous end joining

- Two pathways of Double strand break repair that do not require homology/verrry short microhomologies for repair
- Nonhomologous end joining (NHEJ)
  - NHEJ rejoins DSB ends accurately or leads to small 1-4 bp deletions, and also in some cases to insertion of free DNA, often from mitochondria or retrotransposons
  - *Question*: What is free DNA?
- Microhomology-mediated end joining (MMEG)
  - uses 5 to 25 bp long homologies to anneal to ends of double strand breaks and, leads to deletions of sequences between annealed microhomologies.
- Likely to cause some chromosomal rearrangement by joining nonhomologus sequences

2.1.1.6.2 Breakage-fusion-bridge cycle

- If a chromosome loses its telomere due to a double strand break, there will be two sister chromatids that lack telomeres (during replication)
  - Sister chromatids that lack telomeres will fuse, creating a diecentric chromosome
  - The fused chromosomes will be ripped appart during anaphase
  - this will happen over and over again
  - will lead to large inverted duplications/repeats
  - seen a lot in cancer
  - Barbara McClintock proposed this!

2.1.1.7 Nonhomologous repair: replicative mechanisms

- If you see microhomology at a site of nonhomologous recombination its probably because of nonhomologous end joining this CNV was created
- However, this might be a consequence instead of DNA replication, and Break induced repair instead of NHEj
- Replicative stree might induce CNV
- Aphidicolin: inhibitor of replicative DNA polymerases induces CNVs
- This suggests that replication can cause CNVs
- You see little homolgoy at these endpoints, suggests NOT HR

2.1.1.7.1 Replication slippage or template switching

- Single-stranded sequences that appear during replication (think Okazaki fragments) are often deleted or duplicated

2.1.1.7.2 Fork stalling and template switching

- During replication, forks can be stalled and the 3' primer end uses a single-stranded DNA template of another replication fork
- Microhomology suggest that Homologous recombination is not involved

- They messed with the concentraion of certain exoncucleases and were able to find out which exonucleases involved

2.1.1.8 Microhomology-mediated break-induced replication

- Break induced replication can be mediated by microhomology
- Pol32 which is a non-essential DNA polymerase is needed for Break induced replication
- Some author suggest that this probably causes non-recurrent copy number changes in human
- This author disagrees

2.1.2 Effects of chromosome architecture on CNV

- CNVs are not randomly distributed in the human genome
  - Clustered in regions of complex genomic architecture
  - Complex patterns of direct/inverted Low copy repeats
    - Can cause for stalling in DNA replication
  - heterochromatin near telomeres/centromeres
  - replication origins and terminators
  - scaffold attachment sequences
  - occurence of nonrecurrent changes in regions carrying multiple LCRs
  - inverted repeats and palindromic sequences
  - highly repeated sequences
  - LINEs & SINE
    - Cause CNV by Non-allelic or ectopic homologous recombination
  - non-B conformation able DNA
  - specific consensus sequences associated with CNVs
- *THEME*: Multiple genomic features can affect the probability of their occurence -*Question*: What are the genomic features that can affect the probability of SNPs?

2.1.3 Conclusions and ramifications

- At least two mechanisms for change in copy number -Non-allelic homologous recombination -Formed by classical HR-mediated Double strand break repair via a double holliday junction -Restarts broken replication forks by Homologous recombination -Microhomology-mediated events
  - underlie most copy-number change -Breakage-fusion-bridge cycle operates and mayber importanin in amplification in some cancers
- Don't think that only one mechanism causes one event. There's mediation/interference/synergy in these methods
- CNV could stem from stress reponse.
  - "evolvability"
  - stressed cells can fuel CNV formation and therefore genetic diversity upon which natural selection acts
- Cancer cells loss of heterozygosity drives tumor progression and resistance to therapies
  - *Question*: Cancer cell with runs of homozygosity will be more fit than other cells?
- Probably variants associated with CNVs
  - *Question*: Can we do a GWAS on the phenotype: Total Length of Genome in Run of homoszygosity

2.2 Emerson et al. Natural Selection Shapes Genome-Wide Patterns of Copy-Number Polymorphism in Drosophila melanogaster. (2008) Science

2.2.1 Abstract

- We don't really know how selection affects the distribution/density of CNVs.
- This paper identifies CNP (copy-number polymorphisms) in Drosphila and concludes that the locations and frequencies of CNPs are shaped by purifying selection
- *Strength of Purifying Selection*: Deletions > Duplications. Exon and Intron overlapping duplications and X chromosome duplications > random duplication

2.2.1.1 Paragraph 1

- "CNPs can create new genes, change gene dosage, reshape gene structures, and/or modify the elements that regulate gene expression, understanding their evolution is at the very heart of understanding how such structural changes in the genome contribute to the phenotypic evolution of organisms"

2.2.1.2 Paragraph 2

- Identify CNPs with a custom tiling array and use a HMM trained on a data from a line known to contain specific CNPs.

2.2.1.3 Paragraph 3

- They validated their model with wet-lab procedures.
- Deletions have a relatively high false-ositive rate (47%) because deletions are often near SNPs. This leads to DNA not binding well to the arrays _ *Question*: Wonder what they'd estimate their positve-predictive value to be?

2.2.1.4 Paragraph 4

- They compare predicted and "true" boundaries of CNPs and claim their model can detect small CNPs and estaimate CNP boundaries with precision
- They detect a lot more CNPs than in human with a smaller genome/sample size.
- Human CNPs might include a class that are larger than *anything* found in drosophila. Current studies are missing small-scale variations.

2.2.1.5 Paragraph 5:

- Duplications outnumbered deletions 2.5:1 (Sign test P value <2.22 × 10–16; Fig. 1) and were significantly larger (Wilcoxon rank sum test, P value <2.22 × 10–16; Table 1).
- Nonallelic homologous recombination should either generate a 1:! ratio of Duplications:Deletions OR more deletions than duplications.
- There is deletion bias.
- Suggests that a large proportion of deletions are removed from the population by purifying selection. In this context, the dearth of deletions observed in our data, as well as the smaller size of the deleted variants, suggest that they are far more deleterious than duplications and that larger mutations are more deleterious than smaller ones.

- Deletions == More Deleterious. Larger Mutations More Deleterious than small ones.

2.2.1.6 Paragraph 6:

- Every region of the genome harbors at least low levels of CNPs. The median distance between two events was 12.6 kb (fig. S5).
- Pericentromeric regions were enriched in duplications, though not in deletions (fig. S5)
- Pericentromeric regions are also characterized by extremely low rates of crossing-over, leading to a lower effective population size as a result of linkage (14). Therefore, the higher density of CNPs observed in these regions may be a consequence of the reduced effectiveness of selection in purging deleterious mutations (14). Alternatively, the mutation rate may simply be higher in such regions (15).
- My favorite paragraph so far
- *Question*: Didn't they design the tiling array? So the median distance/density is biased by themselves?
- *Questions*: Positive selection/interference could also cause these high density of enrichment of duplications?

2.2.1.7 Paragraph 7:

- More duplications in general in all categories of the genome
- Deletions relatively deleted in coding regions

2.2.1.8 Paragraph 8:

- 8% of genes partially duplicated
- 2% of genes partially deleted
- Transposable elements and CNPs are arranged similarly with respect to the ends of genes

2.2.1.9 Paragraph 9:

- Estimated demographic parameters, then used the parameters to not reject the standard neutral model, then estimated selection coefficients

2.2.1.10 Paragraph 10:

- Notably, selection differentially influenced CNP evolution among different genomic features as well as among different chromosomes. We compared the patterns of variation between the different classes of variants: both correcting for bias and error and with no corrections.
- Intronic is the most deleterious (splicing?)

2.2.1.11 Paragraph 11:

- Fail to reject neutrality for complete gene duplications

2.2.1.12 Paragraph 12:

- We also found that the autosomes have higher selection coefficients (less deleterious) than the X chromosome (Fig. 2). This observation is compatible with the following models:

- 1. Duplicate mutations on the X chromosome are more deleterious than those on autosomes (X-linked genes may be more sensitive to changes in dosage)

- 2. Duplicate polymorphisms tend to be slightly deleterious and recessive

2.2.1.13 Paragraph 13:

- Genes overlapping toxin respones and known to be under positive selection because of increased rates of gene expression

2.2.1.14 Paragraph 14:

- CNPs are distributed by natural selection

Chapter 3

Week 3

Chapter 4

Week 4

Chapter 5

Week 5

Chapter 6

Week 6