



Sequence Variation & Polymorphism in DNA Sequence

B.Sc. Biotechnology -II year

Subject-Computational Biology & Bioinformatics

Mamta Sagar

Assistant Professor, Department of Bioinformatics

UIET & IBSBT, CSJMU University, Kanpur

- The success of the human genome-sequencing program has stimulated interest in applying genetic analysis on a large scale. The prime motive is to use genetic methods to characterize aspects of human biology, especially those that are relevant to diseases, their treatment, and their cure.
- This endeavor requires more genetic markers than are currently available, but efforts are under way to discover many more.

- For example, an international consortium of 10 pharmaceutical companies and the Wellcome Trust has taken up the task of finding 300,000 random single-nucleotide (nt) polymorphisms (SNPs) spread throughout the genome.
- The National Institutes of Health (NIH) and some commercial concerns have initiated their own efforts to collect large numbers of random SNPs. A second school of thought holds that SNPs should be mined in coding and regulatory regions, which would point more directly to disease-causing mutations (11). Studies could then focus directly on candidate genes.

- Whatever the approach, the task of finding candidate genes by genetic methods requires large-scale analysis, and the task of resequencing to find relevant mutations in candidate genes may also be a large one. Once a disease-associated gene has been discovered and the spectrum of mutations characterized, there may be a need to test large numbers of individuals who are at risk for the presence of these mutations.
- Thus, there is a need for large-scale analysis of sequence variation at four stages in the characterisation of a human trait: discovery of informative sequence variants, mapping by association or linkage, mutation analysis in candidate genes, and diagnosis in individuals at risk. Analysis of sequence variation has applications in fields other than human disease and some of these, such as the study of human origins, also need analysis on a large scale.

Definition of Terms

- Sequence Polymorphism A sequence polymorphism is a variation in a DNA sequence that occurs at a frequency of $>1\%$ in the population, with no implied association with phenotype. A “functional polymorphism” defines a variant that does have an associated phenotype.
- A single-nt polymorphism (SNP) is a polymorphism in which alleles are defined by single or few base changes; in the human genome, these comprise mainly substitutions, but the term also embraces deletions and insertions of one or a few bases.
- The terms aSNPs and cSNPs describe anonymous (i.e. dispersed throughout genome) and coding SNPs, respectively.

- Mutation By contrast with a polymorphism, a mutation is a change in sequence that is directly associated with a change in phenotype.
- Genotyping Genotyping is the analysis of a locus to look for the presence of a known allele.

Resequencing

- Resequencing is the process of determining the sequence of a gene or other sequence known to be related to a reference sequence;
- typically, this process is carried out to look for mutations that have not been previously characterized.

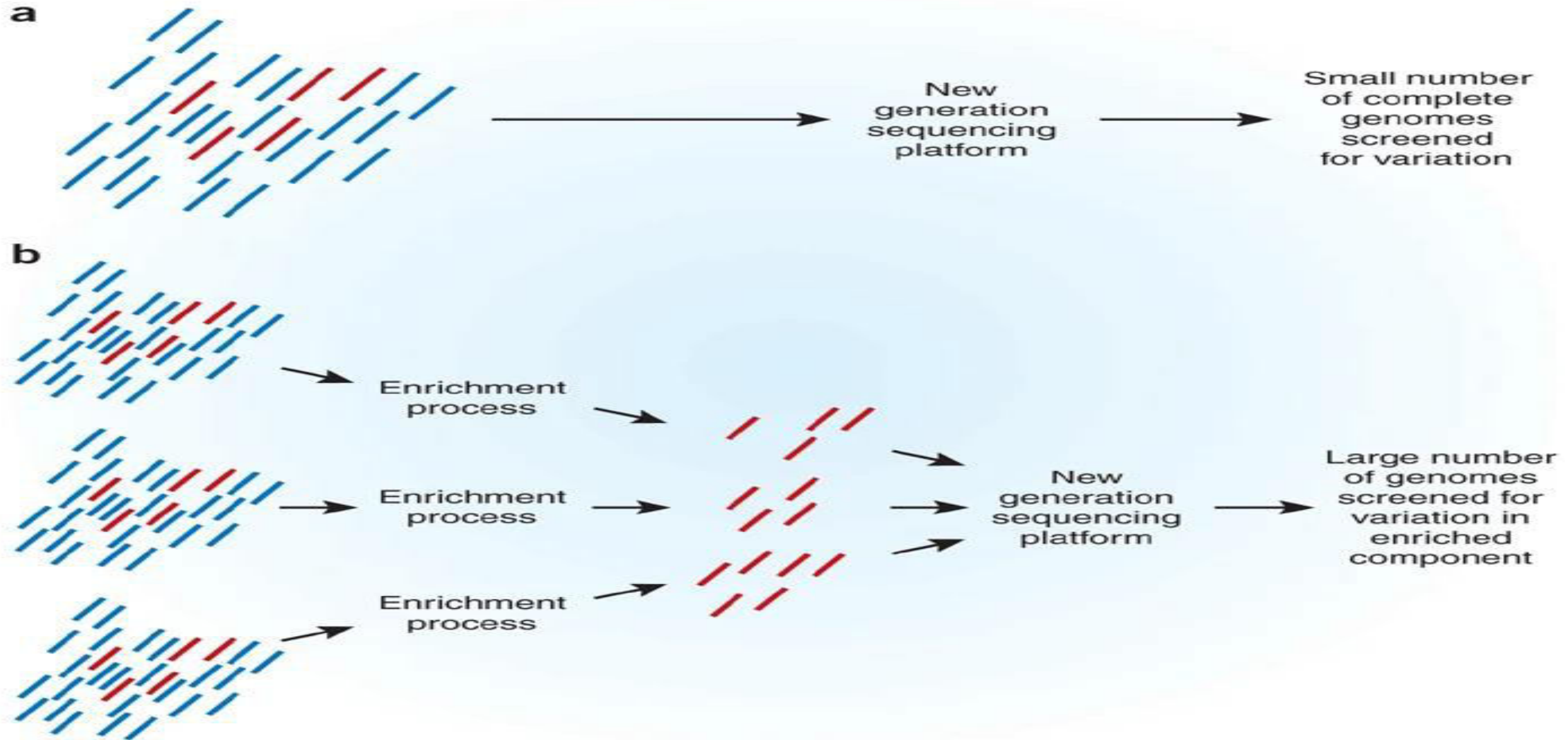
Genome resequencing and genetic variation

- Technologies for selecting segments of large genomes for resequencing will reveal biologically important sequence variation.**
- Now that reference genome sequences for many organisms are available, cataloguing sequence variation and understanding its biological consequences has become a major research aim.**
- However, for large eukaryotic genomes such as the human, even recently developed high-throughput sequencing technologies only allow deep genome-wide sequence coverage of a small number of individuals.**

- Resequencing the genome of many individuals for which there is a reference genome allows investigation of the relationship between sequence variation and normal or disease phenotypes.
- Some of these platforms permit ~100-fold greater rates of DNA sequencing at approximately 100-fold lower cost per base. (Stratton, M., 2008)

- If this new sequencing power could be targeted to limited areas of large genomes, it would become feasible to study variation in these regions in thousands of individuals.
- This configuration of the new sequencing technologies would allow direct and practical application to studies of human disease.
(Stratton, M., 2008)

Enrichment procedures allow redistribution of sequencing throughput from all of a small number of genomes (a) to a small component of a large number of genomes (b).



- The four reports in Nature Methods and Nature Genetics describe novel strategies for processing whole genomic DNA that result in substantial enrichment of a small part (up to 2%) of the genome.
- Coupling of these 'enrichment' approaches to high-throughput sequencing technologies promises to deliver a huge body of information about DNA sequence variation in populations and heralds a new phase in the genetics of humans and other organisms with large genomes.

Short Tandem Repeats Short tandem repeats (STRs)

Short Tandem Repeats Short tandem repeats (STRs) are sequences comprising a number of repeats of short, 2- to 5-nt sub-sequences. Typically, such repeats are scattered throughout the genome and are flanked by unique sequences that can be used to target a polymerase chain reaction (PCR) amplification of the locus. Most STRs are variable in copy number, making them valuable genetic markers. Synonyms for STRs include microsatellites and short sequence length polymorphisms. These are a subset of “variable number of tandem repeats,” which also includes minisatellites.

References

- Kalim U. Mir and Edwin M. Southern, 2000, *Annu. Rev. Genom. Hum. Genet.* 2000.1:329-360. 1527-8204/00/0728-0329\$14.00 **329**,
Downloaded from www.annualreviews.org
- Chakravarti A. 1998. It's raining SNPs, hallelujah? *Nat. Genet.* 19:216–17.
- Stratton, M. Genome resequencing and genetic variation. *Nat Biotechnol* **26**, 65–66 (2008). <https://doi.org/10.1038/nbt0108-65>