Applications of Genomics, Metagenomics and Proteomics in Cariology

Evangelia Morou-Bermudez, DDS, PhD
University of Puerto Rico School of Dental Medicine

Objectives

- To present current technologies available for the study of the human genome, microbiome, and the proteome.
- To discuss applications of genomics, metagenomics, and proteomics technologies in Cariology research.
- To discuss how the study of the human genome, microbiome and proteome can lead to the development of personalized approaches for the prevention and treatment of dental caries.

Definitions

- ***Genetics vs. Genomics***: genetics focus on the effects of single genes, while genomics focus on the interactions of all the genes in the genome and with environmental factors.
- ***Genome***: The entire collection of genetic material in the chromosomes of an organism.
- ***Transcriptome***: The entire collection of genes being expressed in an intermediate RNA template state within a specific cell, tissue or organ at a specific time
- ***Proteome***: The entire collection of proteins contained within a specific cell, tissue or organ in a specific time.
- ***Metabolome***: The collection of metabolites in a biological cell, tissue, organ or organism, which are the end products of cellular metabolic processes.
- ***Microbiome***: The totality of microbes living in a defined environment, their genetic elements (genomes) and environmental interactions. The human body contains 10 times more microbial cells than human cells.
- ***Metagenome***: Genetic material derived directly from an environmental/biological sample.

Most common technologies for generating high throughput data in genomic, metagenomic and proteomic research

- **Microarrays:**
- Use glass or nylon membrane platforms onto which a microscopic grid of molecular probes is arranged. This allows the simultaneous analysis of a hundreds of thousands of DNA, RNA, or protein targets from a **single** sample.
- Best application is to compare expression patterns among health and disease, e.g.: which genes are turned-on or off in a tumor vs. healthy tissue.
- Can also be used to look at an individual's entire genome for DNA variants that are linked to genetic disorders (e.g.. SNPs).
- Disadvantage: focuses on preselected target genes.

Most common technologies for generating high throughput data in genomic, metagenomic and proteomic research

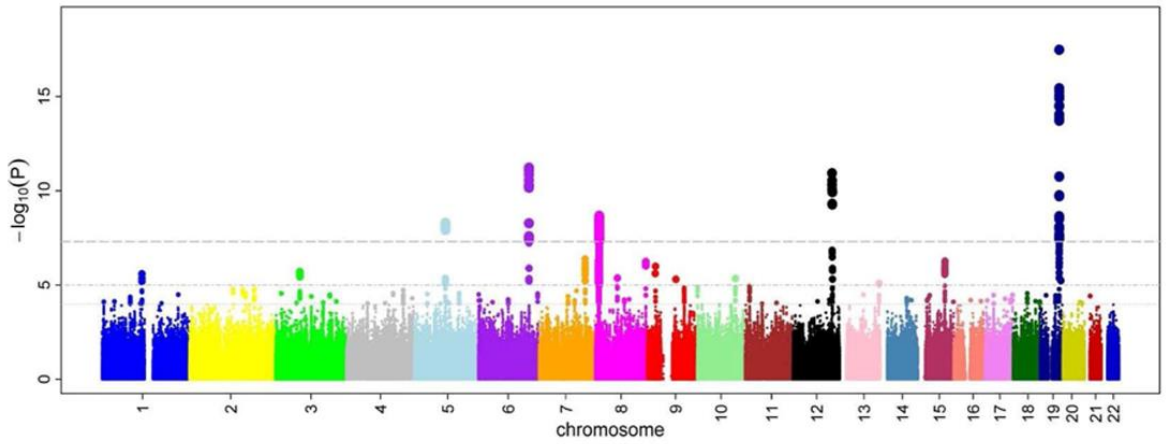
- **Next-generation sequencing:**
- Allows the sequencing of DNA fragments in a massive, parallel way, providing large number of sequence reads in a single run.
- Several methods available; most common:
  - ***454 pyrosequencing*** (Roche Diagnostics): 700 bp reads, 1 million reads per run, about 24 hours/run, most accurate (99.9%). Uses luciferase to detect light from ATP formation after the hydrolysis of pyrophosphate
  - ***Illumina (Solexa)***: 50 to 250 bp reads, 3 billion reads per run, takes 1 to 10 days, low cost per run but expensive equipment. Uses fluorescent nucleotides.
  - ***Ion semiconductor sequencing*** (Ion Torrent Systems Inc., now Life Technologies). 200 bp reads, 5 million reads per run, fast (2 hours), low cost per run and low cost equipment. 98% accuracy. Measures proton release during the DNA polymerization.

Most common technologies for generating high throughput data in genomic, metagenomic and proteomic research

- For metagenomics, there are two main approaches:
  - ***16S rDNA amplicon sequencing***: the hypervariable regions of the bacterial 16S rDNA are first amplified. The generated sequences are processed through a bioinformatics pipeline, which filters them based on quality and generates meaningful groups of sequences, or OTUs. These OTUs can then be compared to 16S rDNA databases, such as the Ribosomal Database Project (RDP), and be assigned a taxonomic classification
  - ***Direct shotgun sequence***: does not involve an intermediate amplification step. DNA gets fragmented to random fragments (about 1,000 bp) and sequenced in parallel. After sequencing, low quality and non-bacterial DNA gets filtered, and then fragments are assembled into contigs, which are then compared to full-genome reference databases to identify the taxa.

Most common technologies for generating high throughput data in genomic, metagenomic and proteomic research

- **High-performance liquid chromatography mass spectrometry**
- Proteins and other metabolites are separated based on their physical characteristics by liquid chromatography. This is followed by mass spectrometry (MALDI-TOF matrix-assisted laser depolarization/ionization), which determines the elemental composition of the sample and elucidates the chemical structure of molecules such as peptides and other chemical compounds.
- Can be used for proteomics and metabolomics.
- In case of complex mixtures where peptide masses may overlap with high resolution mass spectroscopy, the proteins may be separated on as SDS-PAGE gel and then subjected to peptide-mass fingerprinting (PMF) to identify the individual proteins. This is done using computer programs that translate the genome of the organism into proteins.
- Expensive and labor-intensive procedures, still not appropriate for clinical uses.



Genome Wide Association Studies

Genome-Wide Association Studies (GWAS): rapid genomic analysis of patient samples to find genetic variations associated with a particular phenotype

Human Genome Project competed in 2003. 50,000 genes identified.

Genome Wide Association Studies for Dental Caries

- *Vieira et al., 2008*: array-typed genotyping. Pilipino population. 392 genetic markers. Three loci showed suggestive only association with low caries susceptibility; none of them had obvious relationship to caries. They included genes related to stress, reward and feeding and olfactory receptors and caries. Also, a suggestive caries-protective locus was found on the X chromosome, which may provide a genetic explanation to the gender differences in caries susceptibility. The genes with suggestive associations with high caries susceptibility were involved immune responses and an estrogen receptor.

- *Wendell, et. Al, 2010*: About 2,000 participants, in primary, mixed and permanent dentition. Genotyping for three genes involved in taste pathway was done by real-time PCR. Two of these genes showed significant association with caries susceptibility/protection, suggesting that genetic variations may influence dietary preferences of those related to caries development

Genome Wide Association Studies for Dental Caries

- *Shaffer et al., 2011*: GWAS using three different cohorts of white children, 3-12 years. (1,302 from USA, and 1,695 from Denmark) for 1,5 million SNPs. No SNPs met genome-wide significant association (marginal p-value <10E-7). Multiple suggestive loci (p-values between 10E-5 and 10E-7) were observed. They included genes involved in tooth enamel formation, ectodermal dysplasia, response to oral epithelial colonization, salivary enzymes. No reproducibility between US and Danish samples. No genome-wide significance was observed for enamel genes, tuftelin, or taste genes, which had been suggested by earlier genetic studies.

- *Wang et al., 2012*: GWAS involving five independent adult cohorts, totaling more than 7,000 participants. Genotyping was performed with different methods in each cohort, such as SNP arrays and Illumina. Meta-analysis was then used to combine the results of the different cohorts. Again, only suggestive loci were identified, with roles similar to those in the previous study, such as genes involved in tooth development, and immune responses to oral bacteria.

Genome Wide Association Studies for Dental Caries

- *Shimizu et al., 2012*. Correlated genotypic variations in selected genes that are involved in enamel formation (*ameloblastin, amelogenin, enamelin, tuftelin, and tuftelin interacting protein*) with enamel micro-hardness before and after cariogenic challenge (on 48 extracted teeth). Enamel microhardness before and after a cariogenic challenge was significantly associated with TUFT1 and TUIP11. TUIP11 has no obvious relation to caries, and has not been found related to caries in clinical studies. It is involved in diverse cellular activities, such as pre-mRNA splicing and tumorigenesis, and it was speculated that it might influence the enamel's ability to uptake fluoride.

Genome Wide Association Studies for Dental Caries

- Limitations of GWAS:
  - Not well-defined case and control groups, differences in genotyping methods
  - Large potential for false-positive results due to the massive number of statistical tests available
  - Lack of reproducibility across studies and populations
  - Most of the suggestive loci are non-coding, and far from discovered genes; therefore they can not be assigned clear biological roles

Applications of Metagenomics in Cariology Research

- The Human Oral Microbiome Database (http://www.homd.org) includes 619 taxa in 13 phyla. The majority of the species are uncultivated and they are recognized by their 16S rRNA sequence.
- A specific microbiome that is associated with dental caries has not been yet identified.

Applications of Metagenomics in Cariology Research

- *Li et al., 2007*: 12 CF children and 8 with S-ECC. 16S rRNA amplification by PCR. Species richness and diversity is significantly reduced in caries vs. health.

- *Ling et al., 2010*: 60 children 3 to 6 years of age. Pyrosequencing, about 186,000 high quality sequences obtained. The microbial communities of saliva and supragingival plaque differ significantly and can be clustered apart. Caries experience did not correlate with specific pathogens, but rather with pathogenic populations.

Applications of Metagenomics in Cariology Research

- *Alcaraz et al., 2012*: pyrosequencing, 8 plaque samples from individuals with different caries activity. Shotgun pyrosequencing. Results are largely descriptive…no statistics.

- *Gross et al., 2012*: 12 to 36 month old children. 36 with caries and 36 caries-free. 16S sequencing. Total clones sequenced <10,000. Decreasing species diversity with increasing caries experience. *S. mutans*, was dominant in many children with caries but not all; alternative pathogens in the low-mutans children were *S. salivarius, S. sobrinus,* and *S. parasanguiniis*. Veillonela was significantly associated with acid-producers.

Applications of Proteomics in Cariology Research

- Human salivary proteome: essentially completed. About 2,000 different proteins and peptides.
- Salivary transcriptome, metabolome and microRNAs
- Saliva Proteome Knowledge Base: http://www.hspp.ucla.edu/spkbintro.html
- Salivary biomarkers for systemic diseases such as cancer, heart disease, infections diseases have been identified and tools for the rapid-chair side identification of those biomarkers have been developed.
- However, most of those are not actually salivary proteins, but inflammatory biomarkers from serum. Little progress has been made in identifying salivary protein biomarkers for oral diseases, such as caries.
- Salivary proteins, especially the glycosylated ones, play important roles in interacting with the oral bacteria. Salivary glycoproteome and other post-translational modifications need to be studied.
- Technical issues related to sample collection, processing and storage are critical in proteomic analysis of saliva

Applications of Proteomics in Cariology Research

- *Hart et al., 2011:* Developed univariate and multivariate models which combined plaque microbiome and salivary proteome data to discriminate between caries-free children and children with ECC. Sensitivity, specificity all >80%, ROCs with AUCs up to 96 for the combined microbiome and proteome data. Microbiome data gave better AUCs than the proteome. Microbiome was studied with reverse capture checkerboard hybridization with about 80 bacterial probes. Proteomic analysis was done using a protein chip array.

Applications of Proteomics in Cariology Research

- *Siqueira et al., 2012*: Used liquid chromatography electrospray ionization mass spectrometry to compare the composition of salivary pellicle formed on HA discs treated with various concentrations of F. Treatment of the HA disks with fluoride at high concentrations resulted in qualitative and quantitative changes in the pellicle which can potential impact the formation of the biofilm.

- Proteomic technologies can be used to compare the proteomes of oral bacteria or the proteome of a particular organism under different conditions, in particular under those conditions that are relevant to a cariogenic challenge

- (*Martinez et al., 2010*: Characterization of the *S. sobrinus* acid stress response).

Future Directions and Challenges

- The high-throughput technologies present amazing new research opportunities for the study of the genetic and biological factors involved in the development of dental caries.
- They also present unique new challenges, such as:
  - The development of bioinformatics approaches that can handle the massive amount of data that is generated, and that can combine the results generated by different technologies.
  - Ethical issues.
- Dental researchers need to understand the potential and the limitations of these technologies in order to apply them appropriately in their research.
- The knowledge derived from these studies can facilitate the development of novel approaches for the early identification of individuals at risk for developing dental caries and the planning of personalized approaches for caries management.