Imagining the Future of Anthropogeny

Glossary

Agricultural revolution: The transition from **hunting and gathering** to settled **agriculture** starting ~12,000 years ago in Mesopotamia. The development of and transition to agrarian life ways and technology was piecemeal in different places and at different times instead of a full blown "revolution."

Agriculture: A subsistence strategy that relies on **domesticated species** instead of **hunting and gathering** wild animals and plants.

Allele: Alternative **DNA sequence** at the same **locus** (location on the **chromosome**).

Alzheimer's disease: The most common type of dementia in humans. It is a progressive neurodegenerative disease marked by memory loss, language problems, disorientation, and behavioral changes.

Amino acids: Organic compounds that are the building blocks of **proteins** and participate in a number of processes such as neurotransmitter transport and biosynthesis. Amino acids are encoded by the **genome** as different three **nucleotide** codes.

Ancient DNA: **DNA** that is extracted from ancient specimens (skeletons, mummified tissues, frozen specimens, archeological material, archival collections, sediments, and dirt). The current upper age limit for ancient DNA extraction and sequencing is 0.4-1.5 mya.

Anthropocene: The proposed geologic epoch defined by human influence on the Earth. There is yet to be consensus for when the Anthropocene began with suggestions ranging from the start of the **agricultural revolution** to the first atomic explosion.

Anthropogeny: The study of human origins. A compound of Greek words for humans (anthropo) and origins (geny).

Archaic admixture: **DNA** from ancient, divergent, and now extinct populations found in current people.

Artificial intelligence (AI): A form of intelligence in which a machine system is able to make rational decisions based on perception of its environment.

Autism Spectrum Disorder (ASD): A range of conditions characterized by challenges with social skills, repetitive behaviors, speech, and nonverbal communication.

Bipedalism: A form a locomotion involving only an animal's rear, or hind, limbs for propulsion.

Brain organoid: An artificially grown in vitro brain model used for investigating brain development and neurological disease. Brain organoids are derived from **induced pluripotent stem cells** and **embryonic stem cells**.

Chromatin: A complex of **DNA** and **proteins** (histone and adaptor proteins) forming **chromosomes**.

Chromosome: Discrete strands of tightly packaged **chromatin**.

Copy number variation (CNV): A phenomenon in which sections of the **genome** are repeated on the same or different **chromosome** and the number of repeats in the genome varies between individuals in the human population. Such repeats can include functional **genes**.

Deleterious mutation: A genetic change that decreases an individual's fitness and increases susceptibility to or causes disease or disorder. Most deleterious mutations are recessive, i.e. selection can only act on them when an individual carries two copies of the same mutation.

Deoxyribonucleic acid (DNA): The molecule of inheritance, which consists of **sequences** of the four **nucleotide** bases: **A**denine, **T**hymine, **G**uanine, and **C**ytosine.

Diploid: Organisms with two sets of each **chromosome** except for XY sex chromosomes in male mammals.

DNA methylation: A process by which methyl groups are added to the **DNA molecule**. Methylation can change the activity of a DNA segment without changing the **sequence**. When located in a **gene** promoter, DNA methylation typically acts to repress gene **transcription**.

DNA sequence: The specific order of the **nucleotide** bases along a strand of **DNA**.

Domestication: The process of artificial selection by humans for desired traits of plants, animals, fungi, and microorganisms. This implies the complete control of the reproduction of those species.

Embryonic stem cell (ESC): The inner cell mass of the human blastocyst, the rapidly dividing fertilized egg at four to seven days post fertilization. ESCs are pluripotent, meaning they can differentiate into all cell types of the three embryonic tissue layers.

Foraging: Searching for wild food or provisions as opposed to cultivating food crops or breeding livestock.

Frontal lobe (brain): The largest of the four major lobes of the brain in mammals, and is located at the front of each hemisphere. It is devoted to action such as skeletal movement, ocular movement, speech control, the expression of emotions. In humans, the largest part of the frontal cortex is the **prefrontal cortex**.

Gametes: Mature **haploid** sex cells that can unite to form a **diploid** zygote.

GATA binding protein 3 (*GATA3*): A gene that encodes a protein in the GATA family of transcription factors. *GATA3* plays an important role in endothelial cell biology and in allergy and immunity against worm infections. In humans, defects in *GATA3* cause hypoparathyroidism with sensorineural deafness and renal dysplasia.

Gene: A **DNA sequence** which encodes a specific function.

Gene expression: The process by which the information contained within a gene (nucleotide sequence) is used to direct protein synthesis and dictate cell function. Nearly all of the cells in the body contain identical genes, but only a subset of this information is used or expressed at any time. The genes expressed in a cell determine what that cell can do.

Gene pool: The total of all **genes** and their variants (**alleles**) of a population of a **species**.

Genetic diversity: The total of heritable traits within a **species**.

Genetic load: The presence of deleterious **gene** variants (including recessive variants) in a population.

Genetics: The study of genes and their inheritance.

Genome: The totality of **DNA** in a cell. Also refers to the **DNA sequence** that typifies an individual or **species**.

Genomics: The study of **genome** structure/function.

Glia (aka neuroglia): Non-neuronal cells in the central nervous system and the peripheral nervous system that do not produce electrical impulses. Their function is to ensure homeostasis, form myelin, and provide support and protection for neurons. Glia make up ~50% of our brain cells.

Glycoprotein: A class of **proteins** with covalently attached glycans. Glycoproteins play a part in important cellular functions like embryonic development, cell-to-cell recognition, cell adhesion, and immune functions.

Haploid: One set of unpaired **chromosomes**.

Holocene: The current geological epoch, from about 11.7 kya (after the end of the last Ice Age cycle) to the present that is marked by globally warmer and more stable climates.

Hominin: A classification of **species** comprising humans and our extinct relatives following the split with the common ancestor with chimpanzees.

***Homo erectus*:** An extinct **hominin species** with fossil evidence from at least 1.9 million years to 70 thousand years ago and found from Africa to Indonesia. *H. erectus* may have been the first hominin to leave Africa. *H. erectus* DNA may be retrievable from other species due to **archaic admixture**.

***Homo sapiens*:** The **hominin species** comprising all living humans. Meaning "wise man" in Latin, the name was introduced by Carl Linnaeus in 1758. The earliest fossil evidence of *Homo sapiens* appears in Africa around 300 kya (see **Jebel Irhoud hominins**).

Human-specific gene: A protein-coding gene that is present in humans but absent in other non-human ape species.

Hunting and gathering: A subsistence strategy in which most or all food is obtained by **foraging** and is in contrast to **agriculture**, which rely mainly on domesticated species.

Induced pluripotent stem cell (iPSC): Somatic (body) cells that are artificially reprogrammed to an embryonic-like stem cell state and differentiated into other types of cells.

Inter-birth intervals: The time span between live births.

Jebel Irhoud hominins: The oldest known "early" human fossils discovered, dating to roughly 300 kya from an archaeological site in Morocco. The location of this discovery suggests a "pan-African" origin of humans, with a dispersed interbreeding population, likely aided by climactic factors.

Large language models (LLMs): Machine learning algorithms that can recognize, summarize, translate, predict, and generate human languages on the basis of very large text-based datasets.

Linguistics: The scientific study of human language.

Locus (pl. loci): A unique physical position on a **chromosome**.

Long-read sequencing: A **DNA** sequencing technology that generates reads that are typically longer than 10 kbp in length. To be distinguished from short read sequencing technology (eg. Illumina), which is typically 100-250 bp in length.

Microbiome: The totality of all organisms (microbes) that live on and in the body.

Microbiota: Microorganisms, such as bacteria, archaea, protists, fungi, and viruses, that are found in a particular environment.

microRNA (miRNA): A single-stranded non-coding **RNA** that silences RNA and is involved in post-transcriptional regulation of gene expression.

Molecule: A group of two or more atoms covalently bonded together to form the smallest fundamental unit of a chemical compound that can take part in a chemical reaction.

Neoteny: The delay or slowing of development. Compared to other primates, humans are considered neotenous due to the retention of physiological traits typical of juveniles such as facial features (globular skull shape, thinness of skull bones, reduction of browridge, flattened face, larger eyes), limb length ratio, and behavior.

Nervous system: The network of nerve cells and fibers that transmits nerve impulses between parts of the body.

Neural stem cell (NSC): A self-renewing, multipotent cell that generates the neurons and glia of the nervous system of all animals during embryonic development. Some persist in the adult vertebrate brain and continue to produce neurons throughout life.

Neurogenesis: The process by which **neural stem cells (NSCs)** produce neurons.

Neurolinguistics: A branch of **linguistics** that examines the connection between language and the structure and functioning of the brain.

Neuron: A specialized cell that transmits nerve impulses.

Neuroscience: A multidisciplinary science that is concerned with the study of the structure and function of the **nervous system**. It encompasses the evolution, development, cellular and molecular biology, physiology, anatomy and pharmacology of the nervous system, as well as computational, behavioral and cognitive neuroscience.

Neural progenitor cell (NPC): Cells that are capable of dividing a limited number of times and have the capacity to differentiate into a restricted repertoire of **neuronal** and **glial** cell types.

Non-coding RNA (ncRNA): **RNA** that is not translated into a **protein**. Important ncRNAs include transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), as well as small RNAs such as microRNAs, siRNAs, piRNAs, snoRNAs, snRNAs, exRNAs,

scaRNAs and the long ncRNAs such as Xist and HOTAIR.

Nucleic acid: One of the four classes of major biomolecules. The overall name for **DNA** and **RNA**, which are composed of **nucleotides**. DNA is double-stranded and more stable while RNA is single-stranded and less stable.

Nucleotide: Molecular building blocks for **DNA** and **RNA** Specifically, they consist of three components: a 5-carbon sugar, a phosphate group, and a nitrogenous base. The type of sugar, either deoxyribose or ribose, determines if the resulting **nucleic acid** is DNA or RNA.

Pioneer transcription factor: A type of **transcription factor** that can open and bind to **chromatin**. They control enhancer activation and are important in the recruitment of other transcription factors and in controlling DNA methylation.

Population bottleneck: The dramatic reduction in population size, which often results in a loss of **genetic diversity**.

Pre-frontal cortex (brain): The **cerebral cortex** that covers the front part of the **frontal lobe** and is linked to complex cognitive behavior, personality, long and short-term memory, decision making, speech, language, and a person's will to live.

Protein: One of the four classes of major biomolecules. Proteins are molecules encoded by **DNA sequences** and composed of **amino acids** connected by peptide bonds. These range in size from a few amino acids (short peptides) to large molecules (long polypeptides) comprised of thousands of amino acids.

Protein-coding sequence: A section of **DNA** or **RNA** that codes for **protein**.

Recursion (Language): The ability to embed linguistic structures of similar types within each other.

Ribonucleic acid (RNA): A **molecule** essential in **gene** coding, decoding, regulation, and expression. RNA consists of **sequences** of the four **nucleotide** bases: Adenine, Uracil, Guanine, and Cytosine. Types of RNA include messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA), and other non-coding RNAs. Some viruses including Influenza A and SARS-CoV-2 have RNA **genomes**.

Schizophrenia: A mental disorder characterized by delusions, hallucinations, disorganized speech and behavior, and other symptoms that cause social or occupational dysfunction (DSM-V, 2013).

Segmental duplication: Any repetitive portion of **DNA** arising by **genome** duplication that is at least 90% identical and >1 kbp in length.

Sequence: The linear order of the **nucleotide** building blocks, which encodes individual form and function.

Species: A biological population whose individuals can mate with one another to produce viable and fertile offspring. This is a debated definition and the concept is problematic for extinct fossil organisms for which **DNA** is not available. This definition is problematic in regard to bacteria as they can exchange genetic material across widely separate taxa.

Striding bipedalism: The uniquely-human form of bipedal locomotion, which involves the full extension of the hip and knee joints in the support leg during the stance phase, movement of the hip joint over and in front of the knee and ankle joints in the support leg, and a longer stride length compared to ape bipedalism.

Structural variation (Genomics): The variation in structure of an organism's **chromosomes**. It consists of many kinds of variation in the **genome** of one **species**, and usually includes microscopic and submicroscopic types, such as deletions, duplications, **copy-number variants**, insertions, inversions and translocations that are greater than or equal to 50 base pairs in length.

Telomere: A region of repetitive nucleotide sequences located at the ends of chromosomes that functions to protect **chromosomes** from degradation and fusion with other chromosomes.

Transcription: The first step in **gene expression** during which the **nucleotide sequence** of **DNA** is transcribed into an **RNA molecule** that can ultimately be translated into **protein**.

Transcription factors: **Proteins** that initiate and regulate the **transcription** of **genes**. Transcription factors bind to specific sequences of **DNA** called regulatory elements, or other **proteins** that do so, and directly or indirectly affect the initiation of transcription. The activities of transcription factors determine where and when genes are expressed.