LAST CHANCE BIO

JOSEPH BOYD

1. Chemistry

1.1. **The Periodic Table.** The periodic table (Figure 1) organises the chemical elements into a grid, so as to reflect various chemical properties. Foremostly, the elements are arranged by their *atomic number*, that is, number of protons, from left-to-right and top-to-bottom, with Hydrogen ($AN = 1$) in the top left and synthetic element ununoctium ($AN = 118$) in the bottom right. The first 94 elements are naturally occurring, with the others produced in laboratories. Efforts to produce yet heavier elements persist. Other properties are associated with the row (period) and column (group). For example, elements become more metallic towards the left of the table. Such *periodic trends* were historically used to predict properties of then undiscovered elements. The typical colour shadings indicate groups, for example *noble gases* (elements with similar properties in a gas state[1]. Other groupings such as *heavy metals* (metals with relatively high atomic number) are more loosely defined, and have less regular patterns in the table.

[1]Noble gases still have liquid and solid forms, as do all elements, this being a question of the combination of heat and pressure exerted on them.

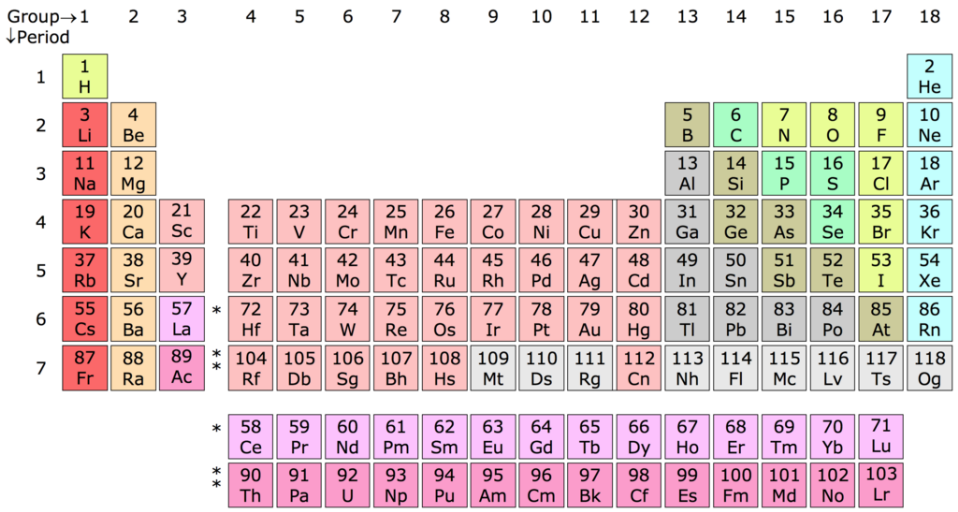

Figure 1. The periodic table of elements.

1.2. **Atoms.** Atoms are the smallest unit of matter having the properties of a chemical element. Thus, any chemical substance consists of one or more atoms. Substances take on one of four states–solid, liquid, gas, and plasma. Atoms are so small that their behaviour is influenced by quantum effects (unlike, in general, meso- or macroscopic objects). All atoms consist of a nucleus, to which are attached electrons by *electromagnetic force*. The nucleus comprises of one or more protons and *a similar number of neutrons*, bound by a *nuclear force*. Protons have a positive electric charge; electrons a negative electric charge. Equality implies electrical neutrality, otherwise the atom is positively or negatively charged (depending on the differential), and is called an *ion* (see *ionisation*). The number of protons determines the chemical element. The number of electrons determines the element's *isotope*. For example, we might compare carbon-12, carbon-13, and carbon-14. All are atoms of the element carbon with 6 protons, but have 6, 7, and 8 electrons respectively.

Radioactivity or radioactive decay occurs when an excess of protons or neutrons in a nucleus overcomes the ability of the nuclear force to hold it together. The nucleus then loses particles and transforms. There are three types of radioactivity: alpha, beta, and gamma, and the process emits electromagnetic radiation, either as light (on the infrared/ultraviolet spectrum), or gamma rays. According to quantum physics, the process is inherently stochastic and cannot be predicted. However, large groups of radioactive isotopes (atoms with inbalanced nucleii) have a predictable *half-life*, that is, the time taken for the radioactive substance to halve in quantity. Half-lives vary wildly across the elements, from mere moments all the way up to billions of years. The radiation from the decay can pose health risks. Radiation poisoning is an immediate risk from overexposure, whereby organs are damaged by the radiation. This can, however, be treated, provided the dosage is limited. Longer-term risks are cancer caused by long-term–though possibly minute–exposure. Background radiation exists at harmless levels everywhere on Earth.

Nuclear fission and nuclear fusion are two ways of unlocking huge amounts of energy from atoms. Both are used in nuclear weaponry (resp. atomic and hydrogen bombs), but only fission is yet used for producing energy (the use of fusion is in its experimental stages). An atom bomb works by firing particles at an unstable isotope, which itself fires off particles in a chain reaction.

1.3. **Molecules.** Molecules are groups of two or more atoms bound together with *chemical bonds*. They are specifically electrically neutral, otherwise they are known as ions[2]. Molecules may be homonuclear (e.g. H_2), or heteronuclear (e.g. H_2O). The word molecule itself comes from the Latin word *mole* meaning *mass*. Note the term mole is used to describe the same number of atoms of a given chemical element as there are in 12 grams of electrically

[2]Ions refer to both charged atoms or molecules.

neutral carbon (carbon-12) (^{12}C). This is $\approx 6.02 \times 10^{23}$ atoms, a figure known as Avogadro's constant. Though molecules are common in nature and biology, not all matter is constructed from recognisable molecules.

The chemical bonds that bind atoms together come in different forms and different strengths. Bonding may be done with the attracting *electrostatic force* between atoms of opposite charge, or with the covalent bond of sharing electrons.

1.4. **Chemical Equations.** Chemical equations are used to model chemical reactions. A chemical reaction is a rearrangement of molecules into a product substance when two or more reactant substances are combined. Solving the equations balances the mass of each element, following the law of conservation of mass. They do not, however predict the molecular form of the new substance.

We start with the chemical formula representation of our reactants, which indicate the proportion of atoms in the molecules. For example, C_5H_{12} and O_2. The chemical reaction is denoted with a plus (+) sign and an arrow ($\rightarrow$),

$$C_5H_{12} + O_2 \rightarrow CO_2 + H_2O,$$

where the chemical products on the right hand side are known in advance. Now we address each element in turn. There are five carbon atoms on the left, and only one on the right. Therefore we update to,

$$C_5H_{12} + O_2 \rightarrow 5CO_2 + H_2O$$

Now we require 12 Hydrogen atoms on the right,

$$C_5H_{12} + O_2 \rightarrow 5CO_2 + 6H_2O$$

but now there are 16 oxygen atoms, therefore we modify the left hand side to,

$$C_5H_{12} + 8O_2 \rightarrow 5CO_2 + 6H_2O,$$

and we have achieved balance.

2. Molecular Biology

[From "Kernel Methods in Computational Biology" (Schölkopf et al.)]

The cell is the atomic unit of living organisms. It is roughly a fluid solution comprising of certain molecules, surrounded by a lipid (fat) membrane. They range in size from $1\mu m$ to $100\mu m$ (micrometers, also known as *microns*). Cells have various properties: metabolism, replication, reaction to environment, etc. Viruses have some of the same properties, but have no metabolism (relying on the host) and are usually not considered to be alive. Thus, they are in a grey area between life and chemical matter.

Living organisms are either: prokarya (single-celled organisms); or eukarya (everything else)[3]. Eukaryotic cells are more complex, containing hereditary information and mitochondria (energy suppliers). Three polymers[4] are present in cells: DNA, RNA, and proteins (polypeptides).

2.1. **Polymers.** Polymers are the bridge between chemical matter and that which can be described as living. They are *macro-molecules*, occupying the smallest scale at which variation exists in like structures (e.g. a water molecule is a water molecule, but there are many forms DNA can take on). This polymorphism is what ultimately leads to the diversity within and without living organisms. Cells of the same type can have very different morphologies and inaccuracies in their replication–something that can lead to problems.

2.1.1. *DNA.* DNA molecules make up the cell's *genome*–heritable information. DNA consists of linearly-linked nucleotides (small chemical compounds), adenine, cytosine, guanine, and thymine. A and T bind together to make base pairs with hydrogen bonds, as do C and G, making them *complementary nucleotides* (see Section 2.1.4). DNA consists of two linear chains of nucleotides arranged in a double helix. This structure improves stability (amongst other things).

In eukarya, the DNA is situated in the cell nucleus. The human genome consists of 3 billion nucleotides, separated into 23 separate DNA molecules called *chromosomes*. In sexual organisms, two (or more) versions of chromosomes are in each cell, one from each parent.

2.1.2. *RNA.* RNA is *ribonucleic acid*–similar to DNA but single-stranded and with uracil (U) rather than thymine. RNA is produced by substrings of the genome called *genes*, in a process known as *transcription*, as part of the process of *gene expression*. This is the mechanism by which genetic information (indirectly) specifies an organism's properties. The RNA is a complement of part of the gene. The majority of genes encode information to produce proteins, following the flow:

$$\text{DNA} \to \text{messenger RNA (mRNA)} \to \text{protein}$$

This classic flow is known as the *central dogma of molecular biology*, and was introduced by British molecular biologist Francis Crick (1916-2004) in 1958.

[3]Life → domain (prokarya, eukarya) → kingdom (plant, animal), phylum, class, order, family, genus, species

[4]Polymers consists of repeating molecular subunits called *monomers*, which are covalently linked.

2.1.3. *Proteins.* Proteins or *polypeptides* are polymers made up of amino acids (as opposed to nucleic acids) and carboxyl acids, which form peptide (special covalent) bonds. They link together to form the protein backbone. Proteins fold themselves into a spaghetti-like mess. Proteins have a diverse range of functions in cells. For example, metabolism[5], energy, communication, and reproduction.

2.1.4. *Nucleotides.* The building blocks of DNA and RNA. They consists of *bases* Adenine (A), Cytosine (C), Guanine(G), and Thymine (T). In RNA, Uracil (U) stands in for Thymine. All contain nitrogen. These bases are connected with a sugar (ribose in RNA, deoxyribose in DNA) and one or more phosphate groups to form *nucleotides*. This creates *phosphodiester linkage* between bases, resulting in a molecule with a certain orientation (5 prime to 3 prime). A strand of DNA thus arises by chaining nucleotides linearly. DNA is formed by two interwoven linear strands forming a double helix. The strands are held together by hydrogen bonds between complementary *base pairs*. Adenine and Thymine form a base pair, as do Cytosine and Guanine. This structure was discovered by Francis Crick and James Watson in 1953, for which they later received the Nobel prize in Physiology or Medicine.

2.1.5. *Histones.* Despite living organisms containing up to several billion base pairs amounting to a few meters of DNA, it is able to be packaged within a nucleus only a few micrometers in diameter with a group of structuring proteins known as *histones*.

2.1.6. *Chromatin.* The groupings of DNA and histones form a beaded chain known as *chromatin*. The beads are the repeated wrapping of DNA around histone proteins, and the chain is the bare DNA.

2.1.7. *Chromosomes.* The chromatin is further packed into chains, constituting what are known as *chromosomes*. Chromosomes are the distinct parts of the genome. In humans, there are 46 chromosomes. Chromosomes define the units of exchange during mitosis and meiosis.

It is only during a particular phase of cell division that chromosomes manifest in the familiar **X** formation through a compaction mechanism. This compaction helps to prevent breaks during cell division. At other times, the chromosomes exist in a sort of soupy arrangement.

3. Cells

Eukaryotic cells are divided into many compartments called *organelles*, each with their own membrane. The nucleus is the most prominent part, in which is stored the (tightly packed) genome. The non-nuclear parts are

[5]Special proteins called *enzymes* catalyse reactions to create small molecules to refresh the cells contents e.g. DNA, RNA, lipid membrane etc. Thus, cells are like chemical factories

known collectively as the cytoplasm–comprising of the organelles and the cytosol liquid solution. The cytoplasm is structured by the *cytoskeleton*. Microtubules, which are essentially polymers shaped as hollow tubes, are an important component of the cytoskeleton, giving the shell its shape.

3.1. **Heredity.** Heredity is the transmission of physical characteristics to offspring, usually in the form of genetic information. The origins of the theory, Mendelian inheritance, are attributed to Austrian friar Gregor Mendel (1822-1884).

3.1.1. *DNA replication.* DNA replication (Figure 2) begins with DNA in its unpackaged chromosomal form. An enzyme[6] called *helicase* runs through the double helix, unzipping the two strands by breaking the hydrogen bonds between the nucleotides. This results in two 'templates': the leading strand, running 5' to 3'; and the lagging strand, running 3' to 5'. Single strand binding (SSB) proteins acting as placeholders attach to the freed strands to prevent reattachment. For the leading strand, another enzyme, DNA Polymerase III runs along the binding proteins and synthesises complementary nucleotides.

For the lagging stand, this synthesis is more complicated. As it runs in the opposite direction, the desired starting end is perhaps not yet separated by the helicase. It is therefore synthesised in pieces called Okozaki fragments[7]. An another enzyme called *primase* creates *RNA primer*, that is, a short sequence of placeholders. The synthesised strand is known as a *primer strand*. DNA Polymerase III again lays down new DNA. This process is done iteratively, creating the fragments. DNA polymerase I replaces the RNA primers with DNA, and finally *DNA ligase*[8] links the fragments together.

3.1.2. *Mutations.* Mutations are aberrations in DNA. They may occur for several reasons. One reason are errors in DNA replication. This occurs with a probability rate[9] of about 10^{-10}. Other causes are the effect of a virus infection and external agents such as radiation. Mutations are essential for evolution, but can also cause cancer.

3.1.3. *Genome.* The genome, like other *-omes* in biology refers to a totality, namely the totality of genetic material found in a cell. DNA sequencing, for example Sanger[10] sequencing is used to read the nucleotides. The Human Genome project was a US$3B project (the largest in the history of biology) running from 1990-2003 to sequence the entire human genome (3.2×10^9 base pairs). The project involved first fragmenting the sequence into millions of

[6]Macromolecular biological catalyst

[7]Named after Tsuneko Okazaki (1933-), a Japanese scientist.

[8]A ligase is a special type of enzyme that catalyzes the formation of new chemical bonds.

[9]Can aging therefore simply be the stochastic accumulation of cell mutations?

[10]Frederick Sanger (1918-2007) was an English biologist and one of only four people to receive two Nobel prizes.

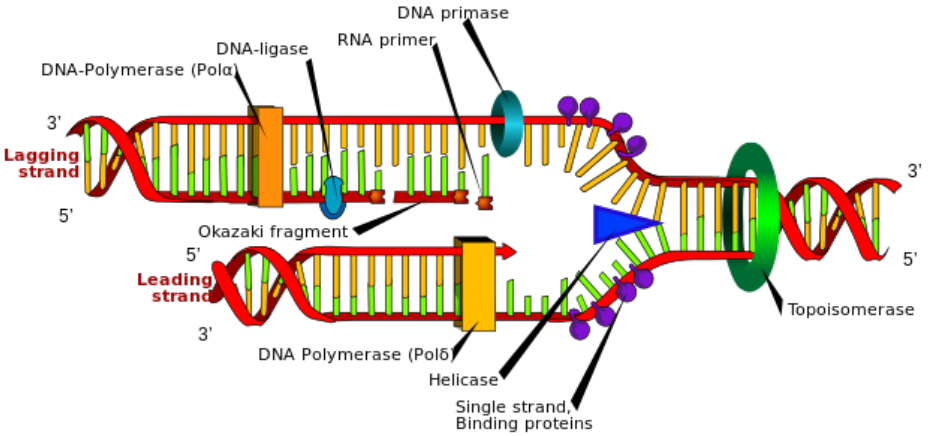

FIGURE 2. The components of DNA replication.

small, yet manageable chunks. It was found that the number of genes is surprisingly low, as is the proportion of *coding DNA*[11] at about 2% of the genome. Individual humans differ by no more than 0.1%.

3.2. **Gene Expression.** Cells read the genome in a process call *gene expression* using RNA. Genes are sequences of DNA encoding a particular function. The linear structure of RNA allows for it to fold into functional structures. There are many forms of RNA, coding and non-coding. Messenger RNA (mRNA), about 5% of the total, are used for coding.

3.2.1. *Transcription.* The transcription process involves RNA polymerase II, which binds to a *promoter*, a region of DNA, upstream of the transcripted gene. The promoter indicates the start site for transcription. This polymerase adds nucleotides to the RNA molecule until the transcription is complete, upon which the RNA is released. The error probability for RNA transcription is much higher than DNA replication, at a probability of about 10^{-4}. The promoter regions are identified by a certain sequence of nucleotides between 100 and 1000 base pairs in lengths. They have various forms, the most important being the TATA box. Note that promoters are examples of non-coding DNA, and it may be appreciated how DNA contains large amounts of information overhead to enable the genetic information to be read[12]. As with DNA replication, processing moves from the 5' end to the 3' end of the strand. Note that either strand of DNA may be used for transcription. RNA is further 'post-processed' after transcription before leaving the cell nucleus for translation (expression). An important process

[11]DNA involved in protein expression

[12]Apart from this there is significant redundancy from our evolutionary past, which we presumably share with many other species? Could it therefore be expected, therefore, that the average genome gets longer and longer as evolution plays out into the future?

is *splicing*, whereby the transcribed RNA is divided to express for multiple proteins. The process is performed by the cell's *spliceosome*. Thus, DNA is something like a class, RNA something like object instances of (parts of) the class.

3.2.2. *Translation.* Proteins are composed of monomers known as amino acids. Amino acids have a certain molecular structure. There are 20 varieties, though all have the same basis. Amino acids are linked with *peptide bonds* to create proteins. Proteins are thus an example of *polypeptide*[13], though are often distinguished as such due to proteins being longer and folding to have a three-dimensional structure. This folding is permitted by the covalent bonds involved. Proteins exhibit various structures, such as an α-helix, and β-sheet, as well as various higher level structures. Proteins have many functions in cells, from enzymatic functionality, to storage, transportation, signaling and reception[14]. The four character nucleotide alphabet of RNA is translated to the 20 character amino acid alphabet of proteins. This translation mapping is known as the *genetic code*. Each nucleotide triplet is known as a *codon*, specifying an amino acid to add during protein synthesis. A start codon is usually 'AUG'. The code can be represented easily with a lookup table. Translation is achieved by a *ribosome*, the set of cellular components responsible for protein synthesis. An important component of this process is *transfer RNA* (tRNA), RNA designed to read the codons of the mRNA in sequence. Though all cells contain the same DNA, they differ with respect to the level of RNA and proteins. Indeed, the composition of proteins is a key part of what gives different cell types their unique characteristics.

3.3. **Cell Structure.** All life forms consist of cells, be it a single cell in the case of bacteria, or hundreds of billions of cells in the case of a human being. Eukaryotic cells contain a nucleus, whereas prokaryotic cells do not. A cell is bound by its membrane. Within the membrane are the cytoskeleton, the structure-providing components (microtubules, actin filaments, etc.), and the cytosol, the aqueous substance in which internal molecules float. The core of the cell is the nucleus, which itself has a porous *nuclear membrane* or *nuclear envelope*. The nucleus contains the DNA and the *nucleolus*, a component that produces ribosome for RNA translation. The mitochondrion (pl. mitochondria) is the energy supply of the cell. There are various other components having the name *organelle* also. Organelles are sub-components of the cell, usually with their own membrane.

3.3.1. *Cell Membrane.* Made mostly of lipid, and some proteins. The cell membrane controls the flow of compounds in and out of the cell.

[13]Simply, polymer linked with peptide bonds

[14]Proteins in the retina, for example, are involved in the detection of light.

3.3.2. *Cytoskeleton.* The cytoskeleton consists of microtubules, actin filaments, and so-called"intermediate filaments". These give the cell its shape, and are important in cell behaviours such as cell division.

Microtubules are hollow tubes of a repeating spiral of 13 α-tubulin and β-tubulin monomers[15].

3.3.3. *Golgi apparatus.*

3.3.4. *Mitochondria.*

3.4. **Cell Cycle.** The lifecycle of a cell has various phases: G_1 (first gap–growth), S (synthesis–DNA replication), G_2 (second gap–growth for mitosis), M (mitosis). The sum of the non-mitotic phases are known as the *interphase*. It is in interphase that a cell spends most of its life. At the end of this cycle, the cell is replaced by two *daughter cells*, otherwise it dies. The lifecycle lasts for anything from a few minutes to more than a year, depending on the cell type.

In the interphase, the cell grows by producing proteins and by duplicating the chromosomes for cell division. Within the interphase, the cell may enter a dormant state called G_0, where it may stay temporarily or permanently. This may be triggered by signals from neighbouring cells. There is a point in G_1 phase after which a cell is committed to mitosis and may not enter G_0.

3.4.1. *Mitosis.* Within the mitotic phase, there is pro-phase[16], pro-metaphase, meta phase, anaphase.

After DNA replication, we have duplicated chromsomes called *sister chromatids* that are bound together by proteins at a point called the centromere. In pro-phase, the chrosomes are tightly packed to avoid breakages during division. Prior to this, that is, during the interphase, the DNA is loosely packed chromatin. Gene transcription ceases, and the nucleolus disappears. *Centrosomes* combine to form a *spindle* aparatus.

In prometaphase, the mitotic spindle is fully assembled with microtubules. Kinetochore microtubules bind to kinetochores (protein structures) on the centromere of each chromatid pair.

In metaphase, the chromosomes are pulled apart, and equally distributed in opposite halves of the nucleus.

In anaphase, the *cohesins* binding the sister chromatids are severed.

In telophase, new cell membranes and nucleoli form around the separated chromatids. Mitosis is complete, with two daughter nuclei formed.

3.4.2. *Cytokenesis.* Cytokenesis is considered separate from mitosis, but is still a part of cell division, and therefore part of M phase.

---

[15]These monomers *poymerise* to create the polymer microtubules.

[16]In plant cells there is an additional initial phase called preprophase that readies the nucleus by placing it in the center of the cell

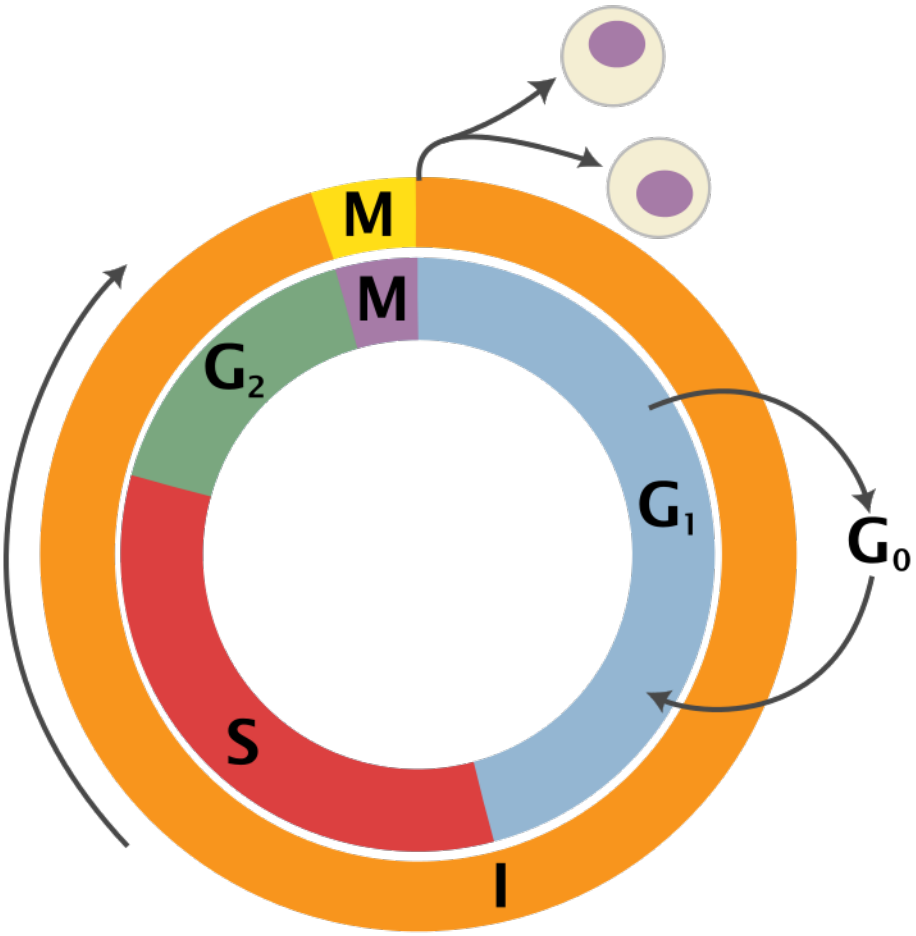

FIGURE 3. The cell cycle

3.4.3. *Meiosis.* Meiosis is a special kind of cell division where, again proceeding the interphase phases, the cell divides twice, leaving four haploid *gametes*, that is, containing one copy of each chromosome. In sexual reproduction, these gametes (a female ovum and a male sperm) meet to form an embryo. Meiosis consists of two divisions, Meiosis I and Meiosis II, including recombination steps assuring random variation. There is nothing particularly special about having four gametes; almost all will never make it to the embryo forming stage. Once gametes meet, an embryo develops over time through many cycles of mitosis. A miscarriage is usually the result of an error in meiosis resulting in an abnormal number of chromosomes, a situation call aneuploidy. Down syndrome (trisomy 21) occurs when a third copy of chromosome 21 is included in the embryo. Humans have 23 pairs (46) chromosomes. They are numbered 1-22, then X and Y. Both females and males

have pairs of each of the first 22, but females have two X chromosomes, and males have an X and a Y.

3.4.4. *Apoptosis.* Apoptosis is *programmed cell death* (as opposed to *necrosis* where a cell dies due to trauma). It is initiated by various signals, and consists of the breaking down of the cell. Problems with apoptosis can lead to diseases, for example, too much leads to atrophy (disintegration of bodily functions); too few can lead to cancer.

3.4.5. *Senescence.* Senescence is a condition by which cells or indeed entire organisms "grow old". On the cellular level, this corresponds to a cessation of cell division.

3.4.6. *Cell Lines.* A *cell culture* is a collection of cells extracted from a plant or animal and grown in an artificial environment. A *cell line* is a cell culture from a single initial cell. An immortalised cell line is one that has undergone a certain mutation such that it can continue in perpetuity, usually *in vitro*. The first such cell line came from the tumour of Henrietta Lacks, an American woman who died in 1951, and is known as the HeLa cell line.

4. Tissue

5. Cancer

Cancer are a group of diseases involving abnormal cells growth. If cancer is malign, it means it spreads to other parts of the body. There are over 100 types, the most common cause being tobacco smoking, infections such as hep B and hep C, and poor eating/lifestyle. Cancer is first detected by screening, then by further imaging and biopsy (extraction of sample cells or tissue). Cancer is treated by some combination of radiation therapy, chemotherapy, and surgery. Palliative care (providing relief) is a big part of treatment. The average five-year survival rate is on the order of $66\% - 80\%$.

- assay
- osteocytes??
- biomarker e.g. γ-H2AX
- antigen/antibody/pathogen
- kinase
- pleitropic
- carcinogenesis
- cell cycle
- immunology
- monoclonal/polyclonal
- metastatic
- cytotoxicity
- buccal = oral epithelial
- Phase I, II, III trials

- olaparib, veliparib
- in vivo, in situ, in vitro, in utero, ex vivo
- radiotherapy
- epigenetic
- S-phase
- senescence
- omics
- genes
- actin, tubulin, keratin
- mitochondria
- adenosine triphosphate
- repressors, activators (gene regulators)
- Model organisms like E. Coli (bacteria), C. Elegans (nematode/roundworm), Drosophila (fruit fly), Saccharomyces cerevisiae (yeast)
- micro RNA etc.
- RNAi (interference)
- polylobed
- p53
- translocation
- pathway
- exon
- intron
- kinase
- allele phenotype
- GWAS
- petri dish

6. Drug discovery

Reference the obvious papers, but also image-based profiling (Loo)

6.1. **High content screening.** In this section we describe the key concepts pertaining to high content screening. We first discuss the living organisms at the heart of the scientific study–cell lines–in Section 6.1.1. We then descibe the specifics of high content screening and fluoresence microscopy in Section 6.1.2.

6.1.1. *Cell lines.* An immortalised cell line (hereafter referred to as a *cell line*) is a population of cells, sustained by cell division without senescence[17]. In this way, a cell line continues replicating indefinitely from a common ancestor. This evasion of senescence is achieved through a certain mutation overriding the normal cell cycle. This mutation may be naturally occurring, as in some cancer cells, or it may be induced by, for example, viral transfection. The first and most well-known cell line generated is the HeLa cell line,

[17]Senescence is the process in which cells "age" and cease to perform mitosis. The limit on repeated mitoses is known as the *Hayflick limit*[**?**].

named after its source in Henrietta Lacks, an American woman who died of cervical cancer in 1951. Since HeLa, many cell lines have been produced. Cell lines are useful biological models due to their longevity in cell culture, and are used extensively in biomedical research, for example in assessing the cytotoxicity[18] of a drug treatment. It must be noted, however, that their accuracy as biological models is curtailed by their character as mutated cells, and the effects of repeated passages, cloning, and biochemical contaminants can lead to significant genetic drift from their *in vivo* ancestors[**?**].

Triple-negative breast cancer (TNBC) is a variety of breast cancer whose cells do not express proteins *estrogen receptor*, nor *progesterone receptor*, nor do they amplify HER2/neu[19]. As these genetic markers are common targets for cancer therapies, TNBC is more difficult to treat than other breast cancers. TNBC is responsible for 10%-15% of breast cancers[**?**]. Our pilot study consists of microscopy of cancer cells from the MDA-231 and MDA-468 cell lines[**?**].

6.1.2. *High content screening.* High content screening (HCS) or high content analysis (HCA), is an approach to drug and target discovery whereby fluorescence microscopy imagery (Section 6.1.3) is the chief source of information. In HCS a cell line population is exposed to *small molecule* drug compounds or RNAi and the effects are assessed visually with microscopy. Note that HCS overlaps with the related *high throughput screening* (HTS). *High throughput* refers to the high volume of trials that may be conducted in an assay. *High content* refers to the richness of image data as a source of biological information. Bioimages provide a richness of perspective on cells that is typically lost in more classical *omics* data [**?**]. Various advancements in fluorescence microscopy and image analysis have pushed HCS to the early "hit-to-lead" stages of the drug discovery process[**?**].

In an assay, a cell line population will be distributed over the wells of a microtiter plate[20], before various small molecule compounds are added and the plate enters incubation. In this way, multiple experiments can be effectuated in parallel. For experimental purposes, a subset of wells will contain the neutral substance *dimethyl sulfoxide* (DMSO), and another subset will contain no substance at all. After incubation, a *fixation* procedure is undertaken whereby the chemical methanol is applied to the wells to protect the cells from decay. The wells are stained with various fluorescent stains that bind to key areas of the cells. Following this, a washing procedure is applied, whereby extraneous dye content is aspirated from the wells, preserving only the chemically bound dye compounds illuminated for the microscopy. One limitation of this necessary washing step is that it may remove cells

[18]Cytotoxicity refers to toxicity (the extent to which something is harmful) towards a cell.

[19]human epidermal growth factor receptor 2

[20]Microtiter plates (hereafter *plates*) are small (usually polystyrene) trays divided into a rectangular grid of wells functioning as individual test tubes.

Stain	Marker
DAPI	A-T regions of DNA
Cyanine 3 (Cy3)	DNA double-strand bbreaks
Cyanine 5 (Cy5)	β-tubulin of microtubules
FITC	Phospholipids in cell membrane

TABLE 1. The stains used in the fluorescence microscopy of our screen and their corresponding biological markers.

dead prior to fixation, ostensibly leaving an absence of cells as a proxy for high levels of apoptosis[21], indicating a cytotoxic compound. Some number of images corresponding to non-overlapping *fields* or positions within each well are taken, typically producing from hundreds to thousands of image perspectives per plate.

6.1.3. *Fluorescence Microscopy.* Fluorescence microscopy uses a laser to excite fluorescent molecules in organic matter. These molecules are known as *fluorophores*. In HCS, cells are exposed to fluorescent stains so as to highlight key regions. The stains used in our screen are listed in Table 6.1.3. *DAPI* (4',6-diamidino-2-phenylindole) is a bright blue stain that permeates a cell's nuclear membrane and binds to regions of DNA rich in A-T[22] base pairs. DAPI is effective in highlighting the cell nucleus, where the DNA is housed[**?**]. Cyanine 3 (Cy3) and cyanine 5 (Cy5) are commonly used dyes belonging to a common family. Here, Cy3 is conjugated to a γ-H2AX antibody. γ-H2AX is a protein that naturally binds to double-strand breaks as a marker for the repair of DNA damage[**?**]. Furthermore Cy5 is conjugated to a β-tubulin antibody, thus highlighting this sub-component of the tubulin polymer that constitutes *microtubules*, which are structure-providing components of cells. Finally, FITC (fluorescein isothiocyanate) highlights phospholipids, molecules in the lipid cellular membrane.

6.1.4. *Phenotypes.* An organism's phenotype is the ensemble of its observable characteristics. This is in contrast to its *genotype* or genetic code, which is fundamentally responsible in determining the phenotype. Other factors contributing to a phenotype are *epigenetic* and environmental factors. While in principle all observable characteristics consistute the *phenotype*, we normally refer to descriptions that are biologically interpretable and thus characterise a certain biological aspect of the biological model used in the study. For instance, we can observe cellular states, such as cell cycle phases or aberrant nuclear morphologies: *interphase*, the longest part of the cell cycle, where the cell nucleus is typically small and convex; *large interphase*, corresponding to an abnormally large interphase nucleus, with an otherwise normal shape, potentially a result of a defect in DNA replication or nuclear

[21]cell death

[22]Adenine and thymine

membrane control; *bright interphase*, where DAPI exhibits higher floures-cence, presumably due to cell cycle deregulation; *prometaphase*, a phase of mitosis just prior to nuclear division only rarely observed in unperturbed cells as the phase is relatively short; *metaphase*, the mitotic phase in which the cell's chromosomes are aligned in the metaphase plate prior to chromo-some segregation; *apoptosis*, cell death; *polylobed*, the cell nucleus takes on an abnormal shape, due to problems in the division process.