JEM 125th Anniversary

From the discovery of DNA to current tools for DNA editing

Pascal Maguin[1] and Luciano A. Marraffini[1,2] (ID)

In 1944, the *Journal of Experimental Medicine* published the groundbreaking discovery that DNA is the molecule holding genetic information (1944. *J. Exp. Med.* https://doi.org/10.1084/jem.79.2.137). This seminal finding was the genesis of molecular biology and the beginning of an incredible journey to understand, read, and manipulate the genetic code.

During the first half of the 20th century, it was hypothesized that proteins carry genetic information, but this changed in 1944, when three scientists at The Rockefeller Institute made the fundamental discovery that DNA is the genetic material and forever changed our understanding of the living world (Avery et al., 1944). Oswald T. Avery, Colin M. MacLeod, and Maclyn McCarty published a study in the *Journal of Experimental Medicine* establishing that DNA purified from virulent type III *Streptococcus pneumoniae* could convert an avirulent type II *S. pneumoniae* into a virulent type III strain (see figure). They based their research on Frederick Griffith's experiment showing that mice injected with avirulent *S. pneumoniae* together with heat-killed virulent *S. pneumoniae* succumb to the infection, and that the bacteria retrieved from the dead mice are of the virulent type (Griffith, 1928). Avery and his team sought to isolate and identify the chemical entity responsible for this transformation. They first undertook a series of careful purification processes of different pneumococcal extracts and isolated a pure solution of the transforming agent. Chemical analysis revealed that the substance had the same carbon, hydrogen, phosphor, and nitrogen composition as DNA. Furthermore, a Dische's chemical test for the presence of DNA was

positive, but Biuret and Millon tests for the presence of proteins were negative. Treatments of the agent with a purified ribonuclease and different proteases did not abolish the substance's transforming capability, ruling out the possibility of RNA and protein as its main components. Next, they wanted to see whether degradation of DNA would eliminate the transforming activity. However, they lacked a DNase, and therefore they tested various sera and organ extracts and found that some could completely inactivate the transforming material. Importantly, only the extracts that degraded a pure sample of DNA abolished the material's transforming activity. Electrophoresis and UV spectroscopy studies also suggested DNA as the transforming agent. Based on their careful and elegantly executed experiments, the authors concluded that "the evidence presented supports the belief that a nucleic acid of the deoxyribose type is the fundamental unit of the transforming principle of Pneumococcus Type III." Although their results were sound, the science community raised the possibility that some trace amount of impurities in their *S. pneumonia* extract could be the real transforming agent. This concern was also raised by the authors in their publication: "It is, of course, possible that the biological activity of the substance described is not an inherent property of the


Insights from Pascal Maguin and Luciano A. Marraffini.

nucleic acid but due to minute amounts of some other substances..." McCarty and Avery published two follow-up articles in 1946, also in *JEM*, to address the concerns raised from their first publication (McCarty and Avery, 1946a; McCarty and Avery, 1946b). In these articles, they refined their purification method and showed that purified DNase could inactivate the transforming activity, thus providing further evidence that the transforming agent is DNA. In 1952, Alfred Hershey and Martha Chase showed that DNA from bacteriophage is the only substance entering bacteria upon infection (Hershey and Chase, 1952), further cementing the idea of DNA as the genetic material. The Avery–MacLeod–McCarty experiment placed DNA in the spotlight of science research and can be considered the birth of

[1]Laboratory of Bacteriology, The Rockefeller University, New York, NY;   [2]Howard Hughes Medical Institute, The Rockefeller University, New York, NY.

Luciano A. Marraffini: marraffini@rockefeller.edu.
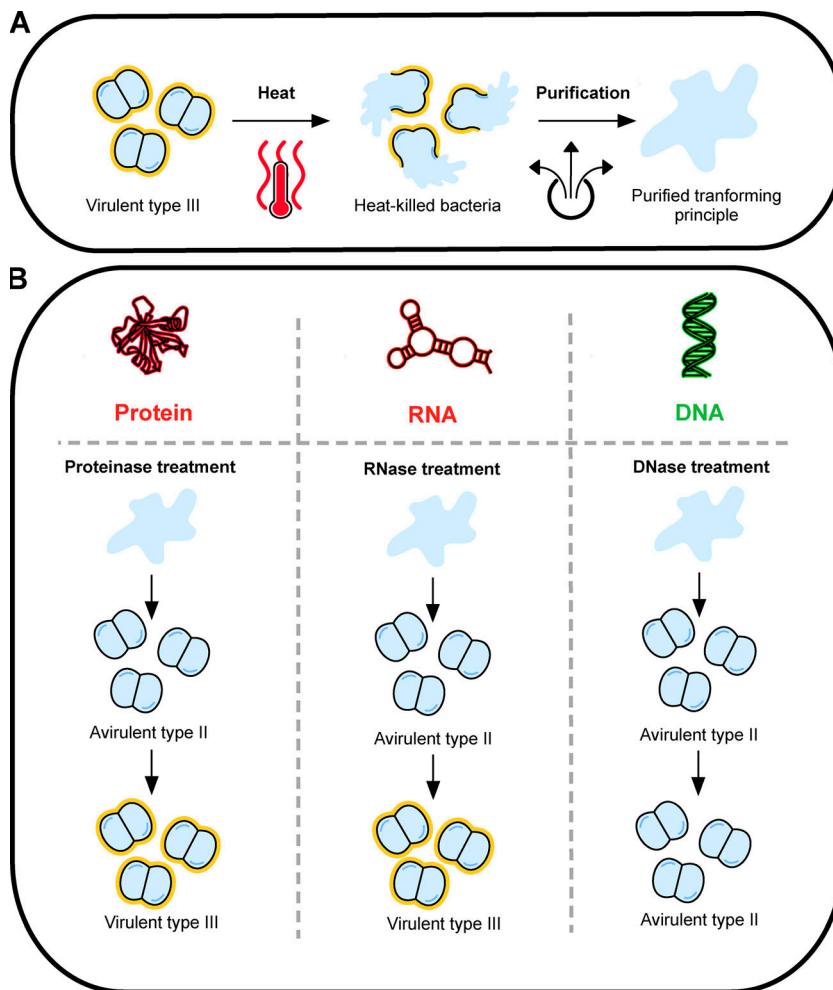
The Avery–MacLeod–McCarthy experiment. **(A)** Avery and his team first isolated a pure extract capable of transforming type II *S. pneumoniae* into virulent type III from heat-killed type III *S. pneumoniae*. **(B)** Then, the transforming material was treated with different enzymes to assess its chemical properties. Proteinase and RNase treatment did not abolish the transformation capabilities of the extract. On the contrary, treating the extract with crude enzyme preparations and organ sera capable of degrading a pure sample of DNA rendered the material incapable of converting *S. pneumoniae* from type II to type III. Thus, the results demonstrated that DNA is the main component of the extract and that it contains the genetic information necessary to convert type II *S. pneumoniae* to type III.

decryption of the 64 codons of the genetic code by 1966 (Szymanski and Barciszewski, 2017).

Once the genetic code was solved and basic questions about DNA such as its replication and transcription were answered, a new era in molecular biology emerged: DNA manipulation. Well before the genetic code was solved entirely, microbiologists had observed the phenomenon of host-controlled modification and restriction in bacteria in the early 1950s (Bertani and Weigle, 1953; Luria and Human, 1952). This led to the discovery of restriction-modification systems, bacterial immune systems capable of recognizing and cleaving incoming viral DNA (Loenen et al., 2014). In 1970, the first restriction enzyme able to cut a specific DNA sequence was isolated (Smith and Wilcox, 1970), and a few years later, recombinant DNA was obtained using these restriction enzymes to cut and paste different pieces of genetic material (Cohen et al., 1973; Jackson et al., 1972). This was the start of molecular cloning, allowing scientists to isolate and study specific genes and to produce proteins from one organism into a less complex organism. Human insulin was produced for the first time in bacteria in 1979 (Goeddel et al., 1979). Quickly after that, genome editing of plants and mice followed, advancing agriculture and medical research.

At the same time DNA manipulation rose, other scientists were looking at whether they could read the information stored in DNA. In 1977, two techniques were developed independently to sequence DNA, the Sanger and the Gilbert methods (Maxam and Gilbert, 1977; Sanger et al., 1977). Using the Sanger method, the genome of bacteriophage phiX174 was sequenced in 1977 (Sanger et al., 1978). These methods were improved and automated in the 1980s, leading the way for the human genome project in the 1990s. In 2001, the first draft of the human genome was published, a tremendous advance for science (Lander et al., 2001).

The beginning of this century marks a new era for DNA research characterized by the rise of next-generation sequencing and the discovery of molecular scissors enabling precise DNA editing. During the last 15 yr, methods to sequence millions of different DNA sequences in one reaction, known as next-generation sequencing, have been

molecular biology. Today, less than 80 yr since this seminal discovery, colossal advances have been made toward our understanding of DNA, from the ability to decode and read DNA to the precise editing of its sequence.

After the discovery of DNA as the molecule holding the code of life, scientists sought to crack the code. Soon after the Hershey–Chase experiment, work from Rosalind Franklin, Francis Crick, and Jim Watson elucidated the iconic double helix structure of DNA (Watson and Crick, 1953). Based on its structure, the basis for DNA replication as a semi-conservative process was hypothesized and then later demonstrated by Matthew Meselson and Franklin Stahl (Meselson and Stahl, 1958). However, the question of how a DNA molecule could encode the richness of the genetic information was still unanswered. A major breakthrough came from the poly-U experiment by Marshall Nirenberg and J. Heinrich Matthaei, which showed that in a cell-free protein synthesis system, adding synthetic RNA made up of only uracil resulted in the synthesis of a poly-phenylalanine amino acid chain (Nirenberg and Matthaei, 1961). This demonstrated that multiple uracil code for the amino acid phenylalanine. More studies from Nirenberg and others resulted in the complete

developed (Shendure et al., 2017). These sequencers rely on the sequencing of small DNA fragments that can be assembled together to reconstruct genomes. Today, third-generation sequencers capable of reading long sequences of DNA exist, which makes the assembly of difficult genomes with repeating sequences of DNA possible. Furthermore, these sequences can read the modification state of DNA, pushing forward the field of epigenetics. At the same time that the quality of DNA sequencing improved, its cost plummeted. This led to widespread access of complete genome sequences, which provided a pathway to an understanding of CRISPR-Cas (CRISPR-associated) systems, the next major breakthrough in DNA manipulation.

In 1995, microbiologists discovered stretches of DNA with short repeating sequences separated by short unique sequences in the genomes of some Archaea (Mojica et al., 1995) and later described them as CRISPR. 10 yr later in 2005, the mystery surrounding the unique sequences between the repeats was solved thanks to rise of DNA sequencing and publicly available genome sequences. A search for these sequences (known as spacers) in public DNA databases revealed that they matched sequences from bacteriophage and mobile genetic elements such as plasmids (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). Later, CRISPR and Cas genes were found to be a novel prokaryotic defense system providing resistance against foreign nucleic acids (Barrangou et al., 2007; Marraffini and Sontheimer, 2008). Guided by a short RNA derived from a spacer sequence, a single Cas protein or a complex of them cleave the foreign DNA at the location matching the spacer sequence (Marraffini, 2015). While in bacteria the double-stranded DNA breaks (DSBs) generated by CRISPR RNA–guided Cas nuclease destroy the invader's genome (Garneau et al., 2010), they are the first step used in most methods to introduce site-specific mutations in eukaryotic organisms. Cells use either nonhomologous end joining to repair the break while creating random mutations at the site, resulting in gene disruption, or homology-directed repair to introduce a specific sequence at the cut site using a DNA template for recombination (Ceccaldi et al., 2016). The potential use of Cas RNA–guided nucleases as molecular scissors for genome editing did not go unnoticed by researchers, and in 2012 it was demonstrated that Cas9 (a Cas protein belonging to a specific CRISPR-Cas system), together with a short RNA guide, could cut DNA in vitro (Gasiunas et al., 2012; Jinek et al., 2012). Shortly after, Cas9 was used to cut DNA and mediate genome editing in human cells (Cong et al., 2013; Mali et al., 2013). CRISPR's adoption by the science community was instantaneous because the method is relatively easy and inexpensive. Today, the technique is used in a wide range of cell types and organisms in laboratories to characterize and study specific genes. In the clinical setting, scientists and doctors are hoping to be able to treat human genetic diseases in the near future. For example, sickle cell anemia is caused by a single nucleotide mutation in the β-globulin gene, and there are ongoing efforts to look at whether hematopoietic stem cells derived from a patient could be edited in vitro to fix the mutation and then readministered in the patient (Ledford, 2019). Last year, a phase 1 clinical trial for the treatment of the eye disease Leber congenital amaurosis started using direct delivery of Cas9 in the human eye to edit the mutation causing the disease and restore vision (Ledford, 2020). Although Cas9 genome editing holds tremendous promises for treating genetic diseases, we are still at the early stage of our understanding of the technology. A lot of diseases are dependent on complex interactions between different genes and will require careful studies to assess where to edit the genome. Also, Cas9 cutting at sites with some sequence similarity to the one specified by its RNA guide, known as Cas9 off-targets, has been documented (Hsu et al., 2013). These can lead to unwanted mutations, and therefore careful selection of RNA guides, with no or minimal homology to nontarget sites, needs to be performed to avoid this problem. Finally, one of the main difficulties of Cas9 genome editing is its accurate delivery to specific organs and cells within the human body, which remains a bottleneck to reach the full potential of this technology.

The Avery–MacLeod–McCarty experiment was the start of an incredible journey to understand how to read, interpret, and edit genetic information. We have reached a stage in which we now need to decide what are the best uses of the knowledge accumulated since their fundamental discovery.

In 2019, against the recommendations of all experts, two babies were born with engineered mutations in their CCR5 receptor (Cyranoski, 2019). This regrettable episode highlights the importance of a careful discussion about the ethics of gene editing, especially of germ cells or embryos. Gene therapy to cure patients, on the other hand, has tremendous potential to change medicine. More than 70 yr ago, Avery, MacLeod, and McCarty triggered a revolution in the biological sciences; it is exciting to wonder where it will lead us in the next 70 yr.

References

Avery, O.T., et al. 1944. *J. Exp. Med.* https://doi.org/10.1084/jem.79.2.137

Barrangou, R., et al. 2007. *Science.* https://doi.org/10.1126/science.1138140

Bertani, G., and J.J. Weigle. 1953. *J. Bacteriol.* https://doi.org/10.1128/JB.65.2.113-121.1953

Bolotin, A., et al. 2005. *Microbiology (Reading).* https://doi.org/10.1099/mic.0.28048-0

Ceccaldi, R., et al. 2016. *Trends Cell Biol.* https://doi.org/10.1016/j.tcb.2015.07.009

Cohen, S.N., et al. 1973. *Proc. Natl. Acad. Sci. USA.* https://doi.org/10.1073/pnas.70.11.3240

Cong, L., et al. 2013. *Science.* https://doi.org/10.1126/science.1231143

Cyranoski, D. 2019. *Nature.* https://doi.org/10.1038/d41586-019-00673-1

Garneau, J.E., et al. 2010. *Nature.* https://doi.org/10.1038/nature09523

Gasiunas, G., et al. 2012. *Proc. Natl. Acad. Sci. USA.* https://doi.org/10.1073/pnas.1208507109

Goeddel, D.V., et al. 1979. *Proc. Natl. Acad. Sci. USA.* https://doi.org/10.1073/pnas.76.1.106

Griffith, F. 1928. *J. Hyg. (Lond.).* https://doi.org/10.1017/S0022172400031879

Hershey, A.D., and M. Chase. 1952. *J. Gen. Physiol.* https://doi.org/10.1085/jgp.36.1.39

Hsu, P.D., et al. 2013. *Nat. Biotechnol.* https://doi.org/10.1038/nbt.2647

Jackson, D.A., et al. 1972. *Proc. Natl. Acad. Sci. USA.* https://doi.org/10.1073/pnas.69.10.2904

Jinek, M., et al. 2012. *Science.* https://doi.org/10.1126/science.1225829

Lander, E.S., et al. International Human Genome Sequencing Consortium. 2001. *Nature.* https://doi.org/10.1038/35057062

Ledford, H. 2019. *Nature.* https://doi.org/10.1038/d41586-019-03698-8

Ledford, H. 2020. *Nature.* https://doi.org/10.1038/d41586-020-00655-8

Loenen, W.A., et al. 2014. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkt990

Luria, S.E., and M.L. Human. 1952. *J. Bacteriol.* https://doi.org/10.1128/JB.64.4.557-569.1952

Mali, P., et al. 2013. *Science.* https://doi.org/10.1126/science.1232033

Marraffini, L.A. 2015. *Nature.* https://doi.org/10.1038/nature15386

Marraffini, L.A., and E.J. Sontheimer. 2008. *Science.* https://doi.org/10.1126/science.1165771

Maxam, A.M., and W. Gilbert. 1977. *Proc. Natl. Acad. Sci. USA.* https://doi.org/10.1073/pnas.74.2.560

McCarty, M., and O.T. Avery. 1946a. *J. Exp. Med.* https://doi.org/10.1084/jem.83.2.89

McCarty, M., and O.T. Avery. 1946b. *J. Exp. Med.* https://doi.org/10.1084/jem.83.2.97

Meselson, M., and F.W. Stahl. 1958. *Proc. Natl. Acad. Sci. USA.* https://doi.org/10.1073/pnas.44.7.671

Mojica, F.J., et al. 1995. *Mol. Microbiol.* https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x

Mojica, F.J., et al. 2005. *J. Mol. Evol.* https://doi.org/10.1007/s00239-004-0046-3

Nirenberg, M.W., and J.H. Matthaei. 1961. *Proc. Natl. Acad. Sci. USA.* https://doi.org/10.1073/pnas.47.10.1588

Pourcel, C., et al. 2005. *Microbiology (Reading).* https://doi.org/10.1099/mic.0.27437-0

Sanger, F., et al. 1977. *Proc. Natl. Acad. Sci. USA.* https://doi.org/10.1073/pnas.74.12.5463

Sanger, F., et al. 1978. *J. Mol. Biol.* https://doi.org/10.1016/0022-2836(78)90346-7

Shendure, J., et al. 2017. *Nature.* https://doi.org/10.1038/nature24286

Smith, H.O., and K.W. Wilcox. 1970. *J. Mol. Biol.* https://doi.org/10.1016/0022-2836(70)90149-X

Szymanski, M., and J. Barciszewski. 2017. *Biochim. Biophys. Acta, Gen. Subj.* https://doi.org/10.1016/j.bbagen.2017.07.009

Watson, J.D., and F.H. Crick. 1953. *Nature.* https://doi.org/10.1038/171737a0