Journal Club

TECHNOLOGY

GENOMIC OUTBREAK SURVEILLANCE IN RESOURCE-POOR SETTINGS

Viral disease outbreaks are common in Africa, often starting in remote areas at the animal–human interface, with the potential to reach epidemic proportions. For example, the West African Ebola outbreak of 2014–2016 received significant international attention owing to its scale (~28,000 infected, ~11,000 dead) and international cases. Coincidentally, 2014 saw the development of an affordable, highly portable sequencing device that measures the disruption of ion charges caused by the different nucleic acid bases when a DNA strand passes through a nanopore. This palm-sized device was unlike conventional whole-genome sequencing platforms, which were large and non-portable, required a steady power supply and, if moved, needed extensive calibration. A paper by Quick et al., who were part of the swift international response to the Ebola outbreak, demonstrates how leveraging a novel sequencing technology challenged the notion that genomic surveillance could not be carried out in resource-poor field settings.

Quick et al. elegantly describe their steps to surmount all previous barriers to conducting on-site Ebola virus genome sequencing in remote settings using this new mobile nanopore sequencing technology. First, to generate enough material for sequencing from patient samples, they optimized a targeted reverse transcriptase PCR protocol using 38 primers sets to amplify the RNA genome of the Ebola virus and identified a minimal 11-amplicon set to cover 97% of the 19 kb Ebola virus genome.

Second, they deployed a 50 kg sequencing tool kit, including the nanopore device, laptops, supplies and reagents, on a commercial airline to Guinea and set it up in an Ebola treatment unit at the heart of the epidemic. The nanopore was unaffected by the power surges and outages, confirming its field readiness.

Third, a well-validated bioinformatics analysis workflow yielded sequence alignment results, assigned genotypes, identified variants and phylogenetic clusters comparable to Illumina sequencing reads. The few exceptions were due to the masked primer binding domains, gaps in primer coverage and challenges in sequencing homopolymer stretches. The investigators noted that the limitations of nanopore sequencing did not significantly alter the key outcomes.

The study demonstrated that the team could get from patient samples to genomic analysis within 24–48 hours. They analysed 142 Ebola viruses over 6 months but, already within 10 days of analysis, could determine that two distinct lineages of strains — the endemic Guinean GN1 strain and a strain from Sierra Leone SL3 — caused the Guinea outbreak, with evidence of cross-country transmission.

Previous strategies, where patient samples were shipped to

> “…leveraging a novel sequencing technology challenged the notion that genomic surveillance could not be carried out in resource-poor field settings”

Journal Club

GENE REGULATION

FIRST GLIMPSE OF ENHANCERS IN GENE REGULATION

The mammalian genome is populated by thousands of genes, and their precise transcriptional regulation is crucial. Historically, promoters and regulatory sequences present within 200 bp of transcription initiation sites were considered to be the sole regulators of eukaryotic gene transcription. However, this assumption was invalidated in 1981 with the discovery by Banerji et al. of distal regulatory elements — which they called ‘enhancers’ — in a viral genome. They and others subsequently confirmed the existence of enhancers in eukaryotic genomes. Since then, enhancers have been shown to be vital for organismal development and homeostasis, with their impairment causing phenotypic variation and disease, and have become a focus of the scientific community studying gene regulation.

Today, the use of sophisticated molecular biology techniques coupled with high throughput genome sequencing has accelerated the discovery and identification of regulatory sequences based on the biochemical markings on DNA and histones. However, it is fascinating to ponder on how enhancers were identified in an era when the choice of molecular biology tools was substantially more limited than now. The Banerji et al. study used simple recombinant plasmid-based assays, in which the rabbit *β-globin* gene was cloned into a construct containing repeat sequences from the region upstream of the SV40 early gene. Surprisingly, this construct expressed the *β-globin* gene at levels several folds higher than a construct without SV40 sequences. Moreover, through various thoughtful iterations — such as flipping of promoter orientation and altering the distances and relative positions of the promoter and SV40 region — Banerji et al. elucidated the fundamental properties of these distal regulatory elements. They coined the term ‘enhancer’ to describe the *cis*-regulatory sequences that enhanced the expression of a related or unrelated gene from a distance and conceived the idea that enhancers act on promoters in an orientation-, position- and distance-independent manner. Furthermore, they even predicted the tissue-type specificity of enhancers, which was then confirmed by them, and separately by others, in 1983.

> “it is fascinating to ponder on how enhancers were identified in an era when the choice of molecular biology tools was substantially more limited than now”

The most impressive aspect of the paper is the insightful discussion of the potential enhancer mechanisms. They could foresee the widespread existence of enhancers beyond viruses and how they could target their promoters by altering the chromatin architecture or via tethering of loci to transcriptionally permissive locations where RNA polymerase can be recruited on the enhancer itself. These ideas, though fully realized now, were ahead of their time — evidence

other countries for testing, had delayed information required to take appropriate treatment and control measures. The extensive network of collaborations and open genomic data sharing by Quick et al. promoted rapid epidemiological analysis and contributed to public health action.

Although the paper emphasizes the deployed genome sequencing capability, it is worth noting that the computational whole-genome sequence analysis was conducted overseas. The authors cite a lack of internet access as the reason and major limitation to running genomic analyses in the field. However, genomic sequencing capacity is incomplete without deploying analysis capacity. Moreover, the exportation of data to other countries for genomic analysis raises problematic ethical issues regarding the ownership of data and creates a dependency on the international community's interest and goodwill, which does not promote genomic equity.

Many epidemics only affect populations in under-resourced environments, which lack the genomic epidemiology capability to control the outbreaks. The paper by Quick et al. demonstrates how the adoption of a new technology in a pandemic can cause paradigm shifts in how and where genome sequencing is conducted. The SARS-CoV-2 pandemic fueled massive growth in sequencing capability, with more countries beginning to conduct routine genomic epidemiological surveillance, mainly because of the accessibility and ease of use of portable nanopore sequencers. But we should not wait for the next pandemic to achieve full genomic equity, enabled by the end-to-end capability to sequence and analyse genomic data for public health in under-resourced settings!

Lillian Musila
Kenya Medical Research Institute/Walter Reed Project, Nairobi, Kenya.
e-mail: lillian.musila@usamru-k.org

Competing interests
The author declares no competing interests.

ORIGINAL ARTICLE Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016)
RELATED ARTICLE Deamer, D. et al. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016)

for the existence of chromatin architecture was just emerging, and polymerase loading on DNA anywhere other than the promoter had not been considered.

The original findings in this paper have inspired generations of scientists across disciplines to investigate how these elements function in a tissue-type specific manner, how they recognize their target promoters in the dense nuclear space and how active enhancers can be identified from the million enhancer elements within the genome. In hindsight, the distal nature of enhancers redirected focus away from the coding genome, which was the norm then, and triggered concentrated efforts to explore the distinct functions of the non-coding genome. As a result, accelerated technological innovations — such as chromatin conformation capture, high resolution microscopy and advances in high throughput genome sequencing — have enabled enhancer mechanisms to be studied in a genome-wide manner. In a nutshell, it is now undisputed that enhancers populate the genomes of most organisms, from small viral genomes to large genomes such as that of humans, and they act in a tissue-type specific manner by virtue of transcription factor binding. Though linearly distant, enhancer and promoter pairs can interact in the 3D nuclear space within the confines of chromatin domains known as topologically associating domains. Moreover, genetic variation within enhancers is associated with disease.

These regulatory elements have fascinated me since I first read Banerji et al. as a PhD student, and they continue to motivate my current research. While we all are enamoured by the latest technologies, this study reminds me that elegant and simple molecular biology experiments can also elucidate novel mechanisms!

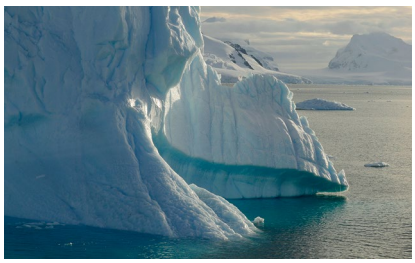
Dimple Notani
Genetics and Development, National Centre for Biological Sciences, Bangalore, India.
e-mail: dnotani@ncbs.res.in

Competing interests
The author declares no competing interests.

ORIGINAL ARTICLE Banerji, J. et al. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981)

METAGENOMICS

Charting the world's microbiomes



Credit: Nature Picture Library / Alamy Stock Photo

Two recent studies report microbial genome and gene catalogues that archive oceanic and glacial genomic and functional diversity at scale and yield insights into their biosynthetic potential.

Paoli et al. reconstructed ~26,000 draft genomes from more than 1,000 seawater samples and integrated these with 10,000 microbial genomes from cultured strains and single cells to build the Ocean Microbiomics Database. By using metagenomic datasets from major oceanographical surveys, such as *Tara* Oceans, and time-series studies, such as the Bermuda-Atlantic Time-series Study (BATS), the authors were able to sample global microbial communities from 215 different sites, across ocean basins, depth layers and time. The team identified 2,700 new microbial species, then mined genomes for biosynthetic gene clusters (BGCs), uncovering 40,000 candidate BGCs. Focusing on one newly identified bacterial family with high BGC content and diversity, named *Candidatus* Eudoremicrobiaceae, the authors characterized two predicted pathways to discover new biosynthetic enzymes and natural products.

In the second study, Liu et al. present the Tibetan Glacier Genome and Gene (TG2G) catalogue, which includes 3,241 genomes from cultured microorganisms as well as newly assembled genomes, spanning 30 phyla and representing 968 bacterial and archaeal species. The samples were sourced from 21 Tibetan glaciers, covering diverse habitats, including snow, ice and cryoconite (a mix of mineral particles and biological material found on glaciers). The team predicted protein functions of microbiomes in different glacial habitats and also identified 15,954 putative BGCs and potential virulence factors.

> “a wealth of data to mine ocean and glacier microbial diversity”

Together, these resources will facilitate the comparison of global microbiomes and provide a wealth of data to mine ocean and glacier microbial diversity, supporting future research and bioprospecting for natural products.

Linda Koch

ORIGINAL ARTICLE Paoli, L. et al. Biosynthetic potential of the global ocean microbiome. *Nature* https://doi.org/10.1038/s41586-022-04862-3 (2022) | Liu, Y. et al. A genome and gene catalog of glacier microbiomes. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01367-2 (2022)
RELATED ARTICLE Medema, M. H. et al. Mining genomes to illuminate the specialized chemistry of life. *Nat. Rev. Genet.* **22**, 553–571 (2021)