This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/ICCMA59762.2023.10374871, 2023 11th International Conference on Control, Mechatronics and Automation (ICCMA)

# Industrial Genomics: A Novel Approach to System Behaviour Discovery

1st Morad Danishvar
*College of Engineering, Design and Physical Sciences*
*Brunel University*
London, UK
Morad.Danishvar@Brunel.ac.uk

2nd Sebelan Danishvar
*College of Engineering, Design and Physical Sciences*
*Brunel University*
London, UK
Sebelan.Danishvar@brunel.ac.uk

3rd Alireza Mousavi
*College of Engineering, Design and Physical Sciences*
*Brunel University*
London, UK
Ali.Mousavi@brunel.ac.uk

***Abstract*— This paper explores a deeper discovery of the concept of industrial genomics which proposes a technique for registering and relating events causing an observable and definable system state and its transfer to another observable state. These industrial genomes are information quanta captured through the digital process to align and represent a chain of activities or processes. They outline the cause-and-effect relationships between events, forming patterns or pathways that ultimately lead to specific outcomes, such as the presence of defects in a product or a machine breakdown. Constructing industrial genomics necessitates understanding the observed or latent parameters of the system's state and how it changes over discrete time intervals.**

**The concept of the proposed industrial genomes, when applied to manufacturing processes, provides a systematic and holistic approach to process optimization, predictive maintenance, and quality control. It has the potential to transform traditional manufacturing processes into smart, efficient, and reliable systems. It could be categorised as a unique method for machine learning.**

**Keywords- Real-time Event Sequencing, Industrial Genomics, DNA sequencing, optimisation, machine learning**

## I. INTRODUCTION

In recent decades, significant progress has been made in the fields of integrated sensors, instrumentation, signal processing algorithms, and internet technology infrastructure culminating in the emergence of 'smart factories' aligned with the 'Industry 4.0' paradigm. Current estimates suggest that between 26 and 50 billion devices are now connected to the Internet and this means the creation of a huge online amount of data [1]. This paradigm instigates today's increasing large-scale developments in online processes and/or machinery health monitoring.

Predictive Maintenance (PdM) and Zero-defect Manufacturing (ZDM) are examples of the most critical components of smart manufacturing and Industry 4.0 [2]. Their benefits include improved efficiency, quality and safety, better compliance and mainly cost reduction due to early detection, and prediction of defects or breakdowns. For example, the PdM strategy for industrial equipment can accurately perceive performance degradation since it was designed to achieve near-zero failures throughout the entire manufacturing process [3].

The Industrial Internet of Things (IIoT) revolution combination with today's advanced data analytics methods enables industries to implement new techniques for the purpose of manufacturing smartifications such as ZDM and PdM. This instigated the authors to present a novel approach for sequential learning of the Genomic of Industrial Process (GIP) in [4] and the Genomic of Machine Breakdown (GMB) in [5]. These studies introduced the theory of process and industrial genomics, inspired by biological genomics, to the field of manufacturing. Herein, we examined a technique for registering and relating events that cause observable and definable system states, and its transfer from one to the other (event-based) and presented a practical implementation in industrial case studies. The presented terminology borrows some of the classical terms and descriptors of DNA (Deoxyribonucleic Acid) sequencing from biology science to the prognosis of the upcoming process defects, breakdowns or component failure. Note that the method is not related to genetic algorithms described in computer science or process optimisation.

In this paper, we further articulate the concept of industrial genomics and its various applications in dynamic industrial environments, where maintaining optimal output quality and/or machine performance is often challenging. The proposal of industrial genomics aims to resolve this challenge, functioning as a language for describing typical processes and connecting occurring events to well-specified outputs.

## II. GENETIC ENGINEERING

Genetic engineering, also known as genetic modification or gene editing, is a biotechnological process that enables scientists to manipulate an organism's genome using modern DNA technology. This manipulation involves inserting, deleting, or modifying DNA to alter genetic traits. Genomics refers to the branch of molecular biology that involves the study of an organism's entire genetic material, or genome. The genome is the complete set of genetic material (DNA in most organisms) that carries the information necessary for

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/ICCMA59762.2023.10374871, 2023 11th International Conference on Control, Mechatronics and Automation (ICCMA)

the growth, development, functioning, and reproduction of that organism [6].

The field of genomics encompasses a wide range of techniques and approaches aimed at understanding the structure, function, evolution, and interactions of genes within a genome. Genomic studies involve not only identifying individual genes but also analysing their organization, regulation, and how they interact with each other and the environment. Genomics involves various sub-disciplines and applications, including Structural, functional, comparative, sequencing, etc. Genomic sequencing determines the complete DNA sequence of an organism's genome [7].

The advent of high-throughput DNA sequencing technologies has revolutionised genomics, enabling researchers to analyse entire genomes more efficiently and at a lower cost. This has led to significant advancements in fields like personalised medicine, evolutionary biology, agriculture, and biotechnology. The DNA labelling process also plays a crucial role in genetic engineering, providing a method for identifying, tracking, and locating specific DNA sequences within a genome. DNA labelling also allows scientists to locate specific DNA sequences within a genome.

Once the DNA has been labelled and the gene of interest is inserted into the host organism's cells, scientists can monitor the genetic modification process. They can track the introduced gene, observe its expression, and evaluate its impact on the organism.

In summary, genomics is a multidisciplinary field that focuses on understanding the complete genetic makeup of organisms, how genes work together, and their roles in various biological processes.

## III. INDUSTRIAL GENOMICS

The concept of genomics, when applies to the field of the manufacturing process, refers to the comprehensive analysis of all parameters of a manufacturing process, their relationships, and their combined impact on the output of the manufacturing system. This concept is introduced by authors in [4]-[5] for manufacturing processes to reduce manufacturing defects and machine breakdowns by identifying defect-causing genes or disruptive sequences. We can rectify or remove these sequences from the process.

In the context of an industrial process sequence, the terms 'Gene', 'Genomic', and 'DNA' can be defined as follows [6]:

- Gene: A gene is a specific sequence of DNA that contains the instructions for building and maintaining a particular functional unit within an organism. In an industrial process sequence, genes are of interest because they can code for process modification/change that contributes to the desired production outcome. Manipulating genes can lead to enhanced product yields or qualities in industrial processes.
- Genomic: 'Genomic' pertains to the entire genetic material of an organism, including all its genes and non-coding sequences. Genomic analysis involves studying the structure, function, and organisation of genes within an organism's genome. In an industrial context, genomic analysis might be used to identify genes responsible for certain production traits or defects or machine breakdown.
- DNA: DNA is a molecule that carries the genetic instructions for the development, functioning, growth, and reproduction of all living organisms. It consists of a double-stranded helical structure, with each strand made up of a sequence of nucleotides. These nucleotides are represented by the letters A, T, C, and G, which stand for adenine, thymine, cytosine, and guanine, respectively. In the context of an industrial process, the term 'DNA labelling' doesn't refer to labelling biological DNA, but it's a metaphorical application of the principles of genetic sequencing to the complex 'genetic code' of industrial processes. It is essential because it encodes the information that determines a production's traits, and it can be manipulated through engineering techniques to optimize production processes.

The underlying principle behind the DNA labelling analogy in the industrial process is to treat each event or state within a process as a 'DNA'. Just as a 'gene' consists of a sequence of DNA that express particular traits, an industrial process comprises a sequence of events or states that lead to a specific output or outcome. By tracking and tracing these process genes, we can accurately predict system behaviour, identify potential errors, and suggest remedial action.

Once these 'genes' have been identified, they can be labelled. Similar to how DNA labelling assigns unique identifiers to different DNA fragments, each unique event or state within an industrial process is assigned a unique 'label' or identifier. These events could be anything from a change in the machine's state, a shift in environmental parameters, or the execution of a particular step in the manufacturing process. This label could take many forms, depending on the nature of the process and the information to be tracked. It could be as simple as a numerical or alphabetical code, or it could be more complex, including information about time, location, resources, etc.

Once each event or state is labelled, they are sequenced based on their occurrence, akin to how the order of DNA bases in a gene determines its function. For example, in a manufacturing process, the sequence could be the order of operations needed to assemble a product. Creating a process genomic sequence requires detailed knowledge of the system's observed or latent parameters, as well as an understanding of how these states change over specific time

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/ICCMA59762.2023.10374871, 2023 11th International Conference on Control, Mechatronics and Automation (ICCMA)

intervals. This step involves monitoring the system, tracking events, and clustering them into discrete units of information, or process genes which can be named parameter identification and discretization [8]-[9].

The labelled sequence of events forms and constructs 'industrial genomes' that reveals patterns and relationships between different events and provides a detailed representation of the system's operation. By identifying, labelling, and tracking these industrial process genomes, This genome can then be analysed and compared to ideal or 'healthy' genomes. Any deviations from the expected sequence can be flagged as potential problems, allowing for early intervention and rectification. Conversely, patterns that consistently lead to positive outcomes can be identified and reinforced. This allows engineers to follow the 'health' of the process in a manner similar to how geneticists might study a DNA strand. Figures 1 and 2 show examples of gene labelling, sequencing and gene prediction.

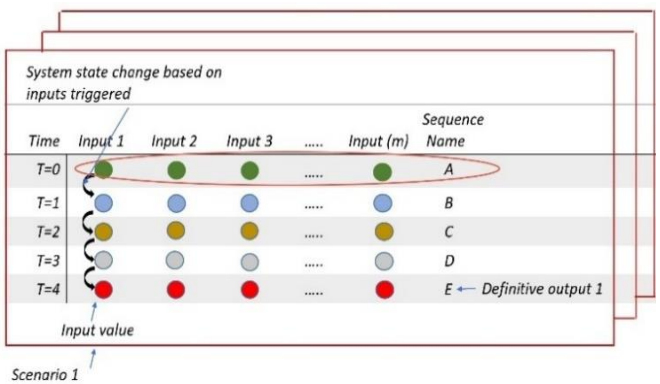


Fig. 1. An example of Gene labelling and sequencing (Borrowed from [4])

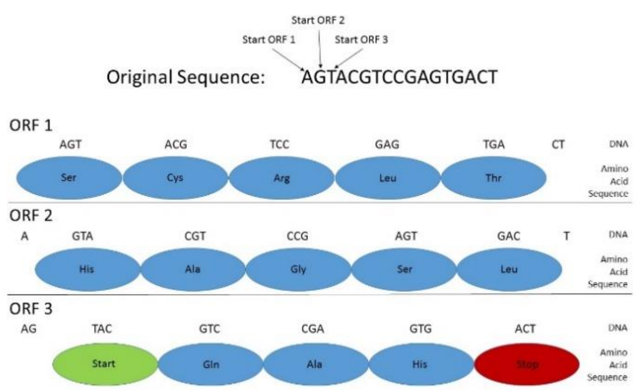


Fig. 2. An example of Gene prediction (Borrowed from [10])

Thus, the principles of industrial genomics offer a powerful tool for accurately predicting system behaviour, enhancing process control, anomaly detection and prediction, predictive maintenance, identifying potential errors, and suggesting remedial action. This can lead to significant improvements in operational efficiency, product quality, and system reliability in sectors such as manufacturing, logistics, energy, and more. It signifies a shift from reactive to proactive process management, enabling organizations to anticipate and address issues before they escalate. In the following, some of the applications of industrial genomics will be elaborated.

- Anomaly Detection: Any deviation from the registered sequence can be easily detected, serving as an indication of potential issues. For instance, if an event occurs out of sequence, it might suggest a machine malfunction or operational error. This is analogous to spotting a mutation in a DNA sequence, which could indicate a genetic disorder.

  For example, if a certain state 'B' in the sequence is always preceded by state 'A' under normal conditions, but suddenly occurs without 'A', the system can flag this as an anomaly. This early warning gives operators a chance to inspect the machinery, identify the cause of the anomaly, and carry out necessary maintenance tasks before a major failure or breakdown occurs.

- Root Cause Analysis: In case of a defect or an anomaly in the manufacturing process, this gene-centric approach can help identify the specific genes (parameters) that caused the deviation. This could help in performing an effective root cause analysis and taking corrective action in a timely manner.

- Optimization of Manufacturing Processes: By defining different combinations of parameters (genes) that influence the manufacturing process, one can determine the optimal state of the system for maximum output and quality. This gene-centric approach can help identify the optimum operating conditions for the process, thereby enhancing productivity and reducing waste.

- Digital Twins (DTs): The event sequence learning approach contributes to the development of digital twins—virtual replicas of physical systems. By providing accurate, real-time operational data, the 'genomic' sequence can keep the digital twin updated, allowing for efficient remote monitoring, simulation, and control.

  In recent years, DT architectures have been equipped with a cognition layer, in addition to access/communication and data analytics layers. The cognitive layer is enabled by the rapid advancement of information, the IIoT, semantic and AI technologies, Big Data Analytics and machine learning methods, ontology engineering and so on. The Cognitive Digital Twin (CDT) is one of the applications of the proposed learning techniques towards higher-level automation and intelligence capabilities and full lifecycle representation of complex systems. The event sequence learning capability, which means transforming the experience of the physical twin into reusable knowledge, besides attributes like attention and memory, is one of the fundamental aspects of cognitive capability [11]. This solution will translate the observed data into objective functions in the form of the Genome of the process. The Decision Support System (DSS), the interpreter of process Genomics, will highlight new state-causing genes and the sequence of events that may result in the development of the event (immediate or further downstream). The gene pool created by real-time gene recognition throughout the learning process allows the nature of the physical process to be transformed into reusable and referenceable knowledge. Genomics sequence can be a powerful tool for interpretation and demonstration of the knowledge graphs in CDT and it will become a major source of information that is required for higher-level decision-making and autonomous reaction abilities within the ecosystem of cognitive digital twins.

- Predictive Maintenance: By studying the sequences of events leading up to equipment failures (akin to genetic predispositions to disease), engineers can predict future breakdowns and conduct preventative maintenance. Through monitoring sensor signals and tracking state changes at defined time intervals, a genomic chain is formed. This 'breakdown genome' can predict potential machine breakdowns and component failures, informing preventive measures.

- Machine Learning and AI: By combining the concept of process genomics with machine learning and AI techniques, manufacturing processes can be made smarter. For instance, self-learning algorithms can be designed to learn from the process genome and predict future states of the manufacturing process. This can help in not only optimizing the process but also in proactive decision-making to prevent machine breakdowns.

The efficacy of this approach lies in its ability to learn and adapt. As more data is collected, the system's predictive capabilities can be refined and optimized. Over time, it will be better equipped to predict the occurrence of a failure, even in complex industrial processes with numerous variables.

## IV. CONCLUSION

The principle of industrial genomics provides an innovative technique for understanding, monitoring, and improving industrial processes. The analogy between the information carried in process sequences allows for a nuanced, predictable understanding of a system's behaviour. By expressing the discrete events of system state transfer as a chain of process genes, we create a detailed map of causal links that lead to specific outputs.

The benefits of this approach are twofold. Firstly, it provides an organized and structured view of the process, simplifying monitoring and control. Secondly, by identifying the sequence of events leading to particular outcomes (e.g., a defect in the product), we can locate the exact point in the process where issues arise, allowing for efficient solutions. By analysing these sequences over time and across multiple processes, we can predict future performance or perform preventive maintenance, reducing the risk of breakdowns and improving overall efficiency.

In summary, the concept of genomics, when applied to manufacturing processes, provides a systematic and holistic approach to process optimization, predictive maintenance, and quality control. It has the potential to transform traditional manufacturing processes into smart, efficient, and reliable systems

## REFERENCES

[1] Kumar, A., Chinnam, R. B., & Tseng, F. (2019). An HMM and polynomial regression-based approach for remaining useful life and health state estimation of cutting tools. Computers & Industrial Engineering, 128, 1008-1014.

[2] Calabrese, M., Cimmino, M., Manfrin, M., Fiume, F., Kapetis, D., Mengoni, M., ... & Toscano, G. (2019, August). An Event-Based Machine Learning Framework for Predictive Maintenance in Industry 4.0. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference 59292 (Vol. 59292, p. V009T12A037). American Society of Mechanical Engineers.

[3] Zhang, W., Yang, D., & Wang, H. (2019). Data-driven methods for predictive maintenance of industrial equipment: A survey. IEEE Systems Journal, 13(3), 2213-2227.

[4] Danishvar M, Mousavi A, Danishvar S. The Genomics of Industrial Process Through the Qualia of Markovian Behavior. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2022 Feb 21;52(11):7173-84.

[5] Danishvar M, Angadi VC, Mousavi A. A PdM framework through the event-based genomics of machine breakdown. In2020 Asia-Pacific International Symposium on Advanced Reliability and Maintenance Modeling (APARM) 2020 Aug 20 (pp. 1-6). IEEE.

[6] Nicholl, Desmond ST. An introduction to genetic engineering. Cambridge University Press, 2023.

[7] Rosenberg, Alex. Reduction and Mechanism. Cambridge University Press, 2020.

[8] Gonçalves M, Sousa P, Mendes J, Danishvar M, Mousavi A. Real-time event-driven learning in highly volatile systems: A case for embedded machine learning for scada systems. Ieee Access. 2022 May 9;10:50794-806.

[9] Danishvar, M., Mousavi, A., & Broomhead, P. (2018). EventiC: A real-time unbiased event-based learning technique for complex systems. IEEE Transactions on Systems, Man, and Cybernetics: Systems, (99), 1-14.

[10] Pandakekok, Gene Prediction.3 December 2016. [Online]. Available https://commons.wikimedia.org/wiki/File:Gene_Prediction.png#filehistory.

[11] Al Faruque, Mohammad Abdullah, Deepan Muthirayan, ShihYuan Yu, and Pramod P Khargonekar. 2021. "Cognitive Digital Twin for Manufacturing Systems." In 2021 Design, Automation & Test in Europe Conference & Exhibition (DATE), 440–445. IEEE.