IDT BIOLOGIKA

# GMP-VALIDATED ADVENTITIOUS VIRUS TESTING BY NEXT GENERATION SEQUENCING

Dr. Hon Q. Tran, Stefan Borutzki, Dr. Izabela Fabianska, Benjamin Richter,
Heiko Julius, Christoph Alarich, Guido Lippert, Guido Schulze,
Stefan Schrapel, Dr. Dietmar Mayer

**CONTENT**

Benefit of reading this text ..... 3

How it all started ..... 3

It is all about regulations ..... 4

US Code of Federal Regulations (CFR) ..... 4

International Council for Harmonisation (ICH) ..... 4

Regulatory science for Adventitious Virus Testing (AVT) ..... 5

NGS for AVT recognized by regulatory authorities ..... 6

Challenges of NGS ..... 6

Bioinformatics ..... 8

Computerized System Validation is the icing on the cake ..... 9

NGS-based AVT as a multidisciplinary approach ..... 11

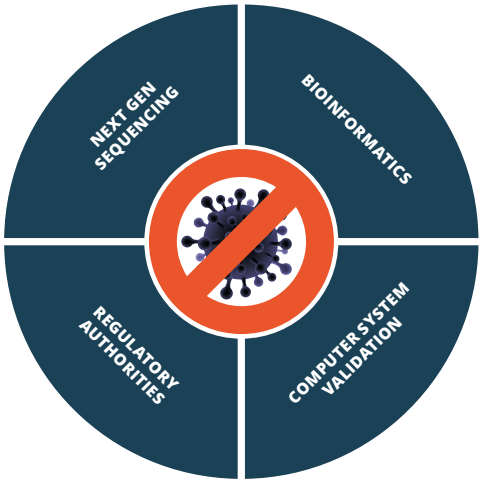
Authors ..... 12

About IDT Biologika ..... 13

References ..... 14

## BENEFIT OF READING THIS TEXT

Safety, quality, and efficacy of biologics are the basis for approving and authorizing through state regulatory bodies. Safety is ensured among other activities by quality controls during the manufacturing process. This also includes testing for unwanted agents which pose a threat to human health. This safety control step is called ***Adventitious Agent Testing***. Here, we specifically deal with viral contamination and name this topic ***Adventitious Virus Testing*** (AVT). This white paper provides you with required knowledge related to the involved regulatory authorities and scientific technologies and disciplines: Next Generation Sequencing (NGS), Bioinformatics and Computer System Validation (CSV). Before starting, let us take a look at the history of AVT, which was not a matter of course.




## HOW IT ALL STARTED

The death of 22 children marks the beginning of regulating the pharmaceutical industry through the U.S. Government. In 1901, a 6-year-old girl in St. Louis (Missouri) was treated with anti-toxin against diphtheria. Likewise, her two siblings as a precautionary measure. All three died, due not to ***Corynebacterium diphtheriae***, but to tetanus [1]. One week later the thirteenth death was reported [2]. This tragedy was due to a lack of organization and serum preparation guidelines within the St. Louis City Health Department. During this period before the regulation, it was commonplace for local health authorities to produce vaccines. The anti-toxin given to the 13 children was sourced from a horse days before and after its infection with ***Clostridium tetani***. A poorhouse was used as a stable for the horse. An assistant bacteriologist was ordered to process the serum at the poorhouse, assisted by the stableman, and release it for use, aware that the horse was bled for serum after infection and the production batches around in those days were not tested on guinea pigs because appropriate ones were exhausted for tests the previous summer. The bacteriologist instructed a janitor to label the anti-toxin bottles with dates of the drawings from memory [3]. Later that same year, tetanus once again led to a number of deaths in Camden (New Jersey), including nine children, following medication with tainted smallpox vaccine [4,5]. These incidents in St. Louis and Camden 120 years ago are considered to mark the starting point of the very first law regulating the production of biologics.

REPUBLIC.

THREE MORE DEATHS RESULT FROM EFFECTS OF INFECTED ANTITOXIN.

Chronicling America: Historic American Newspapers. Lib. of Congress *https://chroniclingamerica.loc.gov/lccn/sn84020274/1901-11-02/ed-1/seq-1/*

President Theodore Roosevelt signed "The Act of July 1, 1902", known since the 1930s as the ***Biologics Control Act***. There is some disagreement as to whether the tragic incidents or the initiative of large biologics manufacturers primarily gave rise to the act [6]. It forced producers to get annual licenses from the ***Laboratory of Hygiene of the Public Health and Marine Hospital Service*** in order to manufacture and sell vaccines, serum and anti-toxins, including inspections of their facilities. In 1906, the ***Federal Food and Drugs Act*** became law, prohibiting the trading of mislabeled and adulterated food and drugs. The regulatory supervision was placed in the hands of the ***Bureau of Chemistry,*** renamed ***Food And Drug Administration*** (FDA) in 1930. In 2006, the FDA celebrated its 100th anniversary. Also in 1930, the ***Laboratory of Hygiene of the Public Health and***

*Marine Hospital Service* was renamed the *National Institute of Health* (NIH), becoming the plural *Institutes* to emphasize its diverse research capacities, one of which was the regulatory control of biologics. In 1972, the latter responsibility was transferred to an FDA division which has gone by the name *Center for Biologics and Evaluation* (CBER) since 1987. Thus, in 2022, the CBER, born with the act of 1902, can look back on 120 years of history [7–9].

## IT IS ALL ABOUT REGULATIONS

Ensuring viral safety during manufacturing of biologics also implies working in a regulated environment. Unlike working in an R&D field, there are numerous rules to be followed and the working procedures and manufacturing facilities are regularly audited by state authorities. This section offers a glimpse of the involved guidelines and regulatory bodies.

### *US CODE OF FEDERAL REGULATIONS (CFR)*

The US Federal Register contains laws signed by the President after legislation passed the Congress. The laws are codified and published as the Code of Federal Regulations (CFR), which comprises about 50 titles. The title 21 is reserved for FDA and exhibits about 1,499 parts [10]. Manufacturers of biologics have to demonstrate the 21 CFR compliance of their work. Parts 210 + 211 outline the Current Good Manufacturing Practices (CGMPs) which control the manufacturing processes and facilities aiming at safeguarding pharmaceutical quality. Along with the advancing digitization and automation in laboratories, Part 11 takes a prominent position [11]. It establishes criteria under which electronic records and signatures are equivalent to paper records. Computer systems should be validated and in case their access is limited to authorized individuals the system must use computer-generated, time-stamped audit trails to record date and time of actions which create, modify, or delete electronic records. Legal demands on hardware and software are an extensive subject, see below for details.

### *INTERNATIONAL COUNCIL FOR HARMONISATION (ICH)*

From the 1960s onwards the first pharmaceutical companies began to go global, with more and more nation states legislating and preparing guidelines for medicinal products. The European countries (EC) were the first whose regulatory bodies aligned their laws, as the states became a single market for pharmaceutical products. Key drivers for this thinking were rising R&D costs and the long time to market due to varying regulatory conditions in the target countries. In 1990, the regulatory agencies from the EC, Japan and FDA and industry representatives from the countries mentioned above came together and launched an initiative which has existed as a legal entity named *International Council for Harmonisation* (ICH) since 2015 [12]. As of December 2021, ICH includes regulatory agencies and industrial representatives from all over the world, of which 19 are members and 35 only have observer status with fewer rights and obligations. The Assembly is the central body of ICH with the most rights. Regulatory and industry members commit to implement and support ICH guidelines, respectively. Each guideline is the result of a five-tier process: (1) a technical document signed off by topic leaders (2) document endorsed by the ICH Assembly (3) public consultation by regulatory members (4) the harmonized guideline adopted by regulatory members of the ICH Assembly (5) implementation by the ICH regulatory members [13]. The last step is optional, depending on the local regulatory authorities. Implementation can occur years after step 4, sometimes never. The guidelines are divided into the categories Quality, Safety, Efficacy and Multidisciplinary topics. The Quality guidelines are subcategorized into Q1 to Q14, which deal with pharmaceutical product purity and stability encompassing, among others, analytical testing, harmonization of major pharmacopoeias, Good Manufacturing Practice (GMP) [14].

The ICH Q5A guideline is the primary instruction for the viral safety of pharmaceutical products originating from animal and human cell cultures: *Viral Safety Evaluation of Biotechnology Products Derived from Cell Lines of Human or Animal Origin*. ICH Q5A passed step 4 in 1997 and was revised two years later [15]. It was implemented by the FDA in 1998, the revised guideline by the EC in 1999. ICH Q5A-(R1) suggests a viral safety program based on a tripartite approach: (1) ensuring raw material such as cell lines and media components is free of adventitious viruses through testing, (2) the production process should have the capacity to remove and inactivate viral contaminants, (3) in-process and final product testing. A number of classical in vitro and vivo testing assays including electron microscopy studies are listed as they were standards more than 20 years ago. This is why a second revision is in preparation.

ICH Q5A(R2) is advocated by ICH management due to four advances since the first revision: (1) new classes of biotechnology products, (2) additional validation approaches for virus clearance, (3) virus clearance validation and risk mitigation strategies for advanced manufacturing like continuous manufacturing processes, (4) new adventitious virus testing methods [16]. The latter issue involves nucleic acid-based assays like Polymerase Chain Reaction (PCR) and NGS. ICH Q5A(R2) envisages the replacement and/or supplementation of existing assays, if possible, and the elaboration of general principles for the continuous inclusion of new assays. According to the business plan, the second revision is scheduled to complete step 1 + 2 in June 2021 and step 3 + 4 in November 2022 [17]. As of March 2022, the ICH website still displays step 1 as current status. The workgroup for this topic is chaired by Johannes Blumel, PhD, (EC, Europe), who also works as workgroup expert together with, among others, Arifa S. Khan, PhD, (CBER, FDA) and Laurent Mallet, PhD, (European Directorate for the Quality of Medicines and Healthcare, EDQM).

## REGULATORY SCIENCE FOR ADVENTITIOUS VIRUS TESTING (AVT)

From the beginning, science laid the foundations of the FDA. Its father and midwife was Harvey Washington Wiley (1844–1930), an analytical chemist and chief of the ***Bureau of Chemistry,*** who developed analytical methods to test foods and drugs for their purity. He saw prevention of food and drug adulteration as his mission in life. Outside of the laboratory he allocated funds for his experiments, gave lectures, and involved manufacturers and groups to generate interest in his mission. His dedication culminated in Food and Drugs Act in 1906, the majority of which he wrote himself [18–20].

Following this tradition of science-based regulation, the FDA, including CBER, continuously developed methods and animal models to test and evaluate biologics. For this scientific work the FDA adopted the term ***Regulatory Science,*** which is ***the science of developing new tools, standards, and approaches to assess the safety, efficacy, quality and performance of all FDA-regulated products*** [21]. CBER uses a staff hiring system named researcher-reviewer model which allows staff members to conduct research alongside their regulatory work. Part of their working time is reserved for research activities. Today, regulatory science is a key concept among the regulatory authorities worldwide. In 2020, the European Medicines Agency (EMA) disclosed its ***Regulatory Science to 2025*** strategy and identified knowledge gaps in about 100 regulatory science topics, one of which is the application of Next Generation Sequencing (NGS) for AVT [22,23].

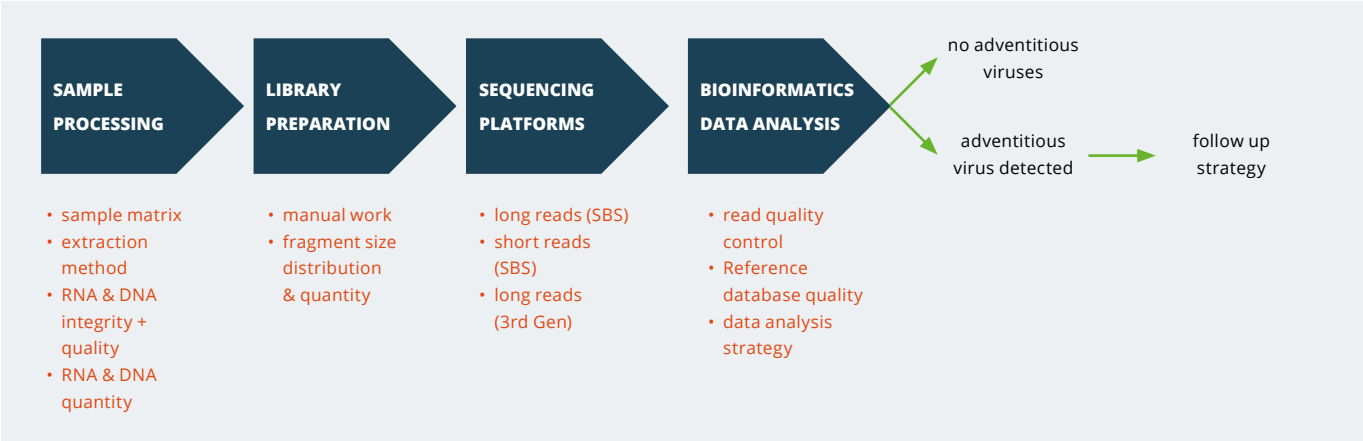
NGS capabilities for broad virus detection can be seen in the regulatory authorities worldwide. The FDA including CBER devotes a lot of resources to advance NGS for applications in biologics. This prioritization greatly illustrates the application of the researcher-reviewer model. Externally, particular efforts are directed to AVDTIG [24] and collaboration with IABS [25]. Arifa S. Khan, PhD, a Senior Investigator from the ***Office of Vaccines Research and Review*** (OVRR) within CBER, is leading these efforts. The Advanced Virus Detection Technology Interest Group (AVDTIG) is an informal forum that brings together more than 180 participants from regulatory and other government agencies, industry and academia to discuss and exchange their ideas in five subgroups: (1) sample selection, preparation, processing, (2) virus standards and reference materials, (3) virus reference database, (4) bioinformatics pipeline analysis, (5) follow-up strategies for result validation. The other effort involves the ***International Alliance for Biological Standardization*** (IABS), founded in 1955 as a non-profit alliance for improving the quality and regulation of biologics. Since 2017 two conferences have been held with a focus on applications of NGS for AVT in biologics [26,27]. These meetings are good for contemplating the perspective of regulatory and health authorities on the one side and industry on the other with regard to the benefits of NGS for adventitious virus detection.

## NGS FOR AVT RECOGNIZED BY REGULATORY AUTHORITIES

Next-Generation Sequencing is a game changer for the detection of unwanted viruses in pharmaceutical products. For the very first time, a vaccine already on the market was shown by NGS to contain adventitious virus. In 2010, an academic group discovered porcine circovirus-1 (PCV1) in Rotarix (GlaxoSmithKline, GSK) by investigating NGS not routinely adopted so far for adventitious virus detection [28]. Rotarix is an oral live-attenuated rotavirus vaccine against acute gastroenteritis in infants and young children and was licensed at this time in more than 123 countries worldwide. GSK confirmed the results and detected PCV1 DNA and viral particles at all manufacturing stages [29]. This finding was a wakeup call for both the regulatory authorities and industry, signaling that there was a gap in the current testing using conventional methods, and that NGS had the power to fill this gap. At the second IABS conference on NGS for AVT FDA, EMA, Paul-Ehrlich-Institute and World Health Organization (WHO) [27] were represented. The latter is not a regulatory agency; however, it provides regulatory standards as a basis for its member states. All agree on the benefits of NGS for securing biosafety of vaccines in human and animal health and see the potential to follow the EMA 3Rs (replacement, reduction, refinement) initiative for animal testing [30]. The same standards are applicable to NGS as to other assays. For their part, the regulatory authorities recognize that they have to gain experience in relation to data content and format to be submitted as well as how to use the data for decision making. The manufacturing industry, Contract Research Organization (CRO) and service providers are expected to address the NGS challenges. These concern the NGS assay itself, the bioinformatics platform, and a follow-up strategy for handling a positive signal out of this approach. The OVRR of CBER is seeing an increasing number of requests from industry to use NGS for adventitious virus detection, not least because of the COVID-19 pandemic, and accepts NGS data on a case-by-case basis [31]. Furthermore, OVRR encourages sponsors to initiate discussion about their intention to use NGS for adventitious virus detection.

## CHALLENGES OF NGS

In a scientific experiment we have a result which, like a school test score, is dependent on variables like sleep duration the night before. Sleep duration is an independent variable, the test score is a dependent one. The Next-Generation Sequencing approach for AVT delivers a result: adventitious virus present or absent. The regulatory authorities expect that this result is independent of the preceding NGS steps and bioinformatics data analysis. There are a number of independent potentially bias-producing variables to be considered.




The NGS-based approach can be divided into four phases, as shown in the figure above. The probe to be tested for adventitious viruses is processed, the extracted nucleotides are prepared for sequencing. The sequencer outputs data which are analyzed by bioinformatic algorithms. Great variability in the composition of viruses complicates the steps prior to sequencing. Viruses can have RNA or DNA, each single-stranded, double-stranded, with or without envelope and with different susceptibility to treatment processes. A sufficient amount of nucleic acid is needed for library preparation. Therefore, its extraction is critical to catch all potential adventitious viruses. The sample matrix can vary, depending on what within the vaccine manufacturing process has to be tested, such as growth media, suspended whole host cells or cell lysate. The physicochemical properties of the viruses in combination with sample matrix becomes further complicated regarding the extraction method [32]. If there are just a few virus particles, their genome could be amplified. Enveloped viruses can be enriched by treating with nuclease [33, 34]. The extracted nucleotides should be checked with respect to their integrity, quality, and quantity. Library preparation might cause variability through hands-on-time to follow sometimes complex protocols to attach adapters and barcodes to the nucleotides [35]. Possible contamination through personnel, DNA extraction kits and other reagents are a permanent cause for the falsifying of results [36]. The quality of the library can be concluded from the size distribution and quantity of the fragments. Further dependencies of the result might stem from the different sequencing platforms in use. The second-generation sequencers (Sequencing-by-Synthesis, SBS) produce short reads like the Illumina devices and longer reads like the 454 Roche system. The advent of the third generation sequencing like the Oxford Nanopore technologies [37] opens door to new ways of data analysis, but also to greater variabilities as well. The quality and length of the reads have an impact on the composition of the bioinformatics data analysis workflow. The latter is addressed below in more detail.

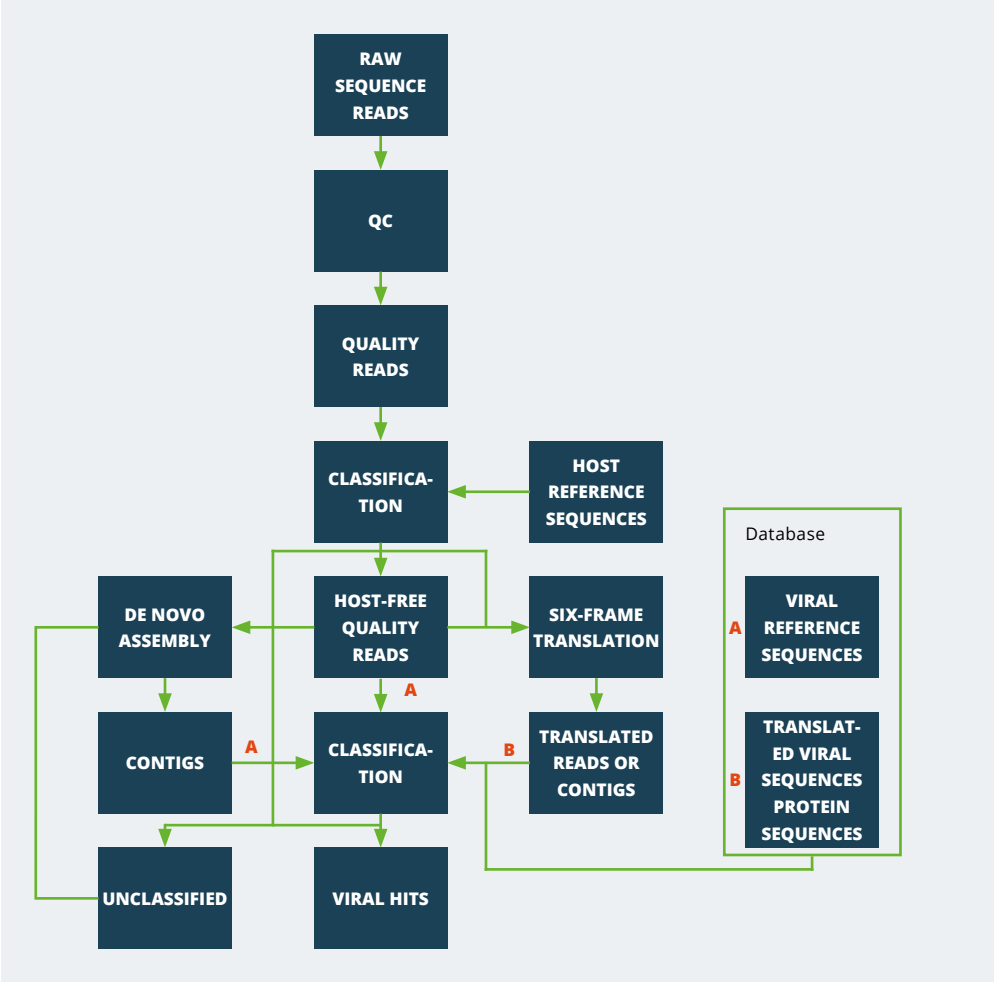
In view of these variables having an effect on virus detection, the regulatory authorities request that the outcome of AVT is independent of the lab and its upstream processes and downstream data analysis of the sequencing step. The road to this objective leads through standardization and validation of the NGS approach. One measure is to establish a panel of reference viruses of different physicochemical properties. Its detection across different labs constitutes the benchmark of the robustness of the technology and reproducibility of the results. In a study involving three labs of AVDTIG a reference panel of four viruses in different matrices was generated [38]. All three labs demonstrated similar sensitivity of virus detection in spite of different sample processing steps, sequencing platforms and bioinformatics workflows. In a larger study, 15 labs were challenged with a panel of 25 viruses [39]. Here, the results differ significantly. Only 6 out of 25 were identified by all participants, the remaining 19 viruses were detected by 4-14 labs. This study shows that standardization of wet lab methods and bioinformatics workflows is urgently required.

Standardization is one of two challenges of NGS applications for AVT. The other one is a follow-up strategy in case of a positive result of detection, a so-called ***hit***. Such a result does not automatically denote a vaccine harmful to human health. It is necessary to check whether the unexpected nucleotides stem from infectious particles. Possible follow-ups include confirmation of full-length viral genome sequences, expression analysis and infectivity assays [40].

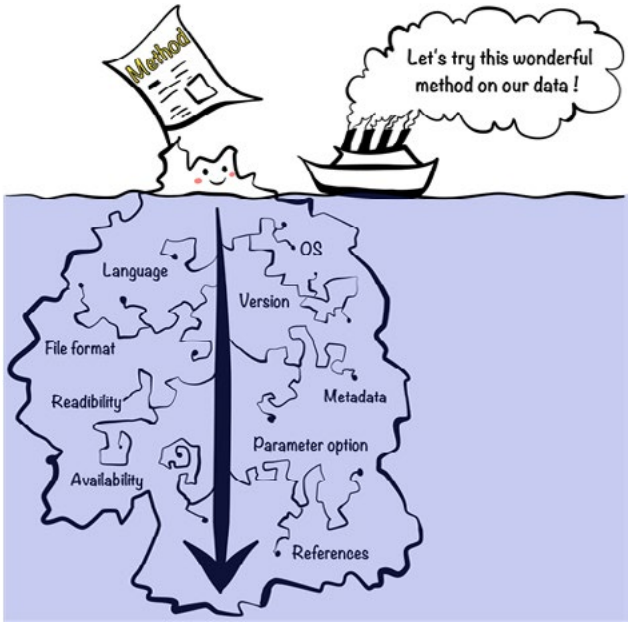
## BIOINFORMATICS

The critical parts of a bioinformatics workflow are the databases and the classification step. The figure below shows the relevant steps found in multiple published workflows [41, 42]. In the first step, quality control (QC), raw sequences with bad metrics are removed through trimming and filtering. Then remaining quality reads are freed from host reads through a classification step. Here, the reads can follow three paths: (1) classification resulting in putative viral hits and unclassified reads, viral references as database, (2) de novo assembly into contigs and classification as first path, (3) six-frame translation and classification, this time translated viral sequences and protein sequences as databases. Unclassified reads from first path can follow the second and third path.

The classification algorithm and the corresponding databases decide whether the reads are classified as host, putative viral hits or remain unclassified. Classification can rely on aligners such as the widely used Bowtie2 [43] or BWA [44]. Reads can also be classified using k-mer based metagenomics tools which also enable a taxonomic classification. One widely used tool is Kraken2 [45]. Most of the published bioinformatics workflow uses databases from the National Center for Biotechnology (NBCI, nr/nt, RefSeq) and/or the Reference Viral Database (RVDB) developed by Arifa Khan at FDA's CBER [46].



The numerous tools, including their even greater number of parameters, lead to a large number of bioinformatics workflows. Standardization is facilitated by testing the published workflows, however, it is impeded by multiple factors, as shown in the figure below.




Source [47]

## COMPUTERIZED SYSTEM VALIDATION IS THE ICING ON THE CAKE

As with the other topics outlined above, this topic demands a white paper of its own. We only briefly touch upon this issue, just enough for you to recognize its importance for the viral safety of pharmaceutical products. NGS- based AVT is simple to undertake: sequence a potentially contaminated probe containing a predefined concentration of a known virus species (spiked-in virus) as control, input these data into a bioinformatics application which is seen working well if the spiked-in virus is found with its expected amount. Adventitious viruses, if any, will be identified. When it comes to human safety of pharmaceutical products, then this procedure is necessary but not sufficient. This is where ***Computerized System Validation*** (CSV) comes into play.

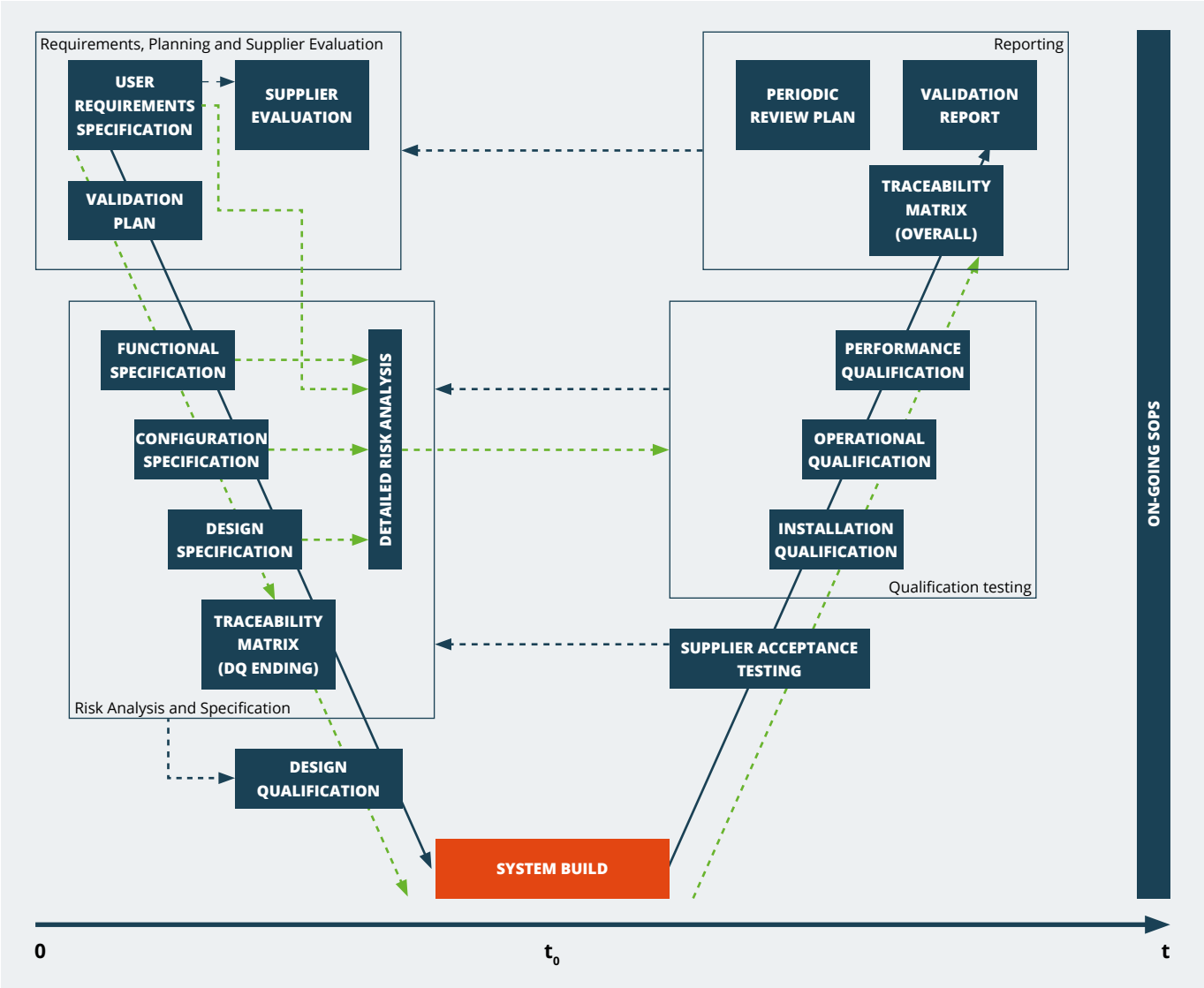
A computerized system ***includes hardware, software, peripheral devices, personnel, and documentation*** [48]. Pharmaceutical manufacturing companies are required to implement a quality system which is fully compliant with EU GMP guidelines and US cGMP requirements. This obligation also applies to the testing of absence of adventitious viruses in biologics. Computerized systems such as computer server and bioinformatics applications involved in AVT have to demonstrate their GxP-compliance. The "x" in GxP is replaced by a single character specific to the corresponding guideline: Good Manufacturing Practice (GMP), Good Laboratory Practice (GLP), and Good Clinical Practices (GCP). Specific to the bioinformatic data analysis example mentioned above, GxP revolves around two facts: (1) who has started the application and when (accountability), (2) each feature of the application can be traced to its user requirement through the intermediate phases testing and design specification (traceability).

Again, it is all about regulation. The US FDA has defined the requirements in 21 CFR Part 11 (see above). The European GMP Guidelines Annex 11 lists terms relevant in the CSV context, including audit trails, risk management and periodic evaluation [49]. The third player is GAMP, which stands for Good Automated Manufacturing Practice, established in 1991 by the International Society for Pharmaceutical Engineering (ISPE) [50]. It is not a state regulatory body, however, it has evolved into a standard in regulated pharmaceutical companies. It defines five software categories. Category 5 includes custom applications, which surely applies to AVT software, as it is not off-the-shelf software. Reference is often made to GAMP 5, which indicates the current version of the guidelines. GAMP is useful because it tells us how to validate computerized systems, whereas 21 CFR Part 11 and EU Annex 11 show what to validate.

Computerized System Validation comprises both qualification of IT infrastructure such as physical computers, virtual machines or storage and validation of software applications. Note the usage of the terms ***Qualification*** and ***Validation***. CSV

passes four phases, which are shown in the following figure.

Hardware qualification includes DQ and IQ, software validation additionally comprises OQ and PQ. The four phases are part of a multi-stage development process which is named *Software Development Life Cycle* (SDLC). As suggested by GAMP 5, one often-used SDLC is the V model, the processes of which are executed in a sequential manner in a V shape. One typical example is shown below.



In the first stage the user requirements are specified, and a decision is made regarding what to validate. DQ is completed after the next stage, in which the application is specified in more detail. After coding, the application is tested against user and regulatory requirements to check it is correctly installed, works fine and delivers the expected results. The software features are checked to see whether they are implemented as designed, indicated by the arrow from the right branch of the V model to the left one. Finally, a validation report is generated stating the validation results. Starting with the user requirements and continuing to the final report, all specifications and testing steps are summarized in a table verifying that all features are tested, implemented as designed and specified as required by the user, and vice versa. This table is called the traceability matrix (TRM). After a defined period, the application is reviewed, and the result is documented in a Periodic Review Report. The cycle is completed. If new requirements are available or modifications needed, the process starts again. The vertical bar on the right indicates that the software validation produces a number of paper documents, including Standard Operating procedure (SOP), Work Instruction, User Manuals, Audit Trail Review and Periodic Review. These documents are required for correct system operation and maintenance. According to the principles of GLP, validation activities are specified as SOPs which outline what is to be done and who is to do it. Work instruction explains how a specific task is to be performed.

## NGS-BASED AVT AS A MULTIDISCIPLINARY APPROACH

Testing for adventitious viruses in regulated industries is a multi-skill discipline involving laboratory work, understanding and application of the NGS technology, knowing relevant bioinformatics algorithms, developing software applications for data analysis, qualifying the required IT infrastructure, and validating the bioinformatics software. All these tasks have to be compliant with the guidelines of regulatory authorities. Now, at the end of this white paper, we hope that you have gained an insight into the exciting and important topic of *Adventitious Virus Testing*.

## REFERENCES


1. Coroner will make an investigation. *The St. Louis Republic 1901, 1 (Nov 7).*
2. Two new cases of tetanus developed. *The t. Louis Republic 1901, 5 (Nov 7).*
3. Serum poisoned, he says. *New-York Tribune 1901, 1 (Dec 11).*
4. The Camden Board of Health takes action because of deaths from lockjaw. *New-York Tribune 1901, 6 (Nov 19).*
5. Another case of tetanus. *New-York Tribune 1901 (Dec 5).*
6. Terry S. Coleman. Early Developments in the Regulation of Biologics. *Food And Drug Law Journal 2016, 544–607.*
7. Linda Bren. The Road to the Biotech Revolution – Highlights of 100 Years of Biologics Regulation (2006).
8. John P. Swann. FDA's Origin. Available at *https://www.fda.gov/about-fda/changes-science-law-and-regulatory-authorities/fdas-origin* (2018).
9. FDA. A brief History of the Center for Drug Evaluation and Research. Available at *https://www.fda.gov/about-fda/fda-history-exhibits/brief-history-center-drug-evaluation-and-research* (2018).
10. CFR. Code of Federal Regulations - Title 21. Available at *https://www.ecfr.gov/current/title-21.*
11. CFR. Code of Federal Regulations - PART 11 - ELECTRONIC RECORDS; ELECTRONIC SIGNATURES. Available at *https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11.*
12. ICH. About International Council for Harmonisation (ICH). Available at *https://www.ich.org/page/history.*
13. ICH. Overview of ICH. Available at *https://admin.ich.org/sites/default/files/2021-12/OverviewOfICH_2021_1202_0.pdf* (2021).
14. ICH. ICH Quality Guidelines. Available at *https://www.ich.org/page/quality-guidelines*.
15. ICH. VIRAL SAFETY EVALUATION OF BIOTECHNOLOGY PRODUCTS DERIVED FROM CELL LINES OF HUMAN OR ANIMAL ORIGIN Q5A(R1). Available at *https://database.ich.org/sites/default/files/Q5A%28R1%29%20Guideline_0.pdf.*
16. ICH. Final Concept Paper Q5A(R2): Viral Safety Evaluation of Biotechnology Products Derived from Cell Lines of Human or Animal Origin. Available at *https://database.ich.org/sites/default/files/Q5A-R2_FinalConceptPaper_2019_1117.pdf.*
17. ICH. Final Business Plan Q5A(R2): Viral Safety Evaluation of Biotechnology Products Derived from Cell Lines of Human or Animal Origin. Available at *https://database.ich.org/sites/default/files/Q5A-R2_FinalBusinessPlan_2019_1118.pdf.*
18. Science History Institue. Harvey Washington Wiley. Available at *https://www.sciencehistory.org/historical-profile/harvey-washington-wiley (2018).*
19. FDA Consumer Magazine. Harvey W. Wiley: Pioneer Consumer Activist. Available at *https://www.fda.gov/about-fda/fda-leadership-1907-today/harvey-w-wiley-pioneer-consumer-activist* (2019).
20. FDA Consumer Magazine. Harvey Washington Wiley. Available at *https://www.fda.gov/about-fda/fda-leadership-1907-today/harvey-wiley* (2020).
21. FDA. Advancing Regulatory Science. Available at *https://www.fda.gov/science-research/science-and-research-special-topics/advancing-regulatory-science* (2021).
22. EMA. Regulatory Science Strategy. Available at *https://www.ema.europa.eu/en/about-us/how-we-work/regulatory-science-strategy#regulatory-science-strategy-to-2025-section.*
23. EMA. EMA launches the Regulatory Science Research Needs initiative. Available at *https://www.ema.europa.eu/en/news/ema-launches-regulatory-science-research-needs-initiative.*
24. Khan, A. S. et al. Advanced Virus Detection Technologies Interest Group (AVDTIG). Efforts on High Throughput Sequencing (HTS) for Virus Detection. PDA journal of pharmaceutical science and technology 70, 591–595; 10.5731/pdajpst.2016.007161 (2016).
25. IABS. International Alliance For Biological Standardization. Available at *https://www.iabs.org/.*
26. Khan, A. S. et al. Report of the international conference on next generation sequencing for adventitious virus detection in biologicals. *Biologicals: journal of the International Association of Biological Standardization 55, 1–16; 10.1016/j.biologicals.2018.08.002* (2018).
27. Khan, A. S. et al. Report of the second international conference on next generation sequencing for adventitious virus detection in biologics for humans and animals. *Biologicals: journal of the International Association of Biological Standardization 67, 94–111; 10.1016/j.biologicals.2020.06.002* (2020).
28. Victoria, J. G. et al. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. *J. Virol. 84, 6033–6040* (2010).
29. Dubin, G. et al. Investigation of a regulatory agency enquiry into potential porcine circovirus type 1 contamination of the human rotavirus vaccine, Rotarix. Approach and outcome. *Human vaccines & immunotherapeutics 9, 2398–2408; 10.4161/hv.25973* (2013).
30. EMA. Guideline on the principles of regulatory acceptance of 3Rs (replacement, reduction, refinement) testing approaches. Available at *https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-principles-regulatory-acceptance-3rs-replacement-reduction-refinement-testing-approaches_en.pdf* (2016).
31. Rebecca Stauffer. CBER Continues to Analyze Next-Gen Sequencing. Available at *https://www.pda.org/pda-letter-portal/home/full-article/cber-continues-to-analyze-next-gen-sequencing*.
32. Sathiamoorthy, S., Malott, R. J., Gisonni-Lex, L. & Ng, S. H. S. Selection and evaluation of an efficient method for the recovery of viral nucleic acids from complex biologicals corrected. *NPJ vaccines 3, 31; 10.1038/s41541-018-0067-3* (2018).
33. Allander, T., Emerson, S. U., Engle, R. E., Purcell, R. H. & Bukh, J. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proceedings of the National Academy of Sciences of the United States of America 98, 11609–11614; 10.1073/pnas.211424698* (2001).
34. Ng, S. H. et al. Current Perspectives on High-Throughput Sequencing (HTS) for Adventitious Virus Detection. Upstream Sample Processing and Library Preparation. *Viruses 10; 10.3390/v10100566* (2018).
35. Hess, J. F. et al. Library preparation for next generation sequencing. A review of automation strategies. *Biotechnology advances 41, 107537; 10.1016/j.biotechadv.2020.107537* (2020).
36. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology 12, 87; 10.1186/s12915-014-0087-z* (2014).
37. Oxford Nanopore. MinION Products. Available at *https://nanoporetech.com/products/minion.*
38. Khan, A. S. et al. A Multicenter Study To Evaluate the Performance of High-Throughput Sequencing for Virus Detection. *mSphere 2; 10.1128/mSphere.00307-17* (2017).

39. Mee, E. T., Preston, M. D., Minor, P. D. & Schepelmann, S. Development of a candidate reference material for adventitious virus detection in vaccine and biologicals manufacturing by deep sequencing. *Vaccine 34, 2035–2043; 10.1016/j.vaccine.2015.12.020* (2016).
40. Charlebois, R. L. et al. Sensitivity and breadth of detection of high-throughput sequencing for adventitious virus detection. *NPJ vaccines 5, 61; 10.1038/s41541-020-0207-4* (2020).
41. Fabiańska, I. et al. LABRADOR-A Computational Workflow for Virus Detection in High-Throughput Sequencing Data. *Viruses 13; 10.3390/v13122541* (2021).
42. Lambert, C. et al. Considerations for Optimization of High-Throughput Sequencing Bioinformatics Pipelines for Virus Detection. *Viruses 10; 10.3390/v10100528* (2018).
43. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods 9, 357–359; 10.1038/nmeth.1923* (2012).
44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England) 25, 1754–1760; 10.1093/bioinformatics/btp324* (2009).
45. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome biology 20, 257; 10.1186/s13059-019-1891-0* (2019).
46. Khan, A. S. Reference Viral Database (RVDB). Available at *https://rvdb.dbi.udel.edu/.*
47. Kim, Y.-M., Poline, J.-B. & Dumas, G. Experimenting with reproducibility. A case study of robustness in bioinformatics. *GigaScience 7; 10.1093/gigascience/giy077* (2018).
48. FDA. Glossary of Computer System Software Development Terminology (8/95). Available at *https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-guides/glossary-computer-system-software-development-terminology-895* (2014).
49. European Union. Good Manufacturing Practice, Medicinal Products for Human and Veterinary Use. Annex 11: Computerised Systems. Available at *https://ec.europa.eu/health/system/files/2016-11/annex11_01-2011_en_0.pdf.*
50. ISPE. GAMP 5 Guide: Compliant GxP Computerized Systems. Available at *https://ispe.org/publications/guidance-documents/gamp-5*.

## AUTHORS

**Dr. Hon Q. Tran**
Expert Bioinformatics

**Dr. Izabela Fabianska**
Bioinformatics Scientist

**Heiko Julius**
Engineer Qualification / Validation

**Guido Lippert**
IT Virtualization Specialist

**Stefan Schrapel**
Senior Expert Computerized Systems

**Stefan Borutzki**
Expert Bioinformatics

**Benjamin Richter**
NGS Lab Scientist

**Christoph Alarich**
Engineer Computerized Systems

**Guido Schulze**
Senior Expert IT Infrastructure

**Dr. Dietmar Mayer**
Director Analytical Development
Corresponding Author

## ABOUT THE COMPANY

In 1921, IDT Biologika started as a state institute with three employees. Today, 100 years later, IDT Biologika is a global biopharmaceutical contract development and manufacturing organization that specializes in the production of innovative live viral vaccines, viral vectors for gene & immune therapeutics, oncolytic viruses, virus-like particles and other sterile liquid or lyophilized biologics to improve human health worldwide. Our Analytical Development Department uses NGS for detecting adventitious viruses in the manufacturing processes. We validated our NGS analytics platform, designed and developed a bioinformatics workflow and validated it in conformance with GAMP 5 after qualifying the required IT infrastructure.