Mutagenesis, 2023, XX, 1–9
https://doi.org/10.1093/mutage/gead012
Advance access publication 10 June 2023
Meeting report

Meeting report

Next Generation Sequencing Workshop at the Royal Society of Medicine (London, May 2022): how genomics is on the path to modernizing genetic toxicology

Anthony M. Lynch[1,2,*,‡], Thalita B. Zanoni[3,‡], Jesse J. Salk[3,4,5], Inigo Martincorena[6], Robert R. Young[3], Jill Kucab[7], Charles C. Valentine, III[3], Carole Yauk[8], Patricia A. Escobar[9], Kristine L. Witt[10], Roland Frötschl[11], Simon H. Reed[12], Anne Ashford[13]

[1]GSK R&D, Stevenage, United Kingdom
[2]School of Medicine, Swansea University, Swansea, United Kingdom
[3]TwinStrand Biosciences, Seattle, WA, United States
[4]Division of Medical Oncology, University of Washington, Seattle, WA, United States
[5]Fred Hutch Cancer Center, Seattle, WA, United States
[6]Cancer, Ageing and Somatic Mutation Programme, Wellcome Sanger Institute, Hinxton, United Kingdom
[7]Genetic and Environmental Toxicology, King's College, London, United Kingdom
[8]Department of Biology, University of Ottawa, Ottawa, ON, Canada
[9]Merck & Co. InC., Rahway, NJ, United States
[10]National Institute of Environmental Health Sciences, Division of Translational Toxicology, Research Triangle Park, NC, United States
[11]Federal Institute for Drugs and Medical Devices (Bundesinstitiut für Arzneimittel und Medizinprodukte), Bonn, Germany
[12]School of Medicine, Division of Cancer and Genetics, Cardiff University, Cardiff, CF14 4XN, United Kingdom
[13]Safety Innovation, Safety Sciences, Clinical Pharmacology & Safety Sciences, R&D, AstraZeneca, Cambridge, United Kingdom

‡AML and TBZ are joint first authors.
*Corresponding author. Department of Genetic Toxicology, GlaxoSmithKline (GSK), Gunnels Wood Rd, Stevenage SG1 2NY, United Kingdom. E-mail: anthony.m.lynch@gsk.com

Abstract

The use of error-corrected Next Generation Sequencing (ecNG) to determine mutagenicity has been a subject of growing interest and potentially a disruptive technology that could supplement, and in time, replace current testing paradigms in preclinical safety assessment. Considering this, a Next Generation Sequencing Workshop was held at the Royal Society of Medicine in London in May 2022, supported by the United Kingdom Environmental Mutagen Society (UKEMS) and TwinStrand Biosciences (WA, USA), to discuss progress and future applications of this technology. In this meeting report, the invited speakers provide an overview of the Workshop topics covered and identify future directions for research. In the area of somatic mutagenesis, several speakers reviewed recent progress made with correlating ecNGS to classic *in vivo* transgenic rodent mutation assays as well as exploring the use of this technology directly in humans and animals, and in complex organoid models. Additionally, ecNGS has been used for detecting off-target effects of gene editing tools and emerging data suggest ecNGS potential to measure clonal expansion of cells carrying mutations in cancer driver genes as an early marker of carcinogenic potential and for direct human biomonitoring. As such, the workshop demonstrated the importance of raising awareness and support for advancing the science of ecNGS for mutagenesis, gene editing, and carcinogenesis research. Furthermore, the potential of this new technology to contribute to advances in drug and product development and improve safety assessment was extensively explored.

Introduction

In recent years, error-corrected Next Generation Sequencing (ecNGS) has emerged as a valuable, highly sensitive, and accurate method for detecting and characterizing mutations in any species/sample from which DNA can be isolated. Accordingly, there have been several reports in the literature where mutagenicity and carcinogenicity studies have applied ecNGS to quantify drug-/chemical-induced mutations and mutational spectra associated with cancer risk [1–4]. ecNGS has potential applications in genotoxicity assessment as a new readout for traditional models, for mutagenesis studies in 3D reconstructed models, organoid cultures ,and microphysiological systems (MPS), and for detecting off-target effects of gene editing tools. Additionally, early data suggest ecNGS can measure clonal expansion of cells carrying mutations in cancer driver genes as a mechanism-agnostic early marker of carcinogenic potential and for direct human biomonitoring.

In May 2022, a Next Generation Sequencing Workshop was held at the Royal Society of Medicine in London and

supported by the United Kingdom Environmental Mutagen Society (UKEMS) and TwinStrand Biosciences (WA, USA). The aim was to bring together experts in the field who have been using or are interested in exploring, ecNGS as a new tool to understand how this novel genomic approach could be used to identify, characterize, and further describe mutagenic and/or carcinogenic modes of action and improve human safety assessment. An assembly of 11 international speakers was invited to present a program organized by the scientific organizing committee (Table 1) with the organization of the event ably supported by Affinity Events (UK) and Pacific Shore Marketing (USA). Approximately 80 delegates attended the workshop from a diverse range of professional backgrounds, including academia, industry, governmental institutions, international consortiums, and regulators. The focus of the workshop was to educate and update delegates on how NGS and ecNGS have been used to advance genetic toxicology and carcinogenicity assessment, inform improved mechanistic understanding, and review its potential to contribute to reducing the risk of genetic damage in humans.

The regulatory evaluation of mutagenesis is currently restricted to bacterial cells (Ames test), or specific gene loci in mammalian cells *in vitro* (e.g. thymidine kinase (TK) locus in the mouse lymphoma assay) and/or rodent transgenic gene mutation (TGR) assays (Muta™Mouse, BigBlue®, and gptDelta) or the Pig-a gene mutation assay. The assessment of mutation at these loci provides little-to-no information beyond a binary yes or no effect on mutagenesis and, in the absence of laborious DNA sequencing of many individually isolated clones, generally no insight into the mechanistic mode of action of mutagenicity. For example, in pharmaceutical development, the International Conference on Harmonization (ICH) S2 guideline on Genetic Toxicology testing indicates the potential for mutagenesis may be addressed using a gene mutation assay in bacteria, that is, the Ames test (OECD 471) and/or an *in vitro* mammalian cell gene mutation test using either the TK gene (OECD 490) or hypoxanthine-guanine phosphoribosyl transferase (HPRT) or a transgene of xanthine-guanine phosphoribosyl transferase (XPRT) genes (OECD 476), although the latter is seldom used. The current paradigm to provide weight of evidence to de-risk an *in vitro* positive mutagenic response is to perform extensive *in vivo* follow-up testing using transgenic rodents (OECD 488) and a carcinogenicity endpoint if this parameter is still of concern (OECD 451).

Given these limitations, there is a need to develop assays that measure induced mutations holistically across the genome and that can be used to better understand the relationship between mutation induction and carcinogenic/heritable disease risk in humans from environmental chemical exposures and medicine safety. As such, there are ongoing efforts by international scientists from consortia, academia, and industry to test whether ecNGS is a fit-for-purpose alternative approach to address the current problems encountered in standard mutagenicity and carcinogenicity testing. As a novel genomic technology, ecNGS allows the evaluation of mutation using genome-wide or genome-representative panels of loci to determine mutation frequency directly, in DNA from any given sample type (i.e. bacteria, cultured mammalian cells, or animal tissues). In addition, ecNGS protocols such as duplex sequencing (DuplexSeq™) (TwinStrand Biosciences) or NanoSeq (Wellcome Sanger Institute) can detect very low-frequency variants, on the order of 1-in-10-million bases sequenced, allowing a precise

Table 1. Next Generation Sequencing Workshop Program & Speakers

Chairs: Anthony M Lynch (am) & Aaron Boswell (pm)	**Aims of meeting and Q&A facilitation**
Session 1. Error corrected next generation sequencing (ecNGS): Current and emerging technologies	
Jesse J. Salk (TwinStrand Biosciences, USA)	*Introduction to ecNGS: DuplexSeq, BotSeq, etc*
Inigo Martincorena (Sanger Institute, UK)	*Introduction to NanoSeq*
Bob Young (GTTC-ecNGS workgroup, USA)	*Duplex Sequencing as its potential to transform cancer safety assessment*
Session 2. ecNGS concepts and mutational signatures	
Jill Kucab (Kings College London, UK)	*Investigating mutational signatures of carcinogens and chemotherapeutics using human tissue-derived organoids.*
Clint Valentine (TwinStrand Biosciences, USA)	*Fundamental concepts with Duplex Sequencing mutagenesis data and trinucleotide signatures*
Session 3. Mutagenesis, carcinogenesis, and regulatory testing	
Carole Yauk (University of Ottawa, Canada)	*Exploring the utility of ecNGS in regulatory toxicology: proof of concept and validation studies in vivo and in vitro*
Patricia Escobar (Merck, USA)	*Genotoxic vs. nongenotoxic carcinogens*
Kristine Witt (NIEHS, USA)	*Duplex Sequencing™: A game-changer in genotoxicity testing and cancer risk assessment?*
Roland Frotschl (BFarm, Germany)	*Regulatory challenges and opportunities*
Session 4. Skills for ecNGS	
Simon Reed (Cardiff University, UK)	*Induce-Seq technology*
Anne Ashford (AstraZeneca, UK)	*Advanced assessment of mutations in an in vitro Pig-a assay through Duplex Sequencing*
Panel Discussion	*Q&A from the day and What next NGS in UK*
	Wrap up

estimate of mutation frequencies around the true spontaneous rate using single-molecule mutation calling. This level of precision also provides the opportunity to evaluate very early clonal expansion of mutations, including cancer driver mutations, as early markers of carcinogenesis initiation. These can occur prior to histopathological changes associated with neoplasia and many months before tumour formation occurs and, uniquely, serve as a marker of cancer risk from carcinogenic agents with a non-mutagenic mode of action. Another great advantage of ecNGS is that the results across different species and samples can be standardized to a simple format for comparison and thereby provide mechanistic insight into the mode of action of mutagenesis by evaluating (i) mutation frequency (ii) simple spectra, and (iii) trinucleotide spectra. Furthermore, compared with traditional mutation assays, the assessment of mutagenesis using ecNGS does not rely on any form of biological selection, which means there are fewer assay elements to contribute to inter-lab variability.

To explore the potential of ecNGS, the workshop focused on biological effects and technologies that would yield the most useful information for evaluating human risk of genetic damage. The workshop was organized into four sessions covering: (i) ecNGS: Current and emerging technologies; (ii) ecNGS: Concepts and mutational signatures; (iii) Mutagenesis, Carcinogenesis and Regulatory Testing; and (iv) Skills required to apply ecNGS. There was extensive opportunity for questions and answers following each session along with a focused poster display on ecNGS topics from participating delegates. As such, the potential of this new technology to contribute to advances in drug and product development and improved safety assessment was extensively explored. Talks highlighted opportunities for improving current assays, including an alternative approach for the 2-year rodent cancer bioassay, and providing more informative and realistic insights from *in vitro*, *in vivo*, and even in-human studies.

Workshop summary

Prof. Anthony Lynch (Chair of the Scientific Steering Committee) opened the workshop, welcoming the speakers and delegates. He outlined several key themes that he hoped would be addressed during the day. These included:

(i) Application of ecNGS to mutational fingerprints in cancer—linking these to potential causal agents and correlating *in vitro* and/or *in vivo* mutagenesis data with cancer driver mutations and human tumour data.
(ii) ecNGS in regulatory genotoxicity testing—what are the opportunities, where are we now and what are the gaps to build confidence and support discipline and regulatory acceptance?
(iii) ecNGS in regulatory carcinogenicity testing—what are the opportunities (e.g. revision of ICH S1), where are we now and what are the gaps to build confidence and support regulatory acceptance?
(iv) Potential ecNGS applications in other sectors (e.g. consumer health, chemical industry, environmental chemicals assessment), and supporting the 3R's.
(v) ecNGS in regulatory testing to support gene editing/therapy approaches and concerns for potential insertional and/or off-target effects.
(vi) Comparison of emerging technology approaches and what are the barriers to advancing these in our science.

In addition, Prof Lynch's introductory comments focused on the importance of raising awareness and support for advancing the science of ecNGS to mutagenesis and carcinogenesis research and regulatory applications in the UK and beyond. A summary of each scientific presentation is provided below.

Session 1—ecNGS: current and emerging technologies

The intention of this session was to introduce the concepts of ecNGS and to discuss how advances in genomic technology can be beneficial to innovate and change the current research and regulatory testing paradigms in mutagenesis and carcinogenesis research.

Prof Salk (TwinStrand Biosciences, WA, USA) opened the session with an overview of general concepts and terminology related to mutagenesis and cancer in the context of ecNGS. The goal was to provide a foundational background understanding on how this novel technology can be used to address mutagenicity and carcinogenicity. The two broad classes of mechanisms that may be involved in the process of cancer formation are genotoxic and non-genotoxic (i.e. non-DNA reactive). The former contributes to the increased formation of mutations across the genome, including those in cancer driver genes, whereas the latter promotes clonal expansion of partially transformed cells with pre-existing mutations in cancer driver genes. Prof Salk described duplex sequencing, a type of ecNGS with error rates on the order of 1 error per 10 million normal DNA bases. This error rate is close to the background somatic mutation load of adult human and rodent tissues. This makes it possible to use duplex sequencing to directly measure the rate and spectra of mutations induced by different mutagens, both *in vivo* and *in vitro*. Insights into a chemical exposure's mutagenic mode of action can be inferred from the patterns of mutations observed (including simple mutation spectra and trinucleotide-context mutational signatures). These signatures can be compared to those present in the catalogue of somatic mutations in cancer (COSMIC) database (see https://cancer.sanger.ac.uk/cosmic), which comprises mutational signatures previously described in published cancer genomes—in some cases linked to known mutagenic processes. Conventional DNA sequencing is a powerful tool that has been regularly used to sequence malignant tumours allowing scientists and oncologists to determine the genes most often responsible for driving many cancers as well as inferring the mutagenic processes active during cancer development. However, this only works once the tumour has been formed. In non-clonal tissues, standard sequencing technology cannot identify *de novo* mutation or clones at early stages because sequencing errors produce background noise that obscures the vastly lower frequency random mutations or rare subclones. In contrast, ecNGS methods like duplex sequencing greatly reduce sequencing errors, and depending on context (e.g. underlying somatic mutation frequencies) they may enable much earlier detection of emerging preneoplastic clones, an important phenotypic marker of carcinogenesis.

The second presentation, by Inigo Martincorena (Sanger Institute, Cambridge, UK), continued the session theme by describing his institute's research effort on mutagenesis and

carcinogenesis in humans and the focus of his own team on understanding somatic mutation in normal tissues using ecNGS. He described how these should be key for understanding the earliest stages of cancer outgrowth and potentially informing risk stratification and early detection efforts. Until recently this had represented largely uncharted territory because of technical limitations for accurate mutation analysis. During his talk, Dr Martincorena introduced NanoSeq, a modified duplex sequencing protocol developed at the Sanger Institute (UK) that avoids library preparation errors to achieve error rates <5e–9 errors per bp [5]. Since the somatic mutation density in human adult tissues is close to 1e–7 mutations/bp, this reduction in the error rates of duplex sequencing enables more accurate studies of somatic mutagenesis *in vivo* and *in vitro*. Dr Martincorena also described their work on studying somatic mutation in normal tissues, starting with their 2015 study of normal human skin. Surprisingly, they discovered a high mutation burden, with more than 10,000 somatic mutations per skin cell and numerous cancer driver mutations (CDM). Their analysis suggested that a quarter of skin cells carry mutations in *NOTCH1-3, TP53, FAT1,* and/or *RBM10* (known cancer genes in squamous cell carcinoma), suggesting a strong positive selection of mutant cells carrying driver mutations [6]. In collaborations with the teams of Phil Jones, Peter Campbell, and Mike Stratton and using cutting-edge technologies such as laser capture microscopy, they were able to show a similar phenomenon in other epithelial tissues, such as oesophagus [7] and bladder [8]. The analysis also revealed some mutational signatures active in normal tissues that corresponded to those identified in the COSMIC database (see https://cancer.sanger.ac.uk/cosmic). Using NanoSeq, Dr Martincorena's team then studied the somatic mutation rates and spectra in a variety of tissues and cell types, including blood, colonic epithelium, smooth muscle, and post-mitotic neurons [4]. They confirmed that dividing and non-dividing cells accumulate mutations linearly throughout life, with modest differences in rates and signatures between highly proliferative and slow-dividing or even post-mitotic cells. This suggests that somatic mutations in most tissues occur independently of cell division. The idea that normal cells are loaded with mutations that accumulate with age even in the absence of cell division was a surprise to many of the genetic toxicologists present and elicited many questions from the audience. Professor David Phillips (King's College, London) asked about the T-to-C mutation signature seen in normal bladder and asked if this had been seen in cancer patients? Dr Martincorena confirmed that this signature was subsequently identified in bladder cancers from smokers by the team of Ludmil Alexandrov (UC San Diego). A second question was asked about the mechanisms responsible for fixing mutations in both strands of DNA in the absence of cell division. Dr Martincorena said his best explanation for this at present was 'fixation of mutations during DNA repair'.

Robert (Bob) Young (Co-chair of the Health and Environmental Sciences Institute (HESI) GTTC-Error Corrected Sequencing Workgroup) concluded the first session by describing the history of the science of mutation detection, for example, in drosophila and mouse spot test and work leading to the development of the transgenic gene mutation (TGR) assays, for example, Big Blue® mouse or rat and Muta™Mouse in the 1990s and their validation in the 2000s [9]. The latter assays have become the default for regulatory evaluation of a compound's potential to induce mutagenicity *in vivo* leading to the development and adoption of a test guideline (TG) by the Organisation for Economic Co-operation and Development (OECD TG488). However, a variety of practical limitations have been recognized with these assays, such as the need for a dedicated single endpoint study using specific transgenic animals, the use of small, non-mammalian, lambda bacteriophage vector-based reporter genes (e.g. *lacZ, lacI, GPT,* or *cII*) as the 'target' for mutation assessment, and the complexity of the methods required to rescue the vector and determine whether mutations have been induced in the reporter gene. Mr Young alluded to the first symposium related to ecNGS at the Environmental Mutagenesis and Genomics Society conference in 2016 (chaired by him and Dr Stephanie Smith-Roe (NIEHS) and entitled 'Analysis of In Vivo Mutation by Next Generation Sequencing: Applications in Cancer Genomics'). This was followed by an early proof of concept investigation in a collaborative project involving Amgen, Bioreliance (Sigma-Aldrich), and TwinStrand Biosciences. The project investigated ethylnitrosourea and benzo(a)pyrene-induced mutations in liver and bone marrow and demonstrated equivalency of duplex sequencing analysis at the *cII* locus to the Big Blue® TGR assay using the *cII* plaque assay. The study also showed how mutation analysis could be extended into endogenous genetic loci, demonstrating that ecNGS could detect both background and induced mutant frequencies and mutation spectra in native mammalian genomic DNA sequences, including proto-oncogenes. The collaboration also reported on urethane mutagenesis in the lungs of Tg-rasH2 transgenic mice, the accepted transgenic mouse strain under ICH S1 for a six-month mouse carcinogenicity assay. Random genome-wide mutagenesis was observed in endogenous genes and the human *HRAS* proto-oncogene carried as a transgene in Tg-rasH2 mice. In addition, hundreds of fold increases in identical (presumably clonal) A·T→T·A transversions in codon 61 of exon 3 in the *HRAS* transgene were observed in tissues harvested 29 days following a urethane treatment that yields nearly 100% incidence of lung adenocarcinomas in mice after 11 weeks. This clonal expansion exhibited the same mutational signature observed in human lung tumours associated with urethane exposure (see https://cancer.sanger.ac.uk/cosmic). He reiterated that detection of clonal expansions of a mutated cancer driver gene reflects a fundamentally different phenomenon than genome-wide mutagenesis [10]. Whereas the latter requires a mutagenic mode of action for a signal to be detected, the former, in theory, should be detectable for all carcinogens, including those without mutagenic modes of action. Mr Young suggested that ecNGS can offer an alternative to existing TGR mutation assays and provide a second *in vivo* genotoxicity biomarker under ICH S2R1 Option 2 and/or be used as a biomarker of cancer risk in repeat dose toxicology studies using wild-type animals possibly under a revised ICH S1. Moreover, he proposed that since ecNGS should be able to detect both genotoxic and nongenotoxic carcinogens in repeat dose studies using wild-type animals, this offers the potential for transforming current approaches to cancer safety assessment. As such, Mr Young argued that ecNGS represents a disruptive new technology for potential applications in pharmaceutical non-clinical safety assessment, as part of drug discovery and development and supporting impurity qualifications. He concluded his presentation by highlighting some of the lessons learnt from the validation of 'new technologies' and their journey to routine implementation in a

regulatory context. These included observations that since industry and regulators have historically been slow to adopt new technology, collaborative qualification efforts are needed with early participation of all stakeholders, coordination of efforts by groups such as HESI Genetic Toxicology Technical Committee (GTTC), and that some type of official recognition of the adequacy of a new technology is needed prior to wide-spread adoption. These should include evidence of robust technology transfer and data integrity, a consensus on data interpretation, and strong regulatory engagement and broad stakeholder management.

Following the morning presentations there was an opportunity for further questions and answers with the session speakers. One question from the audience addressed non-mutagenic carcinogens and how ecNGS could be used as part of a weight of evidence for risk assessment. Mr Young mentioned ongoing studies at the National Center for Toxicological Research and Dr Salk suggested that tumour-specific CDM panels may require organ and species-specific considerations in addition to work defining what 'normal' looked like. He added that translation of animal data to inform human risk would likely be complex although he thought ecNGS, as a platform, can be applied to any species and therefore could potentially bridge 'clinical oncology, pre-cancer biology and non-clinical safety research'. He said as a physician that he thought this was an exciting opportunity and further reiterated the importance of Dr Martincorena's work to establish what 'normal' accumulation of mutations looks like across a cohort of healthy individuals before prematurely concluding that all mutations are abnormal. This was echoed by Dr Patricia Escobar (Merck & Co., NJ, USA).

Session 2: ecNGS concepts and mutational signatures

The intention of this session was to introduce some important concepts of ecNGS, particularly around bioinformatic tools, and to present fundamental research to elucidate mutational signatures of known chemical carcinogens in cultured cells and help determine any association with those previously identified in human cancers.

The first presentation was given by Jill Kucab (Kings College London, UK) who described her work with Prof. Phillips as part of the CRUK Grand Challenge project, 'Mutographs of Cancer'. Dr Kucab is currently investigating mutational signatures of carcinogens and chemotherapeutics using human tissue-derived organoid models. By identifying the mutational fingerprints of environmental carcinogens in experimental systems, she hopes to better understand the contribution of these factors to the mutations we observe in human cancer. Initially, her team used cell culture models to determine mutational signatures induced by environmental and chemotherapeutic agents, where, using whole-genome sequencing of cloned single cells, they identified 41 characteristic single base substitution signatures in human induced pluripotent stem cells treated with 79 DNA-damaging agents [11]. A number of these signatures corresponded with some of those identified in the COSMIC database. However, the necessity for single-cell cloning was a limiting factor in the research and Dr Kucab explained how the team is now using normal human tissue-derived organoids, along with a genome-wide, duplex sequencing (NanoSeq), to examine mutations induced by a further panel of agents. The sensitivity of NanoSeq has enabled the team to bypass the single-cell cloning step, and Dr Kucab presented some of the mutational signatures that her team has recently identified in organoids treated with environmental and chemotherapeutic agents, comparing these signatures to those found in people exposed to the same agents.

Charles (Clint) Valentine (TwinStrand Biosciences, USA) addressed the workshop with a presentation summarizing fundamental concepts of how bioinformatics and data analyses can be carried out to detect mutagenesis using duplex sequencing, while providing clarification of the underlying algorithms and how data are typically generated. Data analysis for general NGS requires a bioinformatics workflow pipeline that contains many steps, which can be exemplified by the TwinStrand Biosciences DuplexSeq™ Software workflow [10]. Mr Valentine stated that advanced algorithms are used to characterize sequencing data, and analysis reduces big data to small data, for example, terabytes of raw sequence reads being reduced to a variant call set of ~100 'mutational' events, suitable for human understanding. During the ecNGS data analysis process, the DNA read sequences are first manipulated so that molecular labelling elements (such as a unique molecular identifier) are extracted and saved. The read sequences are then aligned to the reference genome and a combination of the molecular labelling elements and inferred shear points of the molecule, which are derived from the alignment coordinates, are used to group the reads into families. These families are evaluated to create a consensus sequence that eliminates nearly all read and strand-specific technical artefacts. Finally, a variant caller uses the information from all consensus sequences to call true variant alleles measured directly from representations of the source DNA itself. DNA variants can take many forms but are most commonly base pair substitutions, insertions, deletions, and multi-nucleotide substitutions. Extremely rare inter-species contamination can be a common and troublesome technical artefact in ecNGS experiments, but Mr. Valentine said that by design these can be automatically filtered from the data using a fast and accurate taxonomic classifier.

He added there are various mutagenicity assay metrics and figures that can be generated from ecNGS data, citing simple mutant frequency as an interesting example. This output describes the type and specific location of the mutation in the genome; this easily identifies potential jackpot mutations (i.e. clonal mutations) that are discounted in mutagenesis studies and collapsed down to a single mutation event when all observations are thought to have arisen from a single clonally expanded mutation event. Clonal bias correction can be useful to avoid confounding interpretations that has been a historical challenge with other mutagenesis assays, such as the TGR. Simple base substitution spectra using the 12 transition and transversion mutations are generated by plotting the proportion of all base substitution mutations present in a sample. Unsupervised hierarchical clustering, a method wherein a computer software groups samples with similar types and proportions of mutations together, can reveal qualitative relationships between samples. In addition, analysing the simple mutations in the context of 5-prime and 3-prime flanking bases offers 96 different 3-base pair mutational targets for evaluation termed trinucleotide spectra. Analysis of this trinucleotide spectra offers rich opportunities to investigate class-specific mechanistic modes of action. For example, there

is the potential to distinguish a weak mutagenic dose from a non-mutagenic dose using mutational spectra when the overall mutation frequency in both samples is nearly equal (and assuming no indirect influence on the mutational spectrum). Finally, trinucleotide spectra derived from mutagen-treated samples can provide insights on mutagenic mechanisms of action and associations with human cancer, for example, by comparing the trinucleotide spectrum of a sample to annotated trinucleotide signatures in the COSMIC spectra database. A key learning from Mr Valentine's presentation was the importance of strong bioinformatics support to the deployment of ecNGS technologies and the need for a multidisciplinary approach to data analysis and interpretation that extends beyond classic laboratory scientists.

Session 3: mutagenesis, carcinogenesis, and regulatory testing

The afternoon was chaired by Aaron Boswell (TwinStrand Biosciences, USA). The intention of the third session was to dive deeper into the opportunities for ecNGS in regulatory genotoxicity and carcinogenicity testing for human risk assessment.

Carole Yauk (University of Ottawa, Canada) discussed how her 'Genomics in Regulatory and Applied Toxicology (GReAT) Laboratory' and Health Canada are collaborating to apply genomics in regulatory toxicology testing. Recognizing the limitations of current test battery assays, as described above, the aim of the GReAT lab is to: (i) explore mutagenic responses by duplex sequencing to confirm performance across different genotoxic modes of action in diverse tissues *in vivo* and in cell lines; (ii) establish the degree of qualitative and quantitative concordance relative to conventional mutagenicity endpoints; (iii) define optimal experimental designs; and (iv) investigate the added value of the mechanistic information produced by this approach. Prof Yauk described some early proof of concept and validation work exploring these questions to query modes of mutagenic action associated with adverse outcome pathways (AOPs) developed under the auspices of HESI [12–14]. Her laboratory collaborated with the team of Dr Francesco Marchetti from Health Canada to evaluate the ecNGS mutation spectrum and frequency in bone marrow samples from Muta™Mouse males exposed to benzo(a)pyrene and procarbazine and showed a high correlation with the *lacZ* assay in the same tissue, but reduced sensitivity when clonally expanded mutations were excluded. They also observed inter-locus differences in mutation frequency using ecNGS, with higher susceptibility to induced mutagenesis in inter-genic-regions consistent with a lack of transcription-coupled repair at these sites [10]. Importantly, benzo(a)pyrene and procarbazine induced compound-specific mutational spectra that corresponded to published COSMIC signatures and were consistent with their known mechanisms of action. Next, Prof Yauk described a time-series analysis in TK6 cells that revealed ethylnitrosurea-induced mutations are measurable within 48 h of exposure and that longer durations were not more effective for detecting this potent alkylating agent. The work supports the potential for integration of different endpoints in a single assay format (e.g. mutation assessment and chromosomal/DNA damage) which can be used to inform on mechanism and human relevancy and reduce animal usage. Furthermore, by harnessing AOPs, Prof Yauk suggested these new tests can contribute to a modern integrated approach to testing and assessment for genotoxicity, and she expressed her excitement about the potential of these new technologies to impact regulatory testing going forward.

In the following Q&A, the Chair asked Prof Yauk about the divergent benchmark dose (BMD) modelling seen with duplex sequencing mutation data versus transgenic rodent mutation frequency data and whether scaling to the maximum response had been considered. Prof Yauk said, 'the optimal approach to BMD modelling of ecNGS data remains to be determined and is an important area for future research'. She added 'when clonally expanded mutations were included in the modelling, the ecNGS and TGR assay BMDs seemed to converge', saying 'I was quite impressed with that. But what was really nice was the increased precision on the duplex sequencing'. This was followed by an audience question about Prof Yauk's planned longer-term Muta™Mouse studies (30, 90, and 180-day studies). Prof Yauk explained that she and her colleagues were interested in whether mutagenic potency would be impacted by duration, and if so, by how much—adding that there was a move to make use of quantitative information in genetic toxicity testing to support improved risk assessments.

Dr Patricia A. Escobar (Merck & Co., Inc, NJ, USA) continued the symposium theme, explaining that carcinogenic risk assessment is an essential component of pharmaceutical and industrial/agrochemical safety testing, and that improved biomarkers of carcinogenicity are needed to accelerate detection of carcinogens in advance of the currently required 2-year rodent bioassay. She described how the genetic toxicology testing battery is used for early assessment of genotoxic carcinogens, but this approach was inappropriate for the assessment of non-genotoxic carcinogens (NGCs), that is, substances that induce cancer through indirect stimulation. However, by postulating that atypical cells that appear within a hyperplastic lesion are likely to involve the clonal expansion of cells containing cancer drive gene mutations (CDGMs), Dr Escobar reasoned that the sensitivity and accuracy of ecNGS could be used to identify specific CDGMs associated with clonal expansion and that these could serve as an early biomarker of carcinogenesis induced by NGCs in advance of overt histopathological changes. To test this hypothesis, Dr Escobar helped to establish a multi-disciplinary team to design specific CDGM-targeted gene panels for the evaluation of tumours in Tg-rasH2 mice and wild-type rats, two commonly utilized models for carcinogenicity testing. For Tg-rasH2 mice, they targeted the human *HRAS* transgene and endogenous murine Ras genes. Tg-rasH2 mice were treated for 1, 3, and 6 months with the NGC PCB-126 or vehicle, and the Duplex Sequencing (DS) analysis was performed in the 3 months samples. Clonal expansion was observed in Tg-RasH2 mice treated for 3m with 160 ug/kd-day PCB-126, but not in control Tg-RasH2 mice. For wild-type rat, they determined common CDGMs by using whole-exome sequencing of liver tumours from two different chronic non-genotoxic treatments, nafenopin and phenobarbital. Eleven repeatedly occurring CDMs were identified and combined with orthologs of established human/mouse cancer driver genes for a 27-gene rat duplex sequencing panel. These panels were then evaluated for their ability to detect CDGM-bearing clones in advance of histopathology in wild-type rats treated

for 1, 6, and 9 months with the NGC PCB-126 or vehicle. The mutation frequency was not changed when comparing treated samples with vehicle control, but the incidence of clonal expansion in a subset of genes at 6 and9 months was increased in PCB-126 treated wild-type rats when compared with the vehicle control. The analysis was underway and would be published.

Among questions from the audience, it was asked if anticipated ecNGS would distinguish incidental background lung tumours in control Tg-rasH2 mice and the same tumour type in treatment groups based on mutations in cancer driver genes. Dr Salk said '…with urethane, a strong point mutation signal in the RAS transgene was observed' but added 'COSMIC-like databases do not yet exist to anywhere the same magnitude for rodent tumours, and the fundamental patterns of rodent cancer driver mutations have yet to be defined'. Therefore, comparison with control will have to suffice for the time being and 'we just have to do a lot more studies!'.

The next speaker, Kristine Witt (NIEHS, USA, retired) began her presentation by posing the following question—'Is duplex sequencing a game-changer in genotoxicity testing and cancer risk assessment?' and described how the U.S. National Toxicology Program (NTP; recent name change to Division of Translational Toxicology, DTT) has conducted genotoxicity testing for >40 years, and based on correlations to rodent carcinogenicity, the numerous tests initially evaluated by the NTP had evolved and been refined over the years. Nevertheless, for mutation assessment, she said the Ames assay remains the cornerstone for identifying mutagens and mutagenic carcinogens, even after four decades! She then explained that with the advent of sophisticated molecular technologies, genetic toxicologists are exploring alternative testing approaches that may provide improved human translation, increased precision, and valuable mutagenic mode of action information. In this regard, Ms Witt related how the NTP was evaluating duplex sequencing, describing the technology 'as a promising Next Generation Sequencing method that enables detection and detailed characterization of extremely rare mutations across the genome landscape, potentially revolutionizing mutagenicity testing'. She said proof-of-principle studies had been conducted with the potent tool-mutagen N-ethyl-N-nitrosourea (ENU) in multiple tissues of male rats following a single dose. Patterns of mutation frequency and spectra were assessed at 3 h, 24 h, 7 days, and 28 days after dosing. Mutation frequency increased with time during the first week post-exposure and the characteristic ENU mutation spectrum was established by the 7-day time point in all analysed tissues. The studies in rats were followed by studies in cultured human cells (TK6) where patterns of mutation induction were measured at 24 h intervals over several days and across a broad range of doses. In the *in vitro* studies, significant increases in mutation frequencies were observed at all doses and timepoints, while no evolution in mutation spectrum was observed. The results, she said, were encouraging and correlated well with existing methods of mutation assessment, thereby demonstrating, in her opinion, the value of duplex sequencing in identifying and characterizing mutational events. The data were being prepared for publication and will be available to the community soon. It was evident from this pioneering work that much of the groundwork has been set for developing new approaches to cancer risk assessment and that these new methods are better aligned with the 3Rs principles, especially given the opportunity to integrate into repeat dose toxicity studies, thereby removing the need for stand-alone studies for mutation assessment.

Ms Witt was asked about the observed evolution of the ENU mutational spectrum in somatic cells of male rats during the first 7 days after treatment and she replied that she thought the data showed an initial shift from the background spectrum at the 24 h timepoint 'and at 7 days we see the completion of the process'. She added it would be interesting to go back and sample the tissues at 24 h intervals during the early post-treatment period to further characterize the transition of the mutation spectrum and 'see at what timepoint the characteristic ENU spectrum stabilizes'. There was speculation from participants about the role of cell cycle and DNA repair on the processes, summed up by Ms. Witt as 'a lot to do yet!'.

Dr Roland Frötschl (BfArm, Germany) concluded session 3 by reviewing the regulatory challenges and opportunities for NGS based on his own, individual viewpoint. He began by reminding the audience of the potential consequences of somatic mutation, including carcinogenesis, and citing reviews on somatic mutation and human disease by Erickson [15] and a study investigating Pig-a mutation in patients receiving chemotherapy [16]. He then outlined the objectives of regulatory genotoxicity testing as identifying potentially DNA damaging agents that may be carcinogenic or cause heritable effects, protecting the individual from high risk of developing diseases like cancer, atherosclerosis, some autoimmune diseases, birth-defects, or accelerated aging; and protecting future generations from heritable diseases, such as metabolic diseases, cystic-fibrosis, and increased predisposition/susceptibility for cancer or degenerative diseases. Dr Frötschl said the current testing paradigm provides no or very little information on specific mutations or mutation sites, whereas research had shown that different mutations have differing impacts on cancer risk, citing mutational signatures, the hallmarks of cancer and the impact of mutation on tumour suppressor genes and oncogenes. As such, Dr Frötschl posed the question as to whether it will be possible to develop an early warning system for the development of cancer based on mutational signatures (echoing a theme of cancer driver mutations introduced by Dr Escobar), and which signatures will be the most prognostic. He then switched his focus to address some of the challenges he perceived for the adoption of ecNGS in regulatory testing. He identified the need for validation and agreed implementation standards including data interpretation, understanding variability in mutation background, assay sensitivity, and defining criteria for a positive or negative outcome. He also emphasized the importance of defining the context by which mutational signatures can contribute to the latter. He foresees international guidelines will be required and these will need to address both *in vitro* and *in vivo* applications of the technology, whilst preparing the ground for the training of regulators. Overall, Dr Frötschl said the potential benefits offered by ecNGS, including the potential to inform on both genotoxic and NGCs, represents a golden opportunity to improve regulatory genotoxicity testing, particularly also in the context of the potential for integrating mutation detection into sub-chronic toxicity testing and better aligning with 3R principles.

In the Q&A, Prof Lynch asked about expediated pathways towards regulatory acceptance of ecNGS data, for example, via ICH S1 amendments, revision of OECD guidelines,

and publication of white papers. Dr Frötschl suggested the International Workshop on Genotoxicity Testing could be one vehicle where an expert working group could be assembled to develop a detailed position with data that can 'convince the industry and regulatory stakeholders'. He also suggested working in parallel with OECD engagement and added that ICH M7 guideline mentioned NGS as a potential follow-up of bacterial mutation assays.

Session 4—skills for ecNGS

Prof Simon Reed (Cardiff University, UK) opened by describing how our understanding of the formation and repair of DNA strand breaks in the genome depends on the accurate measurement of the frequency and position of DNA double-strand breaks (DSBs). He then outlined some current methods for assessing DNA strand breaks in cells and explained how such information can help determine a chemical's DNA damaging potential and inform on the safe development of therapies, including those developed using genome editing technologies Prof Reed explained how current DSB sequencing methods suffer from high background levels, the inability to accurately measure low-frequency endogenous breaks and high sequencing costs, all of which are caused by the PCR amplification of labelled breaks during typical NGS library preparation prior to break sequencing. He then described INDUCE-seq, a novel NGS method developed in his laboratory which overcomes these problems, detecting simultaneously the presence of lower-level endogenous DSBs caused by physiological processes, and higher-level recurrent breaks induced by restriction enzymes or those that can be induced by gene editing or chemical exposures [13,17]. INDUCE-seq is a novel PCR-free NGS library preparation that enables the introduction of sequencing adapters in DNA breaks inside the cell (*in situ*). The elimination of PCR steps during library construction avoids the introduction of biases in break quantification associated with amplification. This innovation allows exclusively breaks labelled *in situ* to interact with the NGS flow cell. As a result, every sequencing read obtained from the sample is derived from a unique break present and labelled in the cell sample. The duplex error-correction process used in ecNGS similarly eliminates artifactual errors introduced during the library preparation and amplification process, dramatically improving the sensitivity of off-target edit detection. The HESI working group CT-TRACS has been evaluating INDUCE-seq to screen for off-target gene editing events in a population of edited cells using a range of previously characterized guide RNAs and two different cell types. Previously, by measuring on- and off-target genomic breaks during EMX1 editing by Cas9 in HEK293 cells, Prof Reed's laboratory identified putative off-target edit-sites by INDUCE-seq and investigated the mutational outcome at these sites using amplicon sequencing. Typically, this method has a sensitivity of only ~1 in 1000, far lower than what can be achieved by ecNGS. Consequently, in the HESI study, the off-target sites identified are being deeply interrogated for mutations using a customized DuplexSeq assay to determine the frequency of off-target edits at each site. This confirmation step is important because not all double-stranded breaks result in a mutation, and, therefore, the approach offers the simultaneous advantages of highly specific, unbiased empirical genome-wide screening (necessary when you do not know where an off-target edit might occur in the genome) and highly sensitive confirmation of whether and to what extent off-target edits have been introduced in a population of edited cells. Validation of INDUCE-seq was investigated by comparing intra- and inter-laboratory reproducibility of INDUCE-seq in collaboration with AstraZeneca and Novartis. Prof Reed reported high correlations (R > 98%) in the data generated between each laboratory and he expressed the belief that this new technology would provide a means to accurately assess the potential risk of adverse events occuring following gene editing and/or vector-mediated genotoxicity related to human gene therapy.

The last speaker of the day, Anne Ashford (AstraZeneca, UK) presented a personal 'bench side' view of ecNGS describing her initial experience and scientific insights of introducing ecNGS into a genetic toxicology laboratory and evaluating the DuplexSeq™ platform (TwinStrand Biosciences). She used an *in vitro* version of the Pig-a gene mutation assay (OECD 470) as a comparison. In this assay, cells harbouring a mutant Pig-a gene lose GPI-anchored cell surface proteins (e.g. CD90) that can be detected by the loss of immunofluorescent labelling compared to wild-type cells using flow cytometry. The ratio of labelled and non-labelled cells gives a measure of mutation frequency. AstraZeneca has developed an in-house, *in vitro* version of the Pig-A gene mutation assay to determine the mutagenic potential of a test compound. The same cell populations were treated with two reference mutagens, ethyl methanesulfonate (EMS) and ENU and cell preparations were assessed by the Pig-a assay and the TwinStrand DuplexSeq™ assay. Dr Ashford said she had found establishing DuplexSeq™ to be straightforward and that the methods were reproducible and robust, and capable of accurately detecting rare and low levels of mutation induced by mutagenic compounds. She said the DuplexSeq™ assay was able to recapitulate the mutagenicity assessment at the Pig-a locus, and offered several advantages based on assay performance and important insights, via mutational spectra analysis, into the mutagenic mode of action.

Panel discussion

Following the final presentation, Prof Lynch invited the day's speakers to the podium for an extended Q&A session. The first question was about the COSMIC mutation database and when would the signature set be completed? Dr Kucab said the characterization of chemically induced mutation signatures in iPS cells and their correlation with cancer mutation signatures had 'only just started', with 41 of the 79 agents tested yielding signatures and not all had COSMIC correlates. Her lab was now investigating organoid tissue cultures, but this work was at an early stage. The next questions focused on the use of ecNGS in human environmental exposure studies, especially with complex mixtures. Dr Kucab said there were COSMIC signatures that are not matched by those identified experimentally, although this might be because certain human cohorts have yet to be identified/investigated, citing air pollution as an example, and speculating about possible links with exposure to polycyclic hydrocarbons.

Prof Yauk asked about replacing the cancer bioassay in rodents and whether increased mutation or clonally expanding mutations could be viewed as adverse events in of themselves and relevant to other diseases. Dr Salk pointed out that although the focus of the Workshop had been on genomic mutation, there was the mitochondrial genome to

consider, which is very different. Certain drugs, such as early antivirals, were known to be associated with mitochondrial toxicity. The development of ecNGS platforms could provide the means to investigate mitochondrial mutation, an area neglected by traditional Genetic Toxicology. Prof Yauk suggested that the use of the AOP framework (see OECD link) for linking exposure to an environmental agent and an apical adverse outcome of regulatory concern might be one way to facilitate adoption of ecNGS as data could be used to support key events that are critical for pathway progression.

Prof Lynch concluded the Workshop saying 'It had been a wonderful day. Great to get face-to-face in person again, with fantastic talks, great science, and lots of unanswered questions to address and move this forward'. He added that a Special Interest Group on ecNGS was being established by the UKEMS and that audio-visual recordings of eight presentations from the Workshop would be available to UKEMS members on the society website.

Acknowledgements

The authors would like to thank members of the scientific organizing committee (Robert Smith, David Phillips, Gareth Jenkins, Simon Reed, Darren Kidd, Jo Elloway, and Anne Ashford) for their contributions towards delivering the Next Generation Sequencing workshop and for their review of this meeting report. The authors thank Aaron Boswell for co-chairing the workshop. They are grateful to the United Kingdom Environmental Mutagen Society (UKEMS) and TwinStrand Biosciences for their financial support and assistance in hosting the workshop, along with the staff at the Royal Society of Medicine.

Conflict of Interest

T.B.Z., J.J.S., and C.C.V. declare that they are employees and equity holders at TwinStrand Biosciences, Inc. and are authors on one or more Duplex Sequencing-related patents. I.M. is an academic co-founder and consultant of Flagship Labs 86 Inc.

References

1. Valentine CC 3rd, Young RR, Fielden MR, et al. Direct quantification of in vivo mutagenesis and carcinogenesis using duplex sequencing. *Proc Natl Acad Sci (USA)* 2020;117:33414–25.
2. LeBlanc DPM, Meier M, Lo FY, et al. Duplex sequencing identifies genomic features that determine susceptibility to benzo(a)pyrene-induced *in vivo* mutations. *BMC Genomics* 2022;23:542–57.
3. Wang Y, Mittelstaedt RA, Wynne R. Genetic toxicity testing using human in vitro organotypic airway cultures: assessing DNA damage with the CometChip and mutagenesis by Duplex Sequencing. *Env Mol Mutagen* 2021;62:306–18.
4. Chawanthayatham S, Valentine CC, Fedeles BI, et al. Mutational spectra of aflatoxin B1 in vivo establish biomarkers of exposure for human hepatocellular carcinoma. *Proc Natl Acad Sci (USA)* 2017;114:E3101–9.
5. Abascal F, Harvey LMR, Mitchell E, et al. Somatic mutation landscapes at single molecule resolution. *Nature* 2021;593:405–10.
6. Martincorena I, Roshan A, Gerstung M, et al. Tumor evolution: high burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 2015;348:880–6.
7. Martincorena I, Fowler JC, Wabik A, et al. Somatic mutant clones colonize thehuman esophagus with age. *Science* 2018;362:911–7.
8. Lawson ARJ, Abascal F, Coorens THH, et al. Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science* 2020;370:75–82.
9. Lambert IB, Singer TM, Boucher SE, et al. Detailed review of transgenic rodent mutation assays. *Mutat Res* 2005;590:1–280.
10. Valentine CC 3rd, Young RR, Fielden MR, et al. Direct quantification of in vivo mutagenesis and carcinogenesis using duplex sequencing. *Proc Natl Acad Sci (USA)* 2020;117:33414–25.
11. Kucab JE, Zou X, Morganella S, et al. A compendium of mutational signatures of environmental agents. *Cell* 2019;177:821–836.e16.
12. Sasaki JC, Allemang A, Bryce SM, et al. Application of the adverse outcome pathway framework to genotoxic modes of action. *Env Mol Mutagen* 2020;61:114–34.
13. Dertinger SD, Kraynak AR, Wheeldon RP, et al. Predictions of genotoxic potential, mode of action, molecular targets, and potency via a tiered Multiflow® assay data analysis strategy. *Env Mol Mutagen* 2019;60:513–33.
14. Cho E, Allemang A, Audebert M, et al. AOP report: development of an adverse outcome pathway for oxidative DNA damage leading to mutations and chromosomal aberrations. *Env Mol Mutagen* 2022;63:118–34.
15. Erickson RP. Somatic gene mutation and human disease other than cancer. *Mutation Research* 2010;705:96–106.
16. Dobrovolsky VN, Elespuru RK, Bigger CAH, et al. Monitoring humans for somatic mutation in the endogenous PIG-a gene using red blood cells. *Env Mol Mutagen* 2011;52:784–94.
17. Dobbs FM, van Eijk P, Fellows MD, et al. Precision digital mapping of endogenous and induced genomic DNA breaks by INDUCE-seq. *Nature Communications* 2022;13:3989.