Visual Analytics: A Method to Explore Natural Histories of Oral Epithelial Dysplasia

Stan Nowak [1,*],**Miriam Rosin** [2,3], **Wolfgang Stuerzlinger** [1], **and Lyn Bartram** [1]

[1] *School of Interactive Arts and Technology, Simon Fraser University, Surrey, British Columbia, Canada*
[2] *BC Oral Cancer Prevention Program, Cancer Control Research, BC Cancer, Vancouver, British Columbia, Canada*
[3] *Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada*

Correspondence*:
Stan Nowak
snowak@sfu.ca

ABSTRACT

Risk assessment and follow-up of oral potentially malignant disorders in patients with mild or moderate oral epithelial dysplasia is an ongoing challenge for improved oral cancer prevention. Part of the challenge is a lack of understanding of how observable features of such dysplasia, gathered as data by clinicians during follow-up, relate to underlying biological processes driving progression. Current research is at an exploratory phase where the precise questions to ask are not known. While traditional statistical and the newer machine learning and artificial intelligence methods are effective in well-defined problem spaces with large datasets, these are not the circumstances we face currently. We argue that the field is in need of exploratory methods that can better integrate clinical and scientific knowledge into analysis to iteratively generate viable hypotheses. In this perspective, we propose that visual analytics presents a set of methods well-suited to these needs. We illustrate how visual analytics excels at generating viable research hypotheses by describing our experiences using visual analytics to explore temporal shifts in the clinical presentation of epithelial dysplasia. Visual analytics complements existing methods and fulfills a critical and at-present neglected need in the formative stages of inquiry we are facing.

Keywords: Oral Cancer, Visual Analytics, Artificial Intelligence, Low-grade Oral Dysplasia, Prevention

1 INTRODUCTION

The lack of understanding of the natural history of oral cancer is a major barrier to our ability to impactfully intervene early in the disease. As a collective group, clinicians and scientists have followed patients with clinical lesions and dysplastic disease for decades. There are unused files full of text, pictures, and annotations on these patients. In addition, as our capacity to examine biological change underlying time-varying shifts in lesions has accelerated, there is simultaneously additional, increasingly diverse information from scientists coming in. A key missing component in this effort is methods that allow us to frame and utilize such complex and heterogeneous data. They are highly multi-faceted and demand the integration of diverse clinical and scientific knowledge to generate testable hypotheses informed by the most comprehensive understanding of why lesions shift over time and when such changes may be clinically important.

Traditional statistics or the newer machine learning and artificial intelligence (henceforth ML/AI) methods are ill-suited to address many of the immediate challenges faced. The small sample sizes and complexity of clinical datasets limit the types of questions that can be answered. Additionally, these methods generally rely on well-defined and narrow questions. This is appropriate for summative analyses that aim to evaluate specific hypotheses and expectations. Current research, however, is at an exploratory stage. Instead, formative approaches that aim to understand how clinical data might be interrogated, and that support the scientific inductive process of developing, testing, and iterating over a theory are better suited. This requires the integration of expert knowledge into analysis. Data on their own do not offer explanations of why certain patterns or relationships within them exist. From the understanding of procedures involved in data gathering to theories of how observed data relate to underlying biological mechanisms driving dysplastic disease, clinical and scientific knowledge is key. Unfortunately, statistical and ML/AI methods often require significant training for interpretation and even with sufficient training often remain as difficult-to-understand “black boxes”.

We have faced this problem in British Columbia for some time. The Oral Cancer Prediction Longitudinal (OCPL) study was established over 20 years ago, to follow patients with biopsy-confirmed primary mild and moderate epithelial dysplasia (henceforth low-grade dysplasia, LGD). The presence of epithelial dysplasia is one of the strongest predictors of transformation of LGD to oral cancer; yet there are many unresolved issues around such lesions. The long-term goal of the OCPL study is to use this cohort, with its diverse data on clinical, histologic, and molecular change, and its samples, to help us answer some of the key management questions for these patients: Which of these dysplastic lesions is at risk for progression? Which do we treat, and if we treat, when and how do we do it? There are close to 600 cases in the OCPL study, many with between 10-20 years of follow-up, with over 7000 visits for these patients – a rich resource to identify and study the diverse patterns of temporal change as they occur and look for associations with transformation risk.

The question addressed in this paper is faced by all of us working in this area. How do we deal with this complex and increasingly multi-faceted data pool, especially when dealing with temporal shifts in patient data? How do we use such information to drive meaningful change – to link patterns across data sources and to generate new testable ideas? Where do we begin?

We argue for methods that support the iterative scientific process needed to integrate clinical and mechanistic knowledge. We propose that *visual analytics* (VA) is well-suited to such a niche, providing an approach that can be used to integrate “data-driven” and “knowledge-driven” processes into an iterative analysis that can improve our understanding of the natural history of oral cancer development. In this paper, we describe the challenges of heavily “data-driven” methods and why VA is well suited to complement such methods. We illustrate the value of VA by discussing a simple exploratory visual analysis of lesion shifts in oral dysplasia we conducted using data from the OCPL study.

2 CHALLENGES IN HEAVILY DATA-DRIVEN METHODS

The rapid change in computational capacity has allowed researchers to increase the volume of data analyzed and to employ sophisticated ML/AI to increasingly complex datasets, which have been inaccessible in the past. Computer vision algorithms can identify cancerous nodules from medical imaging with accuracy sometimes exceeding human experts (23). Recent preliminary research has also made headway in making these algorithms more interpretable for clinicians (7). However, state-of-the-art algorithms such as these are applied to *narrow* and *highly specific* tasks and require *large* volumes of highly constrained, well-defined data while relying on a number of assumptions about the statistical properties of these data (28) (Figure 1A).

In contrast, clinical datasets are often complex, heterogeneous, and composed of comparatively much lower volumes of patient data. In addition, patients are diverse and biological processes are ill-understood, and the understanding of how data are gathered is primarily held by clinicians. This creates an ill-defined problem space where the precise questions to ask are not yet known and thus we cannot expect a linear process of well-defined inquiry. Even if there are some well-defined questions, they have not been addressed using existing approaches and progress has been slow. This problem requires iterative and flexible generation and evaluation of practically relevant and knowledge-informed hypotheses (Figure 1B). Presently, natural intelligence is comparatively better than artificial intelligence at dealing with such challenges.

ML/AI algorithms struggle with generalization that goes beyond very constrained problem spaces; they cannot generate causal models of mechanisms underlying the data and translate them to other domains (28). Generalization involves going beyond what is explicit in data and imagining alternative potential mechanisms of explanation. Counterfactual reasoning, the imagining of alternative events and outcomes, has been the foundation of theories explaining causality (19). These theories have been integrated into methods used to analyze observational data in epidemiology in the Bradford-Hill criteria (15). While there is an effort underway to reconcile ML/AI approaches with contemporary causal inference to enable automated discovery of causal structure from data (28), such problems are still largely a human reasoning activity.

3 SENSEMAKING

Sensemaking is a “natural kind of human activity in which large amounts of information about a situation or topic are collected and deliberated upon to form an understanding that becomes the basis of problem-solving and actions” (26). This activity is often described through the data/frame theory of sensemaking which posits that humans organize knowledge and account for new information using explanatory structures called “frames” (16). As humans encounter new information through their environment, or in this case visualization systems, the information is matched and fitted to these frames. These frames are then elaborated upon, questioned, rejected, or otherwise manipulated, in our case through the interactive visualization system, in light of any new information. The scientific process of developing, testing, and iterating over theory closely mirrors sensemaking. This flexible way of thinking is what allows humans to meaningfully understand and act in a variety of natural settings such as the exploratory scientific inquiry of data.

An essential component of sensemaking is the generation of alternative hypotheses or interpretations that are flexibly fitted to and altered by data (27). This process can generate new frames of understanding based on data (data-driven), as well as iterate over existing ones (knowledge-driven) (16). This iterative fitting and manipulation of data and theory (Figure 1B) integrates human knowledge into analysis without being hampered by the limits of what is explicitly contained in the data. The relatively new field of VA specifically supports such human sensemaking activities.

4 VISUAL ANALYTICS

In scientific domains, visualization is commonly thought of as serving a purely communicative role, primarily supplementing text to emphasize a point. Yet, visualizations, especially interactive ones, can also be used to support a method of analysis. Visual Analytics, the "science of analytical reasoning facilitated by interactive visual interfaces" (6), leverages the strengths of computers to improve human analysis. The aim is to make complex computational processes transparent and empower humans to conduct analysis in an interpretable and accessible way. Rather than replacing ML/AI methods, VA complements these approaches and often integrates them in analysis. Addressing the challenges of interpretability and opening the "black box" of ML/AI algorithms has become a burgeoning area of research in VA (4) .

Visualization capitalizes on the innate intelligence of the human visual system. Using external representations as an aid is called "visual thinking" (30). The human visual system can extract complex statistical patterns from scenes while at the same time linking visual information to high-level cognitive processes. The human visual system is not one passive system, but a number of active systems that can both direct attention to important aspects of data in a bottom-up fashion as well as be directed to search for patterns in a top-down fashion (9). This interplay between bottom-up (data-driven) and top-down (knowledge-driven) processes in the visual system creates a dynamic interface between humans and data enabling iterative sensemaking processes. This interaction between prior knowledge and perception enables humans to "complete patterns" and derive meaning based on incomplete or uncertain information. The "Gestalt" school of psychology and the concomitant visual Gestalt laws describe these processes (30).

Just as sensemaking in open-ended problem spaces requires the generation and management of alternative hypotheses, VA systems are designed to support alternative visual representations of data to address these hypotheses and help steer the analysis. Some VA systems also incorporate explicit support for managing alternatives (20, 22). Others have proposed "mixed-initiative" systems that utilize machine learning and data-mining systems that integrate alternative "threads" of analysis as a central system component (21, 29).

VA may seem relatively new, but this approach has already been incorporated in a broad range of domains associated with healthcare and scientific areas. For example, VA has impacted the tracking of disease progression in electronic health records (25), clinical support for blood transfusions (12), decision making in public health (3), genomics (2), chemistry (5), and oncology (14, 24).

5 OUR COLLABORATIVE PROJECT

In this section, we illustrate how VA supports the process of generating, testing, and iterating over alternative hypotheses, using our experiences analyzing a clinical dataset of patients with LGD. We began our analysis around data collected during the examination of clinical lesions. Such assessment is a key initial point in the engagement of a clinician with the patient. It is part of the ascertainment of whether the lesion falls within the "normal" boundaries of change in a tissue, and can thus be triaged back to the community, or instead, requires further follow-up.

Lesions change over time – disappearing, re-appearing, growing in size, altering shape, and changing in texture and appearance. As such, clinical change reflects, in part, alterations occurring at the molecular, cellular, and tissue level. Increasingly there are new developments in clinical approaches and tools used in decision making around lesions. A missing component is our capacity to track changes over time and understand what observable baseline changes, in the absence of intervention, are associated with alterations in progression risk.

Time-based analyses may consider a variety of perspectives or properties of data (e.g., curve fitting, regression, or signal decomposition). When we began these studies, we had no basis to choose any particular type of analysis, and rather than over-constrain the problem-space, we chose to look at sequences which we felt could reveal a variety of patterns in the time-varying data.

Sequences are notoriously challenging for both humans and algorithms to work with (8, 10). As a preliminary step, we consulted ML experts on an appropriate approach. We employed hidden Markov models (HMMs), a set of algorithms commonly used for mining sequence patterns of biological data (32). However, the areas where such models have been particularly effective are where the volume of data is quite high, the variety of patterns is relatively low, and the problem space is also relatively constrained. Examples include sequence mining in genetics (13) or protein structure prediction (31). We discovered early on that we do not have nearly enough data for HMM. Another issue was that our clinical data are relatively complex, reflecting a variety of data-generating processes. The algorithmic output was not strong and we could not find any explanations that could account for the patterns and match existing biological understanding.

We then explored the use of interactive visualizations to analyze these sequences. While algorithmic approaches are often incorporated in VA systems to make sequences and other patterns more tractable (8), for the illustrative purposes of this paper, we will focus on a purely visual approach to highlight how visualizations enable sensemaking and hypothesis generation.

5.1 Investigating Shifts in Lesions

We conducted our analysis using simple dot plot visualizations. In Figure 2A, we provide a simplified diagrammatic version of the interactive visualizations we used in analysis to illustrate our process. We identified patterns in the data which indicated potential explanatory mechanisms (Figure 2B). This is an example of how patterns in data (data-driven) can elicit relevant knowledge and thus also influence how important patterns are perceived (knowledge-driven). Drawing on prior domain knowledge, clinical researchers on our team recognized several sequence patterns and iteratively generated alternative hypotheses that could account for such patterns.

We first identified instances where clinical lesions disappeared completely – establishing when lesions were present or absent for each patient (Figure 2). In some patients, the lesion persisted at all time points (termed “persistent lesions”). In others, the lesion disappeared and did not recur during follow-up (termed “resolved”). In some cases, the lesion disappeared early in follow-up and then ”re-emerged”. A fourth pattern showed lesions disappearing and reappearing, often multiple times, in an “unstable” fashion.

This process triggered some speculative questions around what could explain these perceived patterns. As a preliminary inquiry, we questioned the reliability of these data as they had not been used in this way before. Clinicians associated with the OCPL study went back to the data to confirm these patterns, using clinical charts, pictures, and the database. As a result of this process and dialogue, several errors in the data were identified and corrected, illustrating the value of visualization at such formative stages.

We also questioned whether shifts in “resolving” and “unstable” lesions associated with small lesion size and excision during biopsy could be confounding the lesion’s natural history. We checked. There was no apparent, consistent association with such descriptors. We explored the relationship between these patterns and patient outcomes. Virtually all of the mild or moderate lesions that progressed to severe dysplasia or cancer were persistent lesions. But what intrigued us was the observation that non-progressing lesions fell into two groups: stable, persisting lesions and unstable lesions, with lesions appearing and disappearing multiple times during follow-up. This generated a series of questions: What was causing the “unstable” phenomena, i.e., what is the underlying biology associated with such change? And did it mean anything for risk or future trajectory of patients? Does it have clinical ramifications/value?

One potential hypothesis is that “unstable“ non-progressing lesions could represent those in which protective mechanisms are actively engaged in identifying and removing damaged and genetically altered cells, those with altered signaling pathways, and dysregulated proliferation/differentiation controls. This could involve damage recognition and repair genes, for example, p53-controlled processes, that would trigger events such as senescence or apoptosis. Such changes could also involve cell-cell interactions in the tissue, the local microenvironment, and/or activity of the immune system. These protective systems could switch on and off, as abnormal clones developed and evolved in a lesion. A dysregulation of such systems would result in progression with persistence of the lesions.

The link to the immune system, is particularly attractive, given the rapid evolution of both technology in this area, especially associated with tissue change and risk prediction for cancer development. Recent findings in the esophagus, lung, and oral cavity support the possibility that the immune system is capable of recognizing premalignant lesions and intercepting their progression to cancer (1, 11, 17, 18). Premalignant-specific putative neoantigens have been identified in some such lesions and coupled to tissue infiltration of specific T effector and cytotoxic cells, for example, CD4, CD8, PD-1, and PD-L1 (17). Finally, early data support the association of alterations to antigen processing and presentation pathways and depletion of innate and adaptive immune cells with premalignant lesions that are more likely to progress. The question is, can we now use this knowledge and our current analysis systems to follow the immune system over time, and look for parallel, concordant alterations in unstable lesions that would support their involvement in temporal shifts?

6 DISCUSSION

We have only touched on a small portion of the potential analyses in the research area we have outlined. Even so, our experiences demonstrate the potential for visual analytics to generate and explore new research questions. Conventional methods used in oral oncology research have left many resources, such as complex clinical datasets or the expert knowledge of clinicians, underutilized, and many related questions unasked. It doesn’t need to be this way. Using VA allows us to cast a wider net and catch research trajectories that might otherwise remain unexplored. In the context of early detection and prevention of malignant dysplasia, leveraging the data that are already available through clinics has the potential to transform the standard of care.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

SN, MR, WS, and LB were all involved in setting the conceptual direction of this work. SN and MR wrote the first draft of this article. WS provided critical feedback and insights, and edited the manuscript.

FUNDING

Our research is supported by Simon Fraser University’s Big Data Initiative Next Big Question Fund and the British Columbia Cancer Foundation.

REFERENCES

1 .Beane, J. E., Mazzilli, S. A., Campbell, J. D., Duclos, G., Krysan, K., Moy, C., et al. (2019). enMolecular subtyping reveals immune alterations associated with progression of bronchial premalignant lesions. *Nature Communications* 10, 1856. doi:10.1038/s41467-019-09834-2

2 .Cain, A. A., Kosara, R., and Gibas, C. J. (2012). enGenoSets: Visual Analytic Methods for Comparative Genomics. *PLOS ONE* 7, e46401. doi:10.1371/journal.pone.0046401. Publisher: Public Library of Science

3 .Chishtie, J. A., Babineau, J., Bielska, I. A., Cepoiu-Martin, M., Irvine, M., Koval, A., et al. (2019). Visual Analytic Tools and Techniques in Population Health and Health Services Research: Protocol for a Scoping Review. *JMIR Research Protocols* 8. doi:10.2196/14019

4 .Choo, J. and Liu, S. (2018). Visual Analytics for Explainable Deep Learning. *IEEE Computer Graphics and Applications* 38, 84–92. doi:10.1109/MCG.2018.042731661. Conference Name: IEEE Computer Graphics and Applications

5 .Codorniu-Hernández, E., Wm. Hall, K., Ziemianowicz, D., Carpendale, S., and G. Kusalik, P. (2014). enAqueous production of oxygen atoms from hydroxyl radicals. *Physical Chemistry Chemical Physics* 16, 26094–26102. doi:10.1039/C4CP02959C. Publisher: Royal Society of Chemistry

6 .Cook, K. A. and Thomas, J. J. (2005). *Illuminating the path: The research and development agenda for visual analytics* (Pacific Northwest National Lab.(PNNL), Richland, WA (United States))

7 .Diao, J. A., Chui, W. F., Wang, J. K., Mitchell, R. N., Rao, S. K., Resnick, M. B., et al. (2020). en*Dense, high-resolution mapping of cells and tissues from pathology images for the interpretable prediction of molecular phenotypes in cancer*. preprint, Bioinformatics. doi:10.1101/2020.08.02.233197

8 .Du, F., Shneiderman, B., Plaisant, C., Malik, S., and Perer, A. (2017). Coping with Volume and Variety in Temporal Event Sequences: Strategies for Sharpening Analytic Focus. *IEEE Transactions on Visualization and Computer Graphics* 23, 1636–1649. doi:10.1109/TVCG.2016.2539960

9 .Evans, K. K., Horowitz, T. S., Howe, P., Pedersini, R., Reijnen, E., Pinto, Y., et al. (2011). enVisual attention. *WIREs Cognitive Science* 2, 503–514. doi:https://doi.org/10.1002/wcs.127. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.127

10 .Fournier-Viger, P., Lin, J. C.-W., Kiran, R. U., Koh, Y. S., and Thomas, R. (2017). A survey of sequential pattern mining. *Data Science and Pattern Recognition* 1, 54–77

11 .Foy, J.-P., Bertolus, C., Ortiz-Cuaran, S., Albaret, M.-A., Williams, W. N., Lang, W., et al. (2018). enImmunological and classical subtypes of oral premalignant lesions. *OncoImmunology* 7, e1496880. doi:10.1080/2162402X.2018.1496880

12 .Gálvez, J. A., Ahumada, L., Simpao, A. F., Lin, E. E., Bonafide, C. P., Choudhry, D., et al. (2014). Visual analytical tool for evaluation of 10-year perioperative transfusion practice at a children's hospital. *Journal of the American Medical Informatics Association* 21, 529–534. doi:10.1136/amiajnl-2013-002241

13 .Henderson, J., Salzberg, S., and Fasman, K. H. (1997). Finding genes in DNA with a hidden Markov model. *Journal of Computational Biology* 4, 127–141

14 .Horowitz, T. S. and Rensink, R. A. (2016). enExtended Vision for Oncology. In *Oncology Informatics* (Elsevier). 287–303. doi:10.1016/B978-0-12-802115-6.00015-X

15 .Höfler, M. (2005). enThe Bradford Hill considerations on causality: a counterfactual perspective. *Emerging Themes in Epidemiology* 2, 1–9. doi:10.1186/1742-7622-2-11. Number: 1 Publisher: BioMed Central

16 .Klein, G., Phillips, J., Rall, E., and Peluso, D. (2007). A data-frame theory of sensemaking. *Expertise out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making* , 113–155

17 .Krysan, K., Tran, L. M., Grimes, B. S., Fishbein, G. A., Seki, A., Gardner, B. K., et al. (2019). enThe immune contexture associates with the genomic landscape in lung adenomatous premalignancy. *Cancer Research* , canres.0153.2019doi:10.1158/0008-5472.CAN-19-0153

18 .Lagisetty, K. H., McEwen, D. P., Nancarrow, D. J., Schiebel, J. G., Ferrer-Torres, D., Ray, D., et al. (2021). enImmune determinants of Barrett's progression to esophageal adenocarcinoma. *JCI Insight* 6, e143888. doi:10.1172/jci.insight.143888

19 .Lewis, D. (2013). *Counterfactuals* (John Wiley & Sons)

20 .Liu, J., Boukhelifa, N., and Eagan, J. R. (2020). Understanding the Role of Alternatives in Data Analysis Practices. *IEEE Transactions on Visualization and Computer Graphics* 26, 66–76. doi:10.1109/TVCG.2019.2934593. Conference Name: IEEE Transactions on Visualization and Computer Graphics

21 .Makonin, S., McVeigh, D., Stuerzlinger, W., Tran, K., and Popowich, F. (2016). Mixed-Initiative for Big Data: The Intersection of Human + Visual Analytics + Prediction. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*. 1427–1436. doi:10.1109/HICSS.2016.181. ISSN: 1530-1605

22 .Mathisen, A., Horak, T., Klokmose, C. N., Grønbæk, K., and Elmqvist, N. (2019). enInsideInsights: Integrating Data-Driven Reporting in Collaborative Visual Analytics. *Computer Graphics Forum* 38, 649–661. doi:10.1111/cgf.13717

23 .McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., et al. (2020). enInternational evaluation of an AI system for breast cancer screening. *Nature* 577, 89–94. doi:10.1038/s41586-019-1799-6. Number: 7788 Publisher: Nature Publishing Group

24 .Onukwugha, E., Plaisant, C., and Shneiderman, B. (2016). Data visualization tools for investigating health services utilization among cancer patients. In *Oncology Informatics* (Elsevier). 207–229

25 .Perer, A. and Sun, J. (2012). Matrixflow: temporal network visual analytics to track symptom evolution during disease progression. In *AMIA annual symposium proceedings* (American Medical Informatics Association), vol. 2012, 716

26 .Pirolli, P. (2009). Making sense of Sensemaking in the digital World. In *European Conference on Technology Enhanced Learning* (Springer), 1–2

27 .Pirolli, P. and Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (McLean, VA, USA), vol. 5, 2–4

28 .Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., et al. (2021). Toward causal representation learning. *Proceedings of the IEEE* Publisher: IEEE

29 .Sturzlinger, W., Dwyer, T., Drucker, S., Görg, C., North, C., and Scheuermann, G. (2018). Immersive human-centered computational analytics. In *Immersive Analytics* (Springer). 139–163

30 .Ware, C. (2010). *Visual thinking for design* (Elsevier)

31 .Won, K.-J., Hamelryck, T., Prügel-Bennett, A., and Krogh, A. (2007). An evolutionary method for learning HMM structure: prediction of protein secondary structure. *BMC bioinformatics* 8, 1–13. Publisher: BioMed Central

32 .Yoon, B.-J. (2009). Hidden Markov models and their applications in biological sequence analysis. *Current genomics* 10, 402–415. Publisher: Bentham Science Publishers

FIGURE CAPTIONS

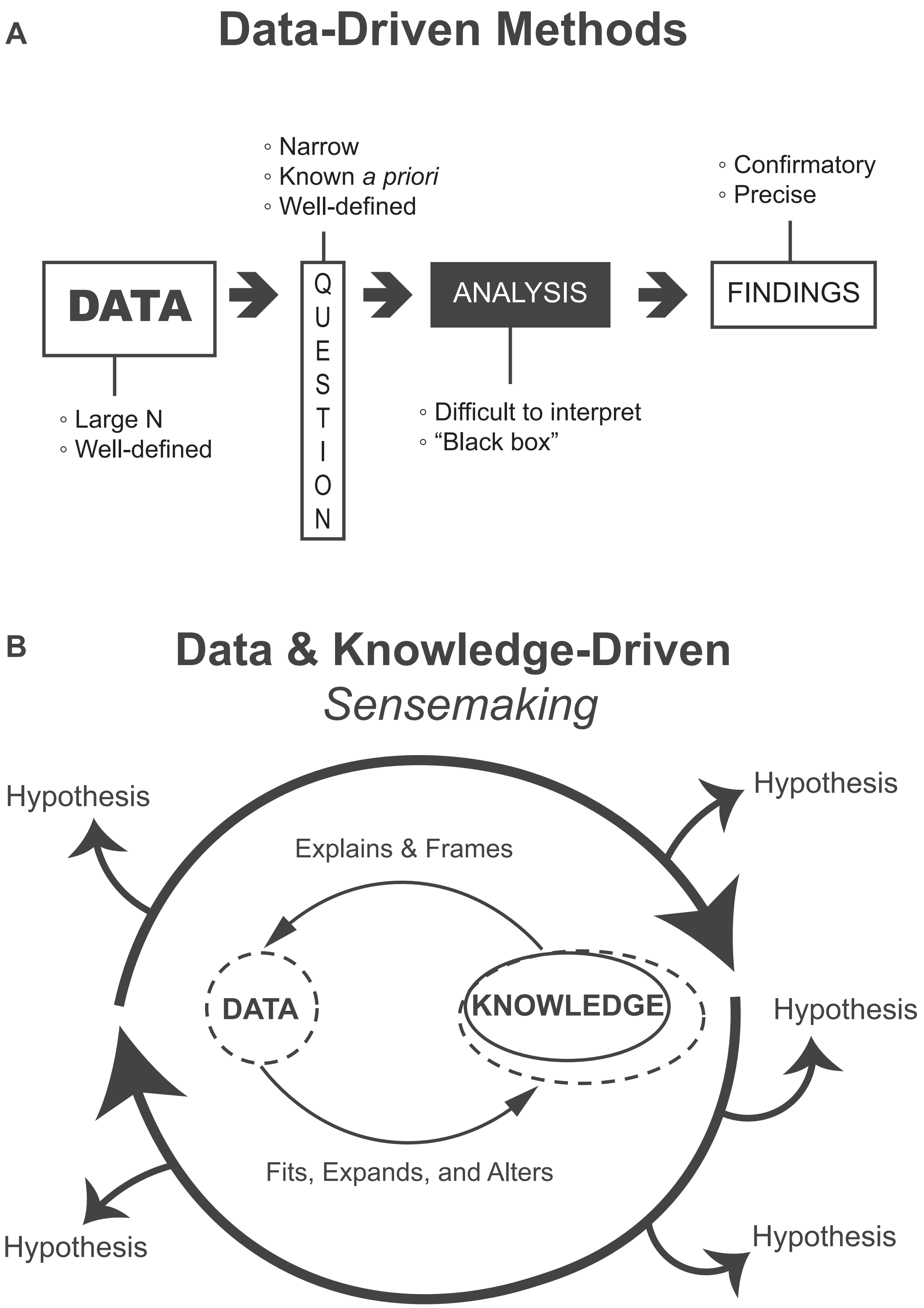

Figure 1. **(A)** Heavily data-driven methods follow a linear flow from data to findings, require voluminous data to address narrow questions that are known ahead of the analysis, and produce confirmatory and precise findings but where analyses may be difficult to interpret ”black boxes”. **(B)** Methods that support the data and knowledge-driven process of sensemaking iteratively generate, evaluate, and refine alternative hypotheses. Such methods are appropriate for exploratory and formative analyses.

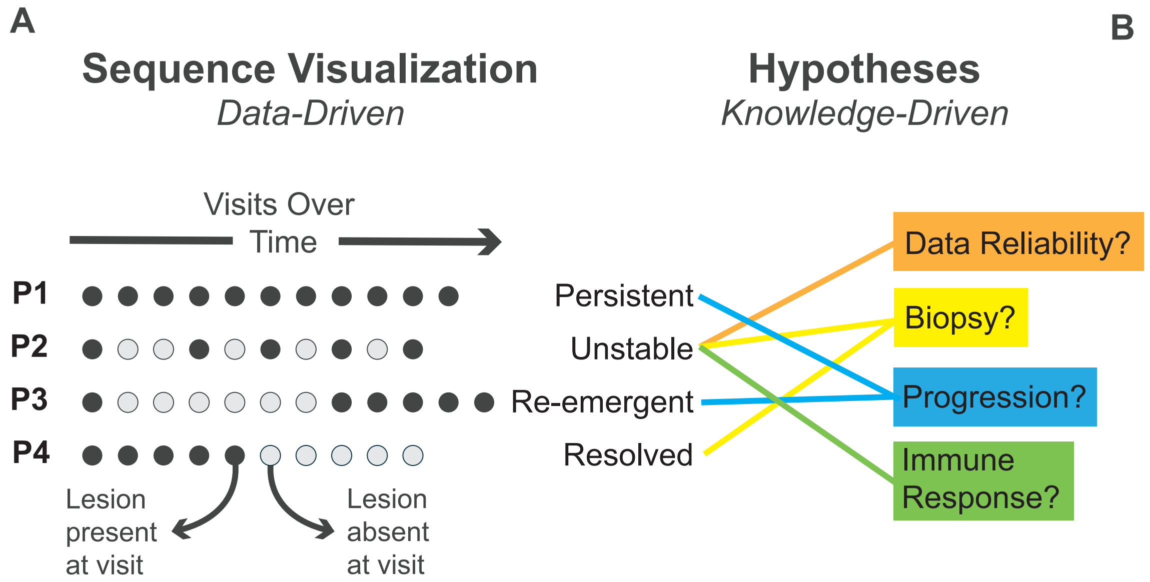

Figure 2. **(A)** Four exemplary sequence patterns in patient visits identified through visual analysis are presented. Circles represent individual visits with time moving left to right. **(B)** Several alternative explanatory mechanisms generated during visual analysis are matched to observed patterns.