Annotation Formats: Why do we need them?

COMP61332: Text Mining
Week 1
Riza Batista-Navarro

Understanding documents

Documents **rarely have a simple structure**

Examples

- news article
- research article

News article

Are you insured?

Guides for Affordable Care Act available

Staff Report

dailytimes@harrisondaily.com

Debbie Johnson a Certified Application Counselor for NARMC, assists residents who want health insurance coverage through the Affordable Care Act. She works for North Arkansas Regional Medical Center.

NARMC in partnership with the Arkansas Insurance Department's Arkansas Health Connector Division is sponsoring a sign up event for those that have questions or would like to sign up for the new Marketplace Insurance.

The Registered Nurse says that she has run across several people with serious health issues that would not receive needed care without signing up for the new insurance.

"Having insurance is a wonderful thing for a lot of people," Johnson stressed. "Some of these people have never had insurance in their life."

Residents without insurance or are interested in checking for insurance options can get information from noon - 6 p.m. Tuesday in the first floor waiting room at North Arkansas

See **HELP**, Page 13A

AFFORDABLE CARE ACT

What: Information and signup for Affordable Care Act;

When: Noon - 6 p.m. Tuesday, Feb. 18;

Where: First floor waiting room at North Arkansas Regional Medical Center.

How much will water rates rise?

By JAMES L. WHITE

jamesw@harrisondaily.com

Harrison water customers will see water/sewer rates increase by 4.2 percent this summer. But just how much will that cost customers and how do Harrison's rates compare with other cities?

Officials said the rate increase would be about 72 cents per month on the minimum water bill, but the minimum bill is $17.38 based on usage of 1,500 gallons of water/sewer.

Alderman Dan Roberts pointed out in a council meeting that many people use well more than the minimum amount of 1,500 gallons.

Using a base of 5,000 gallons used in a month, the water/sewer bill combined would be $49.07. The 4.2-percent increase would raise that monthly bill by a little more than $2.

And just how do Harrison water/sewer rates stack up to some other cities?

The Daily Times obtained water rates for Bentonville, Fayetteville and Mountain Home. Those rates show a disparity in water and sewer rates and put Harrison's rates in the middle of the pack.

For instance, Bentonville's monthly water rate for 5,000 gallons is $16.32, compared to Harrison's rate of $23.90, but sewer rates are a different story.

Bentonville's sewer rate for 5,000 gallons is $38.45, while Harrison's rate is $25.17. With the two rates combined, Harrison's full monthly bill for water and sewer is $5.72 less than in Bentonville.

Fayetteville's rate for 5,000 gallons of water is $21.74, some $2.16 less than in Harrison, but the monthly sewer rate is $35, almost $10 more than in Harrison. The combined bill in Fayetteville is more than $7.60 higher than in Harrison.

Mountain Home's water and sewer rates for 5,000 are both less than in Harrison. The combined monthly bill in Mountain

See **RATES**, Page 13A

CURRENT RATES

Water rates based on 5,000 gallons
- Bentonville — $16.32
- Mountain Home — $21.15
- Fayetteville — $21.74
- Harrison — $23.90

Sewer rates based on 5,000 gallons
- Bentonville — $38.45
- Mountain Home — $25.50
- Fayetteville — $35
- Harrison — $25.17

Combined rates
- Bentonville — $54.79
- Mountain Home — $46.65
- Fayetteville — $56.74
- Harrison — $49.07

WHAT YOU NEED TO KNOW TODAY

PRESIDENT'S DAY HOLIDAY canceled by board. **Page 3A**

DISTRICT BASKETBALL brackets set. **Page 1C**

DINNER THEATER COMING to Twelve Oaks. **Page 1D**

Printed by Ozark Printing, LLC

AREA DEATHS

Fletcher Dean

Darrell Mintle

INDEX

WEEKEND WEATHER

Partly sunny, with a high near 51. South southeast wind 5 to 15 mph.

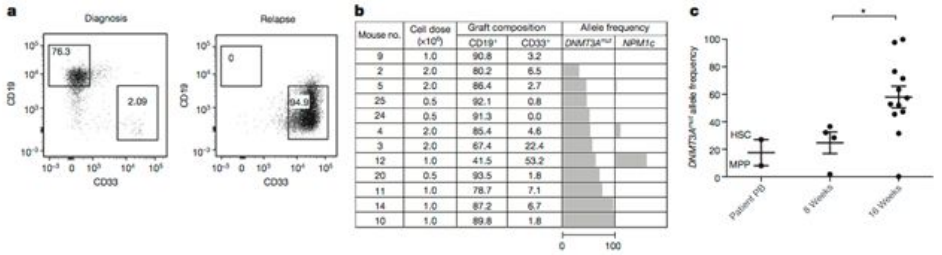
Research article

lymphoid (predominantly CD19$^+$ B cells) and myeloid (CD33$^+$) cells (Extended Data Fig. 6). For patient no. 11, multilineage engraftment was seen in 24 out of 35 mice, giving a calculated frequency of one repopulating HSC in 7.3×10^5 cells (Extended Data Fig. 5a). Only a single graft contained more than 50% CD33$^+$ myeloid cells, consistent with co-engraftment by a leukaemia stem cell (LSC) that was present at low frequency[38]. We analysed by ddPCR 12 of the multilineage xenografts following 16 weeks of repopulation. Ten of these contained a high proportion of cells bearing *DNMT3A*^{mut} without *NPM1c* (mean allele frequency 57%), whereas both *DNMT3A*^{mut} and *NPM1c* were present in the single mouse with significant myeloid engraftment (Fig. 3b). Kinetic analysis demonstrated increasing *DNMT3A*^{mut} allele frequency in multilineage grafts over time (Fig. 3c). Similar results were found for patient no. 55 (Extended Data Fig. 5 and data not shown). In contrast, cells from the relapse sample of both patients generated leukaemic grafts and no multilineage grafts (Fig. 3a and data not shown), consistent with a higher LSC frequency at relapse compared to diagnosis[38]. Together, these data provide evidence that *DNMT3A*^{mut} occurs in HSCs/

Our data predict that *DNMT3A* mutation may occur in healthy adults and pre-date AML diagnosis by months or even years. Through searches of exome sequence databases derived from peripheral blood (https://esp.gs.washington.edu/drupal/) we found that the frequency of the *DNMT3A* R882H variant (rs147001633) was 0.066% (3 in 4,545). Although this was considered to be a germline variant in this healthy adult cohort, our findings raise the possibility that the mutations detected in these studies may have originated from an HSC/MPP containing an acquired somatic *DNMT3A* mutation that underwent clonal expansion.

Discussion

Our study provides a number of key insights into the leukaemogenic process in human AML. Our findings establish the sequential order of mutation acquisition for the patients reported here: *DNMT3A*^{mut} occurs before *NPM1c* and *FLT3*-ITD. Additionally, we provide strong evidence for the presence, at diagnosis, of pre-leukaemic HSCs that are ancestral to the dominant AML clone. Based on our data, pre-leukaemic HSCs are prevalent among patients with *DNMT3A*^{mut}, which account for



Mouse no.	Cell dose ($\times 10^6$)	Graft composition CD19$^+$	Graft composition CD33$^+$	Allele frequency *DNMT3A*^{mut}	Allele frequency *NPM1c*
9	1.0	90.8	3.2		
2	2.0	80.2	6.5		
5	2.0	86.4	2.7		
25	0.5	92.1	0.8		
24	0.5	91.3	0.0		
4	2.0	85.4	4.6		
3	2.0	67.4	22.4		
12	1.0	41.5	53.2		
20	0.5	93.5	1.8		
11	1.0	78.7	7.1		
14	1.0	87.2	6.7		
10	1.0	89.8	1.8		

Figure 3 | Pre-leukaemic HSCs bearing *DNMT3A*^{mut} generate multilineage engraftment and have a competitive advantage in xenograft repopulation assays. **a**, Representative flow cytometric analysis of engrafted human cells collected from NSG mouse bone marrow (BM) 16 weeks after intrafemoral transplantation of peripheral blood mononuclear cells (PBMNCs) from diagnosis and relapse samples of patient no. 11. **b**, Analysis of human graft composition in NSG mouse BM 16 weeks after intrafemoral transplantation of PBMNCs from the diagnosis sample of patient no. 11 across a range of cells doses. The percentage of human (CD45$^+$) B (CD19$^+$) and myeloid (CD33$^+$) cells was determined by flow cytometry. Mutant allele frequency (%) in the human graft was determined by droplet digital PCR (ddPCR) analysis of sorted human cells. The length of the bars is proportional to the mutant allele frequency (the scale bar under the first column applies to all columns). **c**, Summary of *DNMT3A*^{mut} allele frequency in the human graft from mice analysed by ddPCR 8 and 16 weeks after transplantation of PBMNCs from patient no. 11, compared to isolated haematopoietic stem cells/multipotent progenitors (HSCs/MPPs) from the patient's peripheral blood at diagnosis. *$P < 0.05$. Bars indicate mean and standard deviation.

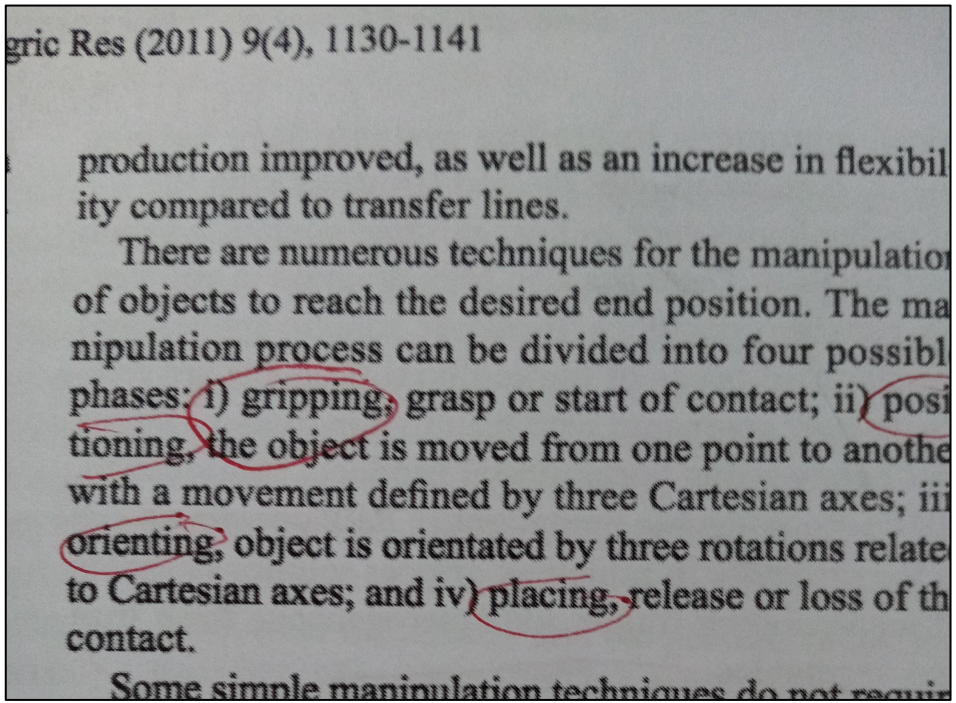
Understanding documents

Documents are meant to be **human-readable**

Easy for us but what about automated systems?

Annotations: Enabling machine-readability

Marking up documents



.... but **following a certain convention**!

Annotations: Enabling machine-readability

Added onto raw text, not original formatted document

There are numerous techniques for the manipulation of objects to reach the desired end position. The manipulation process can be divided into four possible phases: i) gripping, grasp or start of contact; ii) positioning, the object is moved from one point to another with a movement defined by three Cartesian axes; iii) orienting, object is orientated by three rotations related to Cartesian axes; and iv) placing, release or loss of the contact.

Annotations: Purposes

Presentation

4. Conclusions

We reported the isolation and structure elucidation of two novel trialkyl-substitued aromatic acids, solwaric acids A (**1**) and B (**2**), and the known 2,4,6-triphenyl-1-hexene (**3**). The novel compounds demonstrated antibacterial activity against methicillin-resistant *Staphylococcus aureus* (MRSA) and methicillin-sensitive *Staphylococcus aureus* (MSSA). Solwaric acid A (**1**) and B (**2**) were enriched with ^{13}C-labeled glucose that allowed for the acquisition of a ^{13}C–^{13}C COSY and unambiguous assignment of the methyl group location on the phenyl ring. While this example utilized ^{13}C-labeling to determine carbon connectivity for one challenging portion of the structure, this method could be valuable for molecules with multiple tetrasubstituted centers, which make structure elucidation by standard NMR experiments more challenging. Hence, ^{13}C incorporation and subsequent acquisition of a ^{13}C–^{13}C COSY—aided by the increasing sensitivity of NMR spectrometers—could drastically reduce the time for structure determination of microbial-derived natural products, including peptides and terpenes. Compared to the cost and time involved with other methods such as computer assisted structure determination, labeling microbial natural products offers a cost effective solution while providing high confidence in the proposed structure.

Annotations: Purposes

Presentation

```
−<sec sec-type="conclusions">
   <title>4. Conclusions</title>
 −<p>
    We reported the isolation and structure elucidation of two novel trialkyl-substitued aromatic acids, solwaric acids A (
    <bold>1</bold>
    ) and B (
    <bold>2</bold>
    ), and the known 2,4,6-triphenyl-1-hexene (
    <bold>3</bold>
    ). The novel compounds demonstrated antibacterial activity against methicillin-resistant
    <italic>Staphylococcus aureus</italic>
    (MRSA) and methicillin-sensitive
    <italic>Staphylococcus aureus</italic>
    (MSSA). Solwaric acid A (
    <bold>1</bold>
    ) and B (
    <bold>2</bold>
    ) were enriched with
    <sup>13</sup>
    C-labeled glucose that allowed for the acquisition of a
    <sup>13</sup>
    C–
    <sup>13</sup>
    C COSY and unambiguous assignment of the methyl group location on the phenyl ring. While this example utilized
    <sup>13</sup>
    C-labeling to determine carbon connectivity for one challenging portion of the structure, this method could be valuable for molecules with multiple tetrasubstituted centers, which make structure elucidation by standard NMR experiments more challenging. Hence,
    <sup>13</sup>
    C incorporation and subsequent acquisition of a
    <sup>13</sup>
```

Annotations: Purposes

Metadata

Mar. Drugs **2014**, *12*, 1013-1022; doi:10.3390/md12021013

OPEN ACCESS

marine drugs

ISSN 1660-3397

www.mdpi.com/journal/marinedrugs

Article

Solwaric Acids A and B, Antibacterial Aromatic Acids from a Marine *Solwaraspora* sp.

Gregory A. Ellis [1,†], Thomas P. Wyche [1,†], Charles G. Fry [2], Doug R. Braun [1] and Tim S. Bugni [1,*]

[1] Pharmaceutical Sciences Division, University of Wisconsin-Madison, 777 Highland Avenue, Madison, WI 53705, USA; E-Mails: gaellis@wisc.edu (G.A.E.); twyche@wisc.edu (T.P.W.); drbraun1@facstaff.wisc.edu (D.R.B.)

[2] Department of Chemistry, University of Wisconsin-Madison, 1101 University Avenue, Madison, WI 53706, USA; E-Mail: fry@chem.wisc.edu

[†] These authors contributed equally to this work.

[*] Author to whom correspondence should be addressed; E-Mail: tbugni@pharmacy.wisc.edu; Tel.: +1-608-263-2519; Fax: +1-608-262-5345.

Received: 19 December 2013; in revised form: 24 January 2014 / Accepted: 24 January 2014 / Published: 14 February 2014

Annotations: Purposes

Metadata

```
−<article xml:lang="en" article-type="research-article">
 −<front>
  −<journal-meta>
    <journal-id journal-id-type="publisher-id">marinedrugs</journal-id>
    <journal-title>Marine Drugs</journal-title>
    <abbrev-journal-title abbrev-type="publisher">Mar. Drugs</abbrev-journal-title>
    <abbrev-journal-title abbrev-type="pubmed">Marine Drugs</abbrev-journal-title>
    <issn pub-type="epub">1660-3397</issn>
   −<publisher>
     <publisher-name>MDPI</publisher-name>
    </publisher>
   </journal-meta>
  −<article-meta>
    <article-id pub-id-type="doi">10.3390/md12021013</article-id>
    <article-id pub-id-type="publisher-id">marinedrugs-12-01013</article-id>
   −<article-categories>
    −<subj-group>
       <subject>Article</subject>
     </subj-group>
    </article-categories>
   −<title-group>
    −<article-title>
       Solwaric Acids A and B, Antibacterial Aromatic Acids from a Marine
       <italic>Solwaraspora</italic>
       sp.
     </article-title>
    </title-group>
   −<contrib-group>
    −<contrib contrib-type="author">
```

Annotations: Purposes

Understanding content