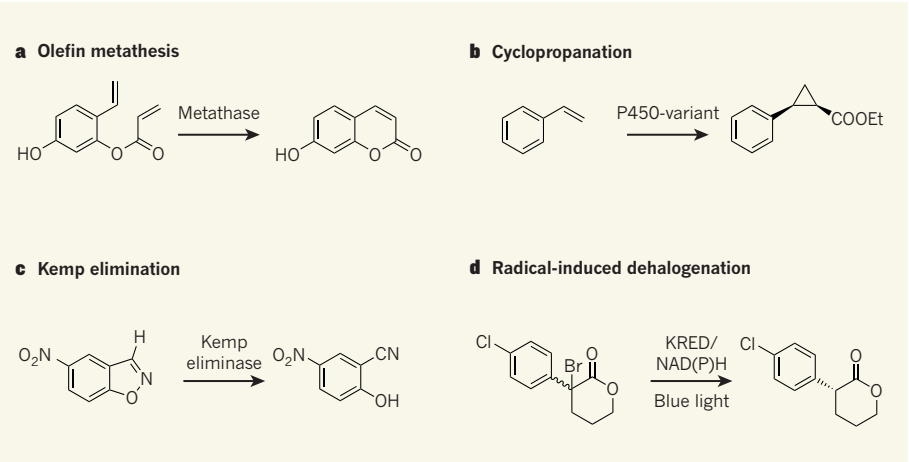


Figure 1 | Non-natural reactions catalysed by enzymes. Several strategies have been devised to develop enzymes that catalyse reactions unknown in nature. **a**, Synthetic catalysts can be incorporated into naturally occurring proteins. This approach was used to prepare metathase enzymes that catalyse olefin metathesis[6]. **b**, Enzymes can be engineered to catalyse non-natural reactions — for example, P450 enzymes have been engineered to promote the cyclopropanation reaction[7]. **c**, Kemp eliminase enzymes have been designed *de novo* computationally[8], to perform the Kemp elimination reaction. **d**, Emmanuel *et al.*[1] now report that the cofactor NAD(P)H can be excited by blue light in ketoreductase (KRED) enzymes. This enables the enzyme to catalyse an unnatural reaction known as radical-induced dehalogenation, which yields the product as predominantly one mirror-image isomer (enantiomer). Et, ethyl; Cl, chlorine; Br, bromine.

forming products predominantly as one enantiomer (Fig. 1d). Moreover, the enantiomer that is formed depends on the preference of the KRED that is used. The authors show that this unnatural reaction can be used to generate either of the enantiomers of products formed from a broad range of halolactones, demonstrating the synthetic usefulness of this approach.

The KRED fulfils two functions in this reaction. First, it ensures productive, coordinated binding of the photoexcited NAD(P)H with the halolactone in its active site. But it also recycles the spent cofactor by reacting it with isopropanol (a component of the reaction mixture), regenerating NAD(P)H. This efficient recycling enables a KRED molecule to mediate multiple catalytic cycles, as would be needed for the enzyme to be used to make gram or kilogram quantities of product for industrial applications.

Not all the KREDs investigated by the authors could catalyse the reaction; Emmanuel and colleagues found that certain point mutations in the enzyme are needed to promote the productive binding of NAD(P)H within the enzyme's scaffold. However, the catalytically active KREDs bind the halolactones perfectly, even though they do not resemble the enzymes' natural substrates. Furthermore, the authors proved that the unnatural reaction occurs only when NAD(P)H is tightly bound to KRED and is irradiated with blue light.

The authors proposed a mechanism for the reaction in which light irradiation causes an electron to be transferred between the NAD(P)H and the substrate, triggering cleavage of the substrate's carbon–halogen bond, and thus generating a radical intermediate that accepts a hydrogen atom to form the final product enantioselectively (see Fig. 3d of the paper[1]). They nicely confirm this mechanism by generating a deuterium donor from NAD(P)H *in situ* in KRED, and observing where the deuterium is incorporated into the reaction products.

Emmanuel *et al.* have demonstrated a completely new strategy for accessing unnatural enzymatic reactions by exploring the interface between photochemistry and protein science. Other synthetic transformations can be envisaged with this approach, by using light-induced changes in NAD(P)H analogues or other cofactors. For instance, the well-studied flavin cofactors (flavin adenine dinucleotide, flavin mononucleotide and their artificial analogues) could be prime candidates for investigation, because various flavin-dependent enzymes are important biological catalysts used by synthetic chemists[10]. In combination with modern tools for protein engineering[11], the authors' concept is likely to have a strong impact on the use of various enzyme classes in biocatalysis. ■

Uwe T. Bornscheuer *is in the Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, Greifswald University, 17489 Greifswald, Germany.*
e-mail: uwe.bornscheuer@uni-greifswald.de

1. Emmanuel, M. A., Greenberg, N. R., Oblinsky, D. G. & Hyster, T. K. *Nature* **540,** 414–417 (2016).
2. Hyster, T. K. & Ward, T. R. *Angew. Chem. Int. Edn* **55,** 7344–7357 (2016).
3. Renata, H., Wang, Z. J. & Arnold, F. H. *Angew. Chem. Int. Edn* **54,** 3351–3367 (2015).
4. Bornscheuer, U. T. *et al. Nature* **485,** 185–194 (2012).
5. Wilson, M. E. & Whitesides, G. M. *J. Am. Chem. Soc.* **100,** 306–307 (1978).
6. Jeschek, M. *et al. Nature* **537,** 661–665 (2016).
7. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. *Science* **339,** 307–310 (2013).
8. Röthlisberger, D. *et al. Nature* **453,** 190–195 (2008).
9. Blomberg, R. *et al. Nature* **503,** 418–421 (2013).
10. Toogood, H. S., Gardiner, J. M. & Scrutton, N. S. *ChemCatChem* **2,** 892–914 (2010).
11. Kazlauskas, R. J. & Bornscheuer, U. T. *Nature Chem. Biol.* **5,** 526–529 (2009).

CANCER

A gene-expression profile for leukaemia

Can simple genetic risk profiles be identified for complex diseases? The development of a gene-expression profile for acute myeloid leukaemia suggests that they can, and that they may improve prognosis prediction. SEE LETTER P.433

GERRIT J. SCHUURHUIS

On page 433, Ng *et al.*[1] report a tool that improves the prediction of prognoses for people who have a form of acute leukaemia. The researchers began by identifying populations of cells that exhibit key properties — collectively known as stemness — that enable the cells to initiate and sustain leukaemia. This allowed the authors to ascertain gene-expression profiles for stemness, and to use them as the basis of a scoring system for risk. The work demonstrates how gene-expression profiles can be used to enable reliable prognoses for complex diseases.

Acute myeloid leukaemia (AML) is characterized by the presence of a huge range of chromosomal and molecular aberrations. This means that there are many subgroups of people with AML who have widely different prognoses[2]. These groups are known as risk groups, and are used to determine which consolidation treatment should be given following initial chemotherapy (induction therapy). For example, transplantation of stem cells from donors is an option for patients judged to be

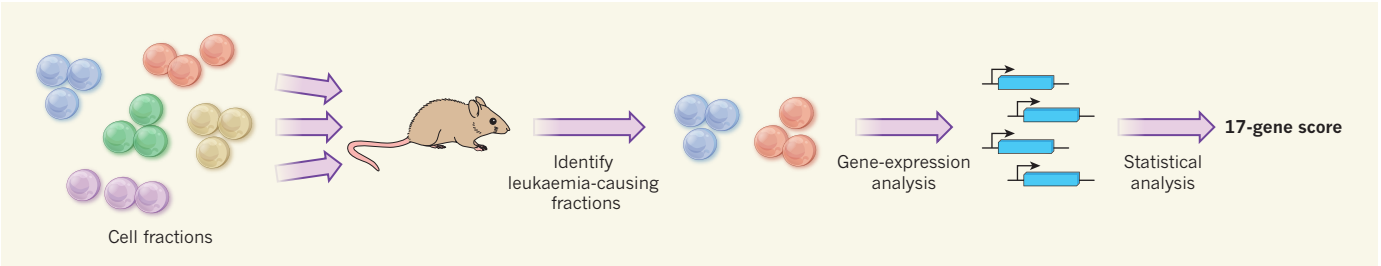


Figure 1 | A 17-gene score for assessing the risk of acute leukaemia. Ng *et al.*[1] took cell samples from people with acute myeloid leukaemia (AML) and divided them into fractions based on the expression of CD34 and CD38 proteins on the cells' surfaces. The researchers transplanted the fractions into mice, and identified which fractions caused leukaemia and which did not. The authors then compared gene-expression patterns in the disease-causing cell fractions with those in the non-disease-causing fractions, and thus identified candidate genes that correlated with tumour formation. This information was used to direct a statistical analysis of gene-expression data that had previously been gathered in a clinical study[7] of people with AML. The analysis identified a score that could be calculated for patients based on the expression of 17 genes. The score provides a reliable system for assessing patients' prognosis.

at the most risk from the disease. But such transplantation often has fatal side effects[3], so is not the best choice for some patients. Improvements to risk assessments are necessary, not only to make decisions about consolidation strategies, but also to choose between different types of induction therapy (which are expected to become available in the future).

Gene-expression profiles could be instrumental in realizing these improvements[4]. Ng and colleagues' approach, which relies on identifying profiles for stemness[5], is a good example of how such profiles could be used. In normal tissue, stemness allows stem cells to self-renew — to sustain the long-term process of normal cell differentiation. In the haematopoietic system, normal haematopoietic stem cells (HSCs) are the origin of blood cells in the circulation and the bone marrow. HSCs express CD34 proteins on their surfaces, but not CD38 proteins, and are thus said to have the $CD34^+CD38^-$ immunophenotype. Leukaemic stem cells (LSCs) have stemness properties similar to those of HSCs, but they can express different patterns of cell-surface proteins: they can be $CD34^+CD38^-$ cells (which are probably derived from HSCs), but they can also have $CD34^+CD38^+$, $CD34^-CD38^+$ or $CD34^-CD38^-$ immunophenotypes. It has previously been shown in animal models that the leukaemia-initiating ability of these different CD34/CD38 subpopulations can differ[6].

In a huge effort, Ng *et al.* isolated 227 CD34/CD38-defined cell fractions from 78 people with AML, and injected the fractions into mice (Fig. 1). They confirmed that the leukaemia-initiating ability of the cell fractions differed: leukaemia could form from all the cell fractions obtained from a patient, from some of the fractions or from none. The authors then compared gene expression in the original cell fractions that caused leukaemia with gene expression in cell fractions that did not, irrespective of the cells' CD34/CD38 immunophenotype. This allowed them to identify gene-expression patterns that were directly related to the ability of cells to form leukaemias *in vivo* in mice.

Ng and colleagues first identified 104 genes for which expression levels differed by at least twofold in leukaemia-initiating cell fractions compared with fractions that didn't initiate leukaemia. The authors then examined a large set of gene-expression data obtained from a clinical study[7] of 495 people with AML, and found that 89 of the 104 genes were present in the set. The cells in that study were not divided into fractions, but displayed gene-expression patterns that were similar to those observed by Ng *et al.* in leukaemia-initiating cell fractions.

Next, the authors used a statistical method to relate gene expression to clinical outcome for these 89 genes, and for a subset of 43 genes that are highly expressed in leukaemia-initiating cell fractions. This allowed them to identify an optimal panel of 17 genes, the expression of which was highly indicative of a poor clinical outcome in a patient subgroup. The authors confirmed this finding in other AML cohorts and found that a scoring system based on their gene panel (called LSC17) offered superior prognoses when compared with other gene-expression profiling systems for AML[5,8]. In fact, Ng and colleagues found that previously reported genetic signatures of AML were not independent prognostic factors when tested in the other cohorts.

Ng *et al.* also found that gene-expression patterns associated with stemness in AML are independent of the chromosomal and molecular aberrations used to assess patient risk, showing that stemness is a factor that crosses the borders of previously identified risk groups. Finally, the authors developed an assay that allows gene-expression data to be rapidly generated, which could form the basis of a fast (24–48 hours) prognostic test for patients.

As the authors indicate, analysis of large data sets from clinical studies in which both extensive information about the mutational status of leukaemia cells[9] and LSC17 scores are available will be needed to assess whether the prognostic value of the LSC17 score is independent of the prognostic value of mutations present at diagnosis. The clinical benefits of the LSC17 score must be assessed, because prognostic value does not always lead to a meaningful clinical advantage. Moreover, small populations of leukaemia cells that have a similar genetic make-up (clones) can be present at diagnosis, survive therapy and proliferate to cause a relapse (in some cases after having acquired additional mutations[10,11]). Only time will tell whether Ng and co-workers' gene-expression profiles account for cell fractions defined by such clones, and thereby predict associated relapses.

The prognosis of a person with leukaemia at the point of diagnosis is only part of the prognostic story. Once treatment has started, factors such as therapy compliance, alterations to drug doses that are made to mitigate side effects, and differences between patients in the concentration of drugs in blood plasma might partly override the effects of prognostic diagnosis parameters such as gene-expression patterns. The consideration of post-treatment parameters such as measurable (minimal) residual disease[12] (a measure of the persistence of small numbers of leukaemia cells in patients in remission) has drastically changed the landscape of risk assessment in AML. Assessing combinations of cellular properties at diagnosis, non-cellular patient-specific factors during therapy, frequencies and properties of cells that remain after treatment and changes in immunological parameters, might offer a more-refined prognosis than is currently possible. This would enable more-personalized induction and consolidation treatments to be used. Ng and colleagues' study is potentially a big step towards such assessments, especially at the diagnosis stage. ■

Gerrit J. Schuurhuis *is in the Department of Hematology, VU University Medical Center, De Boelelaan 1117, 1081HV Amsterdam, the Netherlands.*
e-mail: gj.schuurhuis@vumc.nl

1. Ng, S. W. K. *et al. Nature* **540,** 433–437 (2016).
2. Grimwade, D. *et al. Blood* **116,** 354–365 (2010).
3. Cornelissen, J. J. *et al. Nature Rev. Clin. Oncol.* **9,** 579–590 (2012).
4. Shivarov, V. & Bullinger, L. *Exp. Hematol.* **42,** 651–660 (2014).
5. Levine, J. H. *et al. Cell* **162,** 184–197 (2015).

6. Sarry, J.-E. *et al. J. Clin. Invest.* **121,** 384–395 (2011).
7. Verhaak, R. G. W. *et al. Haematologica* **94,** 131–134 (2009).
8. Gentles, A. J. *et al. J. Am. Med. Assoc.* **304,** 2706–2715 (2010).
9. Papaemmanuil, E. *et al. N. Engl. J. Med.* **374,** 2209–2221 (2016).
10. Ding, L. *et al. Nature* **481,** 506–510 (2012).
11. Jan, M. & Majeti, R. *Oncogene* **32,** 135–140 (2013).
12. Hokland P. *et al. Semin. Hematol.* **52,** 184–192 (2015).

This article was published online on 7 December 2016.

HYDROLOGY

The dynamics of Earth's surface water

High-resolution satellite mapping of Earth's surface water during the past 32 years reveals changes in the planet's water systems, including the influence of natural cycles and human activities. SEE LETTER P.418

DAI YAMAZAKI & MARK A. TRIGG

Everyone appreciates that the water cycle can vary, and can cause floods and droughts at its extremes. On page 418, Pekel *et al.*[1] map the full range of this variability, as evidenced by our rivers, lakes and wetlands, using more than 3 million satellite images collected over the past 32 years. This globally consistent analysis documents both natural water variability and humanity's major influence on Earth's water systems, and will provide a valuable baseline for observations of the effects of future climate change.

Detailed maps describing the location and extent of rivers, lakes and wetlands are needed for many studies of Earth science, but the full global distribution and variability of these systems has not been clearly understood. Scientists have developed methods to map water bodies using satellite observations — for example, by detecting the characteristic reflectance of sunlight from water. But this is a particularly challenging task because the colour of water varies greatly depending on depth, the presence of suspended sediments and dissolved chemicals, and the angle at which sunlight hits the surface. In addition, some land surfaces (such as snow, ice and lava) have similar reflectance characteristics to those of water, which means that water-detection algorithms need to be developed and calibrated carefully.

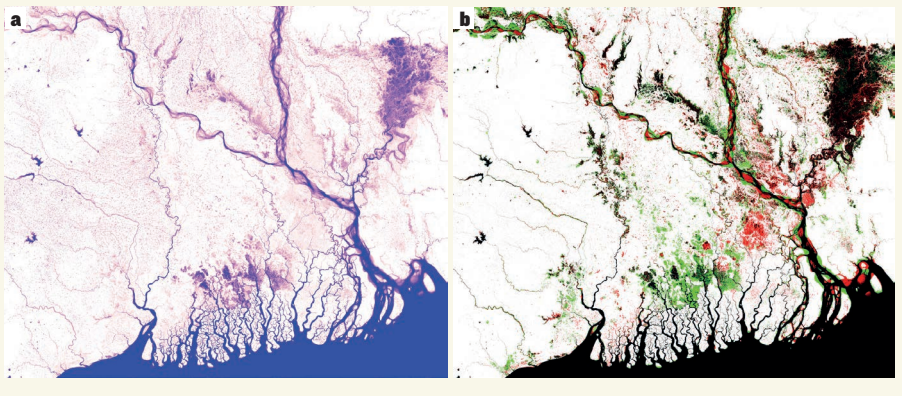
The first global surface-water map[2] made using satellite observations was developed in 2009, but computational power restricted the spatial resolution to 250 metres, which is too low to enable detailed mapping of smaller lakes and rivers. This was a problem, because statistical estimates[3] suggest that millions of lakes less than 1 square kilometre in size could account for about 40% of the global area of inland water. The situation has since improved: a global analysis of water bodies at 30-metre resolution was undertaken recently[4,5] using images from the Landsat programme (the world's longest-running initiative for acquiring satellite images of Earth).

However, the location and extent of water bodies can change with time, in part because of natural processes such as flooding, sedimentation and channel migration, but also because of human processes such as dam construction and water abstraction. This creates a need for a global-scale, high-resolution analysis of information taken at different times — a complete map of surface-water dynamics. Such dynamics have recently been captured in maps that enable scientists to distinguish permanent rivers and lakes from seasonal water bodies such as flood plains[6] and to explore the long-term trends of surface-water changes[7], but these studies used only a subset of all the Landsat images available.

Pekel and colleagues' ambitious work uses the entire Landsat archive[8] to map global surface waters — more than 3 million images collected between 1984 and 2015. To handle this petabyte-scale data set, the authors used Google Earth Engine (go.nature.com/2fdt80k), a freely available cloud-computing platform for analysing big data sets of satellite observations. The Landsat data set was produced using three satellites, and multiple operational issues affected the collection and quality of the data. This presented unique challenges, in addition to those associated with water's variable reflective properties. To overcome these challenges, Pekel *et al.* used a combination of expert systems (computer systems that use artificial intelligence) and visual analytics to identify the existence or absence of surface water for every pixel of Earth imagery, each representing a square of side 30 metres; this was done at monthly intervals over the 32-year period.

An understanding of the frequency with which water occurs at different locations is certainly a useful result of such an analysis. However, more-meaningful information and visualization of global-scale changes are required to cope with gaps in the data set that result from cloud cover and operational deficiencies, and to allow specific interpretation of different surface-water dynamics such as seasonal cycle and long-term trend. Pekel and colleagues therefore provide thematic maps depicting persistence (whether water is always present, or just sometimes), gains versus losses, the consistency of seasonal cycles, permanent versus seasonal water, and transitions between seasonal and permanent water during the period analysed (Fig. 1). The output of the analysis and the thematic maps are available through a user-friendly interface (go.nature.com/2gj81ap), allowing anyone to explore any location and understand what surface-water changes have occurred, without the need for complex analysis or massive computing power.

The authors' high-quality analyses and visualizations of the data reveal that there were 2.78 million km^2 of permanent



J.-F. PEKEL *ET AL.*/REF. 1

Figure 1 | Maps showing variability of surface water in the Ganges delta. Pekel *et al.*[1] have used historical satellite images to produce global maps that depict changes of surface water over the past 32 years. The maps are presented in different ways to enable different information to be visualized. **a**, This map shows the average water-occurrence frequency over 32 years; blue represents water that is always there, pink is water that is sometimes there. **b**, Here, red regions indicate where water occurrence has decreased during the period studied, whereas green indicates increased occurrence. These maps, along with others that depict seasonal variations, help to distinguish different causes of water dynamics, such as seasonal inundation, channel migration and reservoir construction.