Kent Academic Repository
Full text document (pdf)

Citation for published version

DOI

Link to record in KAR

https://kar.kent.ac.uk/77329/

Document Version

Author's Accepted Manuscript

High content imaging of unbiased chemical perturbations reveals that the phenotypic plasticity of the actin cytoskeleton is constrained

Nicole S. Bryce[1], Tim W. Failes[2], Justine R. Stehn[1], Karen Baker[3], Stefan Zahler[4], Yulia Arzhaeva[5], Leanne Bischof[5], Ciaran Lyons[1], Irina Dedova[1], Greg M. Arndt[2], Katharina Gaus[1,6,7], Benjamin T. Goult[3], Edna C. Hardeman[1], Peter W. Gunning[1@] and John G. Lock[1#@].

1. School of Medical Sciences, UNSW Sydney, NSW 2052, Australia
2. ACRF Drug Discovery Centre, Children's Cancer Institute, Lowy Cancer Research Centre, UNSW Sydney, NSW 2052, Australia
3. School of Biosciences, University of Kent, Canterbury, Kent, CT2 7NJ, UK
4. Department of Pharmacy, Ludwig-Maximilians-University, Munich, Germany
5. Quantitative Imaging Research Team, CSIRO Data 61, Marsfield, NSW, Australia
6. EMBL Australia Node in Single Molecule Science, UNSW Sydney, NSW 2052, Australia
7. ARC Centre of Excellence in Advanced Molecular Imaging, UNSW Sydney, NSW 2052, Australia

Lead contact

john.lock@unsw.edu.au

@ Denotes equal contribution

Summary

Although F-actin has a large number of binding partners and regulators, the number of phenotypic states available to the actin cytoskeleton is unknown. Here, we quantified 74 features defining F-actin and cellular morphology in >25 million cells after treatment with a library of 114,400 structurally diverse compounds. After reducing the dimensionality of these data, only ~25 recurrent F-actin phenotypes emerged, each defined by distinct quantitative features that could be machine learned. We identified 2003 unknown compounds as inducers of actin-related phenotypes, including two that directly bind the focal adhesion protein talin. Moreover, we observed that compounds with distinct molecular mechanisms could induce equivalent phenotypes and that initially divergent cellular responses could converge over time. These findings suggest a conceptual parallel between the actin cytoskeleton and gene regulatory networks; where the theoretical plasticity of interactions is nearly infinite, yet phenotypes *in vivo* are constrained into a limited subset of practicable configurations.

Introduction

Our current knowledge of the actin cytoskeleton and its regulatory system – incorporating 6784 proteins with actin-related functionality (> 1/3 of the annotated human proteome, based on 363 distinct gene ontology labels; http://geneontology.org) – suggests no clear limits to the organisational plasticity of this complex system. Plasticity in actin organisation physically manifests in the numerous distinct forms and combinations of sub-cellular filamentous actin (F-actin) structure (stress fibres, lamellipodia, filopodia and invadopodia etc) (Burridge and Wittchen, 2013; Lehtimaki et al., 2017; Leijnse et al., 2015; Skau and Waterman, 2015). From a regulatory perspective, the multiplicity of direct actin-binding partners, distal regulatory proteins and impinging signalling pathways that collectively modulate actin organisation (Pollard, 2016) also suggest a high level of organisational diversity. Taken together, this information suggests the potential for virtually unlimited phenotypic plasticity in actin organisation at the cellular scale.

Considering the possibility of such wide-ranging plasticity, the diversity of functional demands on the actin cytoskeleton highlight a significant challenge for the evolution and adaptation of this system. How are changes amongst the repertoire of actin-binding and actin-regulatory proteins accommodated without widespread disruption of the cytoskeleton and its numerous parallel functions? A potential clue is provided by the dynamics of a contrasting class of system: gene regulatory networks. Specifically, it has been observed that diverse patterns of genetic mutations (Huang et al., 2009) or mechanistically diverse chemical perturbations (Zhou et al., 2016) may be channelled into strongly convergent patterns of gene expression, suggesting the existence of high-dimensional attractor states (Huang et al., 2005) within the landscape of potential gene regulatory network configurations. Thus, the virtually infinite theoretical plasticity of gene regulatory network states (implying evolutionary instability) is likely constrained into a much more limited subset of practicable configurations. Recent theoretical works now indicate the potential for attractor states to effectively constrain the plasticity of protein signalling and interaction networks (Huang et al., 2017).

61 Here, we have adopted a chemical biology approach to estimate the number of ways in which cells

62 can (re)organize their F-actin cytoskeleton given a large spectrum of perturbations (Bryce et al., 2019).

63 Cells with well-defined, quantifiable actin phenotypes were challenged with 114,400 structurally

64 diverse compounds, previously demonstrated to affect a broad range of cell biological processes. In

65 contrast to both the large number of compounds and large number of actin regulators, we observed

66 a low number of emergent actin phenotypes, each with distinct quantitative features that can be

67 machine-learned. Subsequent selective mechanistic analyses, coupled with temporal analyses of

68 progressive perturbation effects, support the idea that the actin cytoskeleton contains phenotypic

69 attractor states that would be conceptually analogous to those proposed for gene regulatory

70 networks. Taken together, our findings provide an estimate of the degree of organisational plasticity

71 available to the F-actin cytoskeleton, indicating this to be relatively constrained. We suggest that this

72 constraint on plasticity may reflect an emergent process of canalization favouring evolutionarily

73 competent phenotypes and functions within this pivotal biological system.

74 ## Results

75 **Using diverse chemical perturbations to assess the degree of plasticity in F-actin organisation**

76 We challenged a total of 25,619,680 cells using the 114,400 compound WECC structural diversity

77 library (Baell, 2013; Baell and Holloway, 2010), which was established to efficiently sample the

78 chemical space of lead-like compounds with minimal structural similarity (< 85-90%). This approach

79 allowed us to estimate the number of ways in which cells can (re)organize their F-actin cytoskeleton

80 in response to chemical perturbation. This constituted 21 separate biological experiments including

81 more than 320 individual experimental repeats (fifteen to twenty 384-well plates per experiment),

82 each of which included 3 positive and 1 negative control, with these controls replicated 8-fold each

83 per plate. Including both unknown (114,400) and replicated control treatments (>10,000), a total of

84 124,767 experimental conditions were assessed. To read out phenotypic responses we employed

85 quantitative imaging and statistical analyses, with an overview of this approach summarized in Figure

4

86   1A. As a cell model, SK-N-SH neuroblastoma cells were utilized due to their large, flat morphology and

87   the relative homogeneity of their actin organisation phenotype, which makes drug-induced changes

88   more clearly discernible.

89   **Quantifying F-actin organisation**

90   To observe actin phenotypes, F-actin (phalloidin) and nuclear (Hoechst) markers were assessed via

91   automated spinning-disc confocal imaging (Figure 1B, upper panels). Visual inspection confirmed that

92   compounds from the diversity library induced a variety of distinct actin organisation phenotypes

93   (Figure 2A).

94   Image segmentation (Figure 1B) and analysis extracted 75 quantitative features defining cell

95   morphology, texture as well as properties of actin stress fibres and actin puncta. Measurements from

96   wells with extreme cell numbers or duplicated values were excluded (<10 & >700 cells; 124,343

97   conditions retained; ~200 cells average per condition), removing cytotoxic conditions or overgrown

98   wells. Z-score normalisation of data was performed for each of 74 features (1 excluded due to low

99   variance; features defined in Table S1) relative to all unknown compounds (Birmingham et al., 2009)

100   grouped per experimental batch. This greatly improved data superimposition (Figure S1A, B) and

101   revealed a conserved phenotypic signature across the 74 normalized features (Figure S1C, D) as well

102   as tight alignment of individual feature value distributions across experimental batches (Figure S1E-

103   H).

104   **Estimating F-actin organization phenotype number**

105   Next, we estimated the number of distinct and recurrent phenotypic 'clusters' that exist within the

106   complete chemical treatment dataset. To ensure the robustness of this estimation, we combined four

107   complementary methods into a "quadrangulation" strategy (Figure 1C) that progressively narrowed

108   the estimation range to a final approximation of phenotype number. Starting with the possibility that

109   the number of phenotypes may range from two or three up to thousands, method 1 of our

110 quadrangulation strategy involved visual inspection of tens of thousands of images spanning the

111 complete dataset, producing a human expert-guided range estimate for the number of phenotypes

112 that were readily discernible (~15 to ~40). This range estimate was then complimented by method 2;

113 iterative, sub-sampling-based multivariate statistical estimation of phenotypic cluster numbers using

114 hierarchical clustering interpreted via two alternative cluster-detection indices (Dunn's index or SD

115 index) (Figure S2A). Outcomes based on both indices suggested phenotype cluster numbers ranging

116 between ~17 and ~35, supporting the range estimate from visual inspection.

117 The range estimated above was next used to initialise parameterization of dimension reduction and

118 unsupervised clustering approaches for more precise estimations of phenotype number. As detailed

119 below, these approaches were combined to underpin methods 3 and 4 of the quadrangulation

120 strategy introduced above. Dimension reduction was first performed via t-distributed stochastic

121 neighbour embedding (t-SNE) (Van der Maaten and Hinton, 2008) after principle components analysis

122 (PCA) was used to reduce the computational complexity of the data. t-SNE enabled projection of high-

123 dimensional quantitative phenotype observations into a 2D space according to their local similarity

124 (Figure 2B), with unsupervised cluster detection then applied via the OPTICS algorithm to objectively

125 define distinct and recurrent organisational states of F-actin. This revealed an estimate of just 25

126 distinct phenotypes (Figure 2B, phenotypes colour-coded). As schematised in Figure 1C, this result

127 followed extensive exploration of t-SNE tuning parameter (i.e. 'perplexity') values whilst monitoring

128 the trade-off between phenotype number (quadrangulation method 3) (Figure 2C, upper) and positive

129 control 'self-clustering' (quadrangulation method 4) (i.e. representation of homogeneous positive

130 control drug treatments as forming only one major phenotype each; Figure 2C, lower). To further

131 validate this surprisingly low level of F-actin phenotype plasticity, we repeated our entire dimension

132 reduction analysis using the uniform manifold approximation and projection (UMAP) algorithm

133 (McInnes et al., 2018). This produced a recognisably similar projection of the dataset, with OPTICS

134 cluster detection suggesting just 16 distinct phenotypes (Figure 2D, phenotypes colour-coded). Again,

135 this final UMAP projection and phenotype number estimation followed extensive optimisation of the

136   UMAP tuning parameter (i.e. 'nearest neighbour') guided by the trade-off between phenotype

137   number (quadrangulation method 3) (Figure 2E, upper) and positive control 'self-clustering'

138   (quadrangulation method 4) (Figure 2E, lower). The lower phenotype number estimate from UMAP

139   corresponds with recent comparisons of t-SNE and UMAP dimension reduction methods, wherein

140   UMAP tended to identify fewer discernible clusters than t-SNE (Becht et al., 2018). This in part reflects

141   the preservation by UMAP of medium- and long-distance spatial relationships and trajectory / lineage

142   structures that are expected in datasets based on, for example, RNAseq analyses of cell differentiation.

143   Such datasets comport with the progressive manifold assumption (Moon et al., 2018), whereas such

144   progressive manifolds (i.e. trajectories) are not implied by our endpoint analysis of chemical

145   perturbation effects. Hence, we do not here draw interpretations from the distances between

146   phenotype clusters (in either t-SNE or UMAP projections), instead only assessing phenotype number.

147   For this reason, and because t-SNE more accurately preserves local similarities in phenotype (Becht et

148   al., 2018) (by not compromising local *versus* long-distance similarity relationships - thus more faithfully

149   identifying co-clustered compounds that may have similar effects), we continued our analyses using

150   the optimised data projection based on t-SNE dimension reduction.

151   Pairs of images depicting selected cellular phenotypes (Figure 2F) confirm phenotypic homogeneity

152   within each cluster. Individual observations that did not fall into a defined phenotype were excluded

153   from further analysis (1,235 observations excluded; 123,108 observations retained; Figure S2B). To

154   determine which of the 25 t-SNE-estimated phenotypes are spontaneous within the SK-N-SH cell

155   population, i.e. naturally occurring, and which are *bona fide* drug-induced phenotypes, we assessed

156   the frequency of DMSO negative control observations within each phenotype relative to the expected

157   random rate (Figure S2C). This indicated 5 spontaneous and 20 chemically induced phenotypes (Figure

158   S2D). For clarity, the 20 induced phenotypes are shown in Figure S2E.

159   We next sought to characterize the major differences between defined phenotypic clusters. To this

160   end, we first performed machine learning-based forward feature selection using all 74 Z-normalized

161   features, thereby delineating a subset of 15 features that optimally predict phenotype for each

162   observation. The >98% accuracy of the resulting model supports the significance of the 15 feature-

163   subset. Comparison of phenotypes based on these key features reveals a recurring signature that

164   characterizes similarities between most phenotypes while also highlighting specific features that

165   differentiate each (Figure 2G).

166   Visible similarities between cell images within phenotype clusters confirm the efficacy of our

167   clustering strategy. Given that similar image data are clustered by our analysis approach, two key

168   questions remain: **1)** do distinct molecular mechanisms reproducibly generate distinctive F-actin

169   phenotypes? And **2)** do individual phenotypes reflect singular molecular mechanisms?

170   **Known perturbation mechanisms induce reproducible and distinguishable phenotypes**

171   To address question 1, we assessed in detail the clustering of three positive control compounds

172   (jasplakinolide, latrunculin A, TR100) with mechanisms of action towards F-actin that are known and

173   also distinct. Jasplakinolide binds to and stabilizes actin filaments (Bubb et al., 1994), thereby

174   increasing F-actin within the cell relative to the DMSO control (Figure S3A, B). Latrunculin A binds actin

175   monomers, thus inhibiting filament assembly (Spector et al., 1983) and reducing F-actin within the cell

176   (Figure S3C). TR100 binds to tropomyosin isoform Tpm3.1 dimers and is incorporated into actin

177   filaments, thereby accelerating filament depolymerization and reducing organization into stress fibres

178   (Bonello et al., 2016; Janco et al., 2019; Stehn et al., 2013) (Figure S3D). The different mechanisms of

179   these positive control compounds are known to induce F-actin organisation phenotypes that are

180   reproducible, homogeneous and distinct. As such, we expected these conditions to not only self-

181   cluster, but also to cluster separately from each other.

182   As predicted, each of the embedded positive control drugs overwhelmingly clusters within a single

183   distinct phenotype (Figure 3A, shown individually in Figure S3F-I). Specifically, Jasplakinolide,

184   Latrunculin and TR100 induce phenotypes in 91%, 96% and 66% of cases (Figure S3E), and of these

185 induced phenotypes, Jasplakinolide, Latrunculin A and TR100 recurrently fall into specific F-actin

186 phenotypes 3, 4 and 5 in 89%, 95% and 87% of cases, respectively (Figure 3B). This highlights the

187 experimental and analytical reproducibility achieved throughout this study, wherein more than 320

188 individual experimental replicates (i.e. 384-well plates) were processed across 21 independent

189 biological experiments. In contrast to the positive controls, DMSO is broadly distributed, as expected

190 (Figure 3A; Figure S3F). Taken together, these results validate the robustness and sensitivity of the

191 experimental and analytical methods applied herein, since each positive control compound

192 overwhelmingly clustered within a single phenotype (as optimised via the 'quadrangulation' strategy,

193 Figure 1C), and these phenotypes are clearly distinguishable. Moreover, comparison of key feature

194 values between these control compounds and phenotype cluster 1 (Figure S3J, the main spontaneous

195 phenotype; large grey cluster in Figure 2B) confirms the expected effects of these control compounds.

196 For instance, latrunculin A reduces actin filament length and intensity whilst increasing actin spot

197 intensity and size (Figure S3J), as observable in Figure S3C. In contrast, TR100 substantially reduces

198 cell area, whilst maintaining long peripheral actin fibres (Figure S3D, J).

199 **Defining the properties of key phenotypic clusters**

200 In total, 1.77% (2003) of unknown compounds were associated with induced phenotypes (Figure S3E),

201 and these were distributed across all 20 induced phenotype clusters (Figure 3C), implying diverse

202 effects and a wide variety of molecular mechanisms. We next focused on understanding the biological

203 differences between key phenotypic clusters. Specifically, we compare phenotype cluster 4

204 (Latrunculin-like phenotype), 5 (TR100-like phenotype) and 9 (the largest unknown induced cluster)

205 wherein we identified significant numbers of co-clustered unknown compounds (cluster 4, 87

206 compounds; cluster 5, 55 compounds; cluster 9, 1166 compounds) (Figure 3C, D). Cell image examples

207 for unknown compounds from each of these phenotype clusters are depicted in Figure 2F.

208 Differences between biological states (here induced by chemical perturbations) can be defined not

209 only by observable changes in phenotypic features, but also by more complex (and less readily

210   discernible) changes in the statistical linkages between these features, which may reflect changes in

211   biological dependencies. To explore such changes in inter-feature linkages, we assessed asymmetries

212   in the mutual information shared between each possible pairing of the 15 machine learning-selected

213   features, using an estimation of the uncertainty coefficient (Figure 3E). This revealed large differences

214   in the number of inter-feature linkages (above a constant minimal threshold) detected in clusters 1

215   and 9 (low connectivity) and clusters 4 and 5 (high connectivity). Comparisons of connectivity changes

216   induced from cluster 1 to clusters 4, 5 and 9 showed that, despite having distinct patterns of

217   phenotypic effect, clusters 4 and 5 reflected extensive and similar alterations in inter-feature

218   connectivity. In contrast, despite a pronounced phenotype, cluster 9 reveals limited effects on inter-

219   feature connectivity.

220   **Do individual phenotypes reflect singular molecular mechanisms?**

221   An important implication of the small number of recurrent phenotypes detected in this study – relative

222   to 114,400 diverse chemical perturbations – is that multiple underlying molecular mechanisms may

223   result in the same phenotypic end-state. To this end, we performed secondary mechanistic analyses

224   of selected unknown compounds drawn from clusters 4, 5 and 9.

225   We first selected the unknown compounds producing latrunculin A-like phenotypes shown in Figure

226   2F (cluster 04), and assessed their mechanistic activities based on two key assays. First, we monitored

227   reductions in cellular F-actin levels using a G:F actin ratio assay (Figure 4A). Mimicking the effects of

228   latrunculin A, unknown compounds L2 and L3 reduced F-actin levels. In contrast, and despite

229   producing a quantitatively equivalent phenotypic end-state, unknown compound L1 had no effect on

230   the G:F actin ratio. To further assess whether L2 and L3 precisely mimic the mechanisms by which

231   Latrunculin impacts the G:F actin ratio, i.e. by inhibiting F-actin assembly, we performed an *in vitro*

232   pyrene actin filament assembly assay to measure F-actin assembly rates (Figure 4B). Neither L2 nor L3

233   directly inhibited F-actin assembly, indicating differences in their precise mechanism of action relative

234   to Latrunculin, again despite their phenotypic similarity. Thus, at least three different mechanisms

235   (latrunculin A, L1, L2/L3) can induce the singular phenotypic end-state defined by cluster 4.

236   We next selected two compounds from the TR100-like phenotypic cluster and used

237   immunofluorescence imaging to determine if Tpm3.1 was removed from F-actin stress fibres. Relative

238   to DMSO-treated control cells (Figure 4C), Tpm3.1 was removed from stress fibres by the compound

239   designated Tr1 (Figure 4D), thus mimicking the effects of TR100. In contrast, the compound designated

240   Tr2 induced a similar actin phenotype without displacing Tpm3.1 from actin filaments (Figure 4E).

241   Thus, once again, equivalent actin phenotypes emerge as a result of distinct underlying mechanisms.

242   Phenotype cluster 9 is the most populous of the induced phenotypes and does not correspond with

243   any of the embedded control drugs (Figure 2B and 3B). Visual inspection (Figure 2F) shows that cluster

244   9 cells are small and round, with thick stress fibres and intense actin puncta. They tend to be isolated

245   rather than in close contact, resulting in low cell density. Interestingly, this phenotype corresponds

246   closely with that induced by overexpression of the focal adhesion protein talin (Figure S4A-C). Talin is

247   critical for integrin activation and mechanically links integrin cytoplasmic tails with F-actin (Klapholz

248   and Brown, 2017; Yao et al., 2016), thereby influencing actin organisation.

249   We hypothesized that compounds found in phenotype cluster 9 may influence talin. Four such

250   compounds (structures in Figure 4F-I; phenotypes Figure S4D-G) were randomly selected for

251   saturation transfer difference (STD) NMR to test for direct talin-binding (Figure 4F-I). Of these, two

252   compounds (designated T1 and T2) bind talin directly and selectively, since they did not also bind

253   another core adhesion component, vinculin (example in Figure S4H). Notably, this expands upon the

254   single talin-binding compound previously identified (Yang et al., 2017). In contrast, compounds

255   designated T3 and T4 did not bind talin despite inducing an indistinguishable phenotype, suggesting

256   possible targeting of a related focal adhesion component. More broadly, these results again confirm

257   the capacity for distinct molecular mechanisms to drive the emergence of singular phenotypic end-

258   states.

Temporal analyses reveal different phenotypic trajectories to the same end-state

We next sought to test whether different compounds that yield the same phenotypic state after 24 hours exposure undergo the same phenotypic changes with time or arrive at the 24-hour state via different trajectories. As the phenotypic equivalence of mechanistically distinct perturbations was observed in several instances, and because the limited phenotypic plasticity we observe constitutes a global property of the actin system, we focused on 'systems-level' mechanisms that may explain both observations in a coherent manner. Given this perspective, we considered whether the recurrent actin phenotypes we observe might represent particular stable equilibria, or 'attractor states', which reflect limited phenotypic options available to the actin cytoskeletal system. This hypothesis was based on two factors. First, such attractor states tend to arise in complex systems composed of extensive, highly inter-dependent networks – characteristics found not only in the context of gene regulation (Huang et al., 2005) but also in the actin regulatory network (Agarwal and Zaidel-Bar, 2018; Senju and Lappalainen, 2018; Steinbacher and Ebnet, 2018). Second, attractor states tend to constitute relatively infrequent points of stable equilibrium within state-space landscapes comprising much larger repertoires of unstable states. Thus, the low phenotypic diversity we have observed after 24 hours of compound exposure mirrors the theoretical view that stable attractors are low-frequency phenomena.

Given this reasoning, we investigated whether observed actin phenotypes may actually demarcate stable systems-level attractor states. Specifically, we sought evidence of a third phenomenon predicted by attractor state theory, namely, that distinct transitional states may dynamically converge towards coincident end-states – as has been observed in the behaviour of gene regulatory networks over time (Huang et al., 2005). To test this possibility, we chose two compounds (T1 and T2) that generate equivalent talin-like phenotypes at 24 hours but are also structurally divergent (Figure 4F-G). We mapped actin phenotypes induced by each of these compounds at 12 timepoints post-treatment (1, 5, 15, 30, 45 minutes and 1, 2, 3, 4, 6, 20, 24 hours), enabling comparison of their phenotypic trajectories (Figure 4J).

285   Despite both binding to talin, these compounds induced phenotypes that diverged strongly after just

286   a few minutes of treatment, remaining clearly distinct for several hours (Figure 4J; 4-hour cell images

287   in Figure S5A, left), before converging after ~20 hours of treatment (Figure 4J; 24-hour cell images in

288   Figure S5A, right) (outliers excluded as detailed in Figure S5B-D). These observations support the

289   attractor state hypothesis as an emergent mechanism shaping and constraining phenotypic plasticity

290   within the actin cytoskeletal system.

Discussion

292   In this study, we estimated the degree of cellular-scale phenotypic plasticity available to the F-actin

293   cytoskeletal system by using a large, structurally diverse library of chemical perturbations to induce a

294   high degree of variability in adaptive responses. This unbiased approach revealed a comparatively low

295   number of (approximately 25) distinct F-actin phenotypes, strongly indicating that F-actin

296   organizational plasticity is constrained to far less than the hundreds or thousands of phenotypes that

297   might have been expected given the large number and diversity of chemical challenges applied

298   (~115k), and the known complexity of the actin regulatory system.

299   Our findings thus support the notion that actin phenotypes reflect a limited set of stable equilibria, or

300   attractor states, in the organisation of the actin system. This is reminiscent of dynamic perturbation

301   responses mapped in high-dimensional analyses of gene regulatory networks (Huang et al., 2009);

302   (Zhou et al., 2016). Crucially, channelling of divergent gene expression profile trajectories towards

303   equivalent end-states is increasingly attributed to the attractor state hypothesis (Huang et al., 2005),

304   with support from Boolean (gene) regulatory network modelling and the broader 'canalization'

305   concept embodied, for example, in Waddington's theory of cellular differentiation (Huang, 2012;

306   Waddington, 1942). Notably, these theories draw upon the emergence of a systems-scale 'logic' from

307   the highly interconnected and interdependent characteristics of gene regulatory networks.

308   Given the connectivity and interdependence of the actin regulatory network, we suggest that parallels

309   may exist between the emergence of, and constraints on, plasticity in this molecular system.

310   Specifically, mutation-driven changes in actin regulation might be projected to have strongly

311   deleterious consequences, due to the highly interconnected nature of this biological system and its

312   pivotal involvement in so many essential functions. However, our results suggest that by channelling

313   diverse perturbations into a narrow array of recurring attractor states, the actin system achieves an

314   intrinsic buffering capacity that prioritises useful organisation of actin filaments, thereby reducing the

315   impact of deleterious perturbations. Such a buffering capacity may facilitate evolutionary change and

316   diversity at the molecular level while maintaining the integrity of a structural system involved in

317   essentially all cell functions.

318

Author Contributions

326   Conceptualisation: NSB, JRS, ECH, PWG and JGL. Chemical perturbations: TWF, GMA and NSB. Image

327   analysis and data curation: NSB. Statistical analysis: JGL. Target validation: NSB, JGL, KB, ID, CL, BTG

328   and SZ. Software: JGL, YA and LB. Supervision and Funding: NSB, KG, ECH and PWG. Writing – original

329   draft: JGL and NSB. Writing – review and editing: All authors.

Declaration of Interests

331   PWG and ECH are Directors of TroBio Therapeutics, a company that is commercialising anti-

332   tropomyosin drugs for the treatment of cancer and their labs receive funding from TroBio

333   Therapeutics to evaluate anti-tropomyosin drug candidates.

Figure Legends

Figure 1. *Overview of experimental and analytical strategy.* **A)** Flow chart detailing major steps in the

phenotypic analysis workflow from cell biology to imaging to statistical analyses. **B)** Raw images of

Atto488-phalloidin and Hoechst 33342 staining (upper panels). Image analysis algorithms segment

nuclei to seed single cell boundary detection (middle panels). Stress fibres and actin rich puncta are

then segmented and measured per cell (bottom panels). Scale bar 50 µm. **C)** "Quadrangulation"

Strategy for Phenotype Number Estimation. To objectively estimate the number of distinct

phenotypes induced across the 114,400 compounds applied, we combined inferences from 4 distinct

methods (see grey boxes). Method 1 involved visual inspection of raw image data, providing a range

estimate for the number of visually discernible phenotypes. In method 2, hierarchical clustering was

applied using two alternate criteria (the SD index and Dunn's index) to provide additional range

estimates for phenotype number (Figure S2A). Both method 1 and 2 suggested that phenotype

number lay in the range between ~15 and ~40. This range guided subsequent phenotype number

estimation based on two alternate dimension reduction techniques (t-SNE [*exploring the tuneable*

'perplexity' parameter-space] or UMAP [*exploring the tuneable 'nearest neighbours' parameter-*

space]) combined with cluster (i.e. phenotype) detection (OPTICS algorithm). Using this progressive

exploration of the t-SNE and UMAP tuneable parameter-spaces, method 3 seeks the maximum

phenotype number and/or the point where this estimate plateaus (Figure 2C *(t-SNE)*, 2E *(UMAP)*;

upper). Method 3 tends to push phenotype number estimates higher (see dashed lines with hollow

arrow heads). Conversely, method 4 sets an upper bound on the phenotype number estimate by

monitoring "self-clustering" of positive controls (Jasplakinolide, Latrunculin A, TR100). These drugs are

known to induce relatively homogeneous phenotypes, providing a ground-truth expectation that

these treatments should predominantly "self-cluster" into a single phenotype. "Fragmentation" of any

of these controls into multiple phenotypes (resulting in reduced self-clustering) therefore implies

over-clustering/over-fitting, i.e. an excessive estimate of phenotype number (Figure 2C *(t-SNE)*, 2E

(UMAP); lower). Thus, method 4 tends to constrain phenotype number estimates (see dashed lines

360   with solid arrow heads). When combined, the trade-off between method 3 and 4 defines a final

361   optimised phenotype number estimate (using either t-SNE or UMAP dimension reduction).

362   Significantly, both UMAP (~16 phenotypes) and t-SNE (~25 phenotypes) suggest the emergence of a

363   surprisingly low number of distinct phenotypes.

364

365   **Figure 2.** *Assessing the diversity of F-actin organisation phenotypes using unbiased chemical*

366   *perturbations.* **A)** Example images of the diverse F-actin organisation phenotypes induced by chemical

367   perturbations. Scale bar 50 μm. **B)** Scatter plot of observations (123,108 observations, n = 22

368   independent experiments) distributed according to t-SNE dimension reduction (t-SNE tuneable

369   parameter 'perplexity' = 60) with 25 distinct phenotype clusters (colour-coded) defined by

370   unsupervised clustering via the OPTICS algorithm. **C)** To select the optimal t-SNE projection, 17

371   perplexity values were tested with 11 random seeding-replicates per value (187 total t-SNE

372   projections) as part of the 'quadrangulation strategy' (described in Figure 1C) for t-SNE projection

373   optimisation. Monitoring of changes in detected phenotype number (Y-axis, *upper panel*) and positive

374   control 'self-clustering' (Y-axis*, lower panel*; Jasplakinolide, aqua; Latrunculin, yellow; TR100, pink)

375   identified the optimal perplexity value (60, as used in B; dashed vertical line; X-axes). Smoothed means

376   (thin solid lines), 95% confidence intervals of smoothed means (pale envelopes) and median values

377   (dashed lines) are shown. **D)** Scatter plot of observations distributed according to UMAP dimension

378   reduction (UMAP tuneable parameter 'nearest neighbour' = 21) with 16 distinct phenotype clusters

379   (colour-coded) defined by unsupervised clustering via the OPTICS algorithm. **E)** To select the optimal

380   UMAP projection, 14 nearest neighbour values were tested with 11 random seeding-replicates per

381   value (154 total UMAP projections) as part of the 'quadrangulation strategy' (described in Figure 1C)

382   for UMAP projection optimisation. Monitoring of changes in predicted phenotype number (*upper*

383   *panel*, Y-axis) and positive control 'self-clustering' (*lower panel*, Y-axis; Jasplakinolide, green;

384   Latrunculin, red; TR100, orange) allowed identification of the optimal nearest neighbour value (21, as

385 used in D; dashed vertical line; X-axes). Smoothed means (thin solid lines), 95% confidence intervals

386 of smoothed means (pale envelopes) and median values (dashed lines) are shown. **F)** Example screen

387 images of phenotypes induced by two unknown compounds from each selected cluster. Scale bar 50

388 µm. **G)** Parallel coordinates plot depicting similarities and differences between all 25 phenotype

389 clusters (colour-coded as in D) across 15 key feature values. Key features determined by forward

390 feature selection based on random forest machine learning.

391

392 **Figure 3.** *Distinct clustering of compounds confirms robustness and sensitivity of the phenotypic*

393 *analyses and allows comparative analyses of feature values and inter-feature relationships to highlight*

394 *specific differences between major phenotypes.* **A)** 2D t-SNE plot showing 123,108 observations (n =

395 22 independent experiments) highlighting the distribution of the 4 control compounds: DMSO (pink);

396 Jasplakinolide (yellow); Latrunculin A (red); TR100 (green). **B)** Sunburst plot showing how control or

397 unknown compounds map to drug-induced cluster phenotypes. **C)** Sunburst plot showing the number

398 of unknown compounds in induced phenotype clusters. **D)** Enlarged views of clusters 4, 5 and 9 with

399 co-clustered unknown compounds denoted as black dots and the number of unknown compounds

400 listed. E) Upper row; for each of 4 major phenotypic clusters (1, 4, 5, 9), networks depict the wiring of

401 information theoretic dependencies between pairs of key features based on calculation of uncertainty

402 coefficients. Arrows indicate the direction of an inferred dependency, based on the mutual

403 information between each feature pairing, corrected for asymmetries in the entropy of each feature.

404 The relative strength of dependencies is coded by line thickness and colour (yellow, weak; red, strong).

405 Lower row; networks depict plasticity in uncertainty coefficient dependencies between each feature

406 pair based on comparison of phenotypic clusters 4, 5 or 9 to cluster 1 (the main non-responsive

407 phenotype). This highlights relationships that are 'gained' (present in phenotype 4, 5 or 9 but not in

408 1; purple), versus relationships that are 'lost' (present in phenotype 1 but not in 4, 5 or 9; cyan).

409

410 **Figure 4.** *Compounds producing co-clustered phenotypes can have distinct mechanisms of action.* **A)**

411 Western blot of a representative G:F actin assay (3 indepedent experiments) probed with total actin

412 antibody. Changes in actin composition between G-actin (G) and F-actin (F) for 3 compounds identified

413 from the Latrunculin-like cluster 4 as well as the positive control Latrunculin A and vehicle control

414 DMSO are shown. **B)** Representative line graph (3 independent experiments) of pyrene actin filament

415 assembly in response to Latrunculin-like compounds L2 and L3, compared with the positive control

416 Latrunculin and vehicle control DMSO. **C-E)** Representative widefield fluorescent images of SK-N-SH

417 cells immunostained with the γ9d antibody that detects Tpm3.1. Cells were treated with DMSO **(C)**,

418 or two compounds from cluster 5 that had differential effects on Tpm3.1 incorporation into stress

419 fibres **(D,E)**, scale bars C-E 20 μm. **(F-I)**. Talin binding assay. Representative NMR spectra of compounds

420 T1-T4 (blue lines) and STD-NMR spectra (red lines) of compounds T1-T4 in the presence of full length

421 talin. Chemical structures of compounds T1 (PubChem CID 42913180), T2 (PubChem CID 850363), T3

422 (PubChem CID 3351420) and T4 (PubChem CID 1211816) are shown inset. **J)** Comparison of time-

423 resolved phenotypic trajectories in t-SNE-space for cells treated with T1 (10 μM, duplicate wells; small

424 yellow to large red dots indicative of treatment time from 1 minute to 24 hours) vs T2 (10 μM,

425 duplicate wells; small cyan to large dark blue dots indicative of treatment time from 1 minute to 24

426 hours). Bezier-fitted lines (T1 – dark red; T2 – purple; arrows indicate the direction of time) highlight

427 the equivalence of initial phenotypes (adjacent to negative controls – grey dots). Trajectories rapidly

428 diverge in phenotypic space during the initial hours of treatment before converging again at 20 and

429 24 hours. **K)** Convergent phenotypic responses; a schematic summary of evidence for an evolutionary

430 buffering capacity in actin phenotype control in the face of diverse chemical challenges. By challenging

431 the actin cytoskeleton to respond to an overwhelming diversity of chemical perturbations ('High

432 Perturbation Diversity; drug 'perturbation diversity' hierarchy illustrative only), we have revealed how

433 such diverse challenges are channelled into a limited number of organisational phenotypes ('Low

434 Phenotypic Diversity'; calculated 'phenotypic diversity' hierarchy shown, see also Figure S6). This

435 implies that numerous molecular mechanisms of action may translate into singular phenotypic

436 outcomes, a suggestion confirmed by comparisons of the mechanisms-of-action within several pairs

437 of chemical compounds drawn equivalent phenotypic clusters ('Distinct mechanisms with same

438 phenotype'). This convergence of mechanistic diversity into phenotypic similarity suggests that

439 recurrent actin phenotypes may constitute attractor states within the broader actin phenotypic

440 landscape. This attractor state hypothesis is supported by the final convergence of initially divergent

441 phenotypic trajectories, as detailed in Figure 4J.

References

443 Agarwal, P., and Zaidel-Bar, R. (2018). Principles of Actomyosin Regulation In Vivo. Trends in cell
444 biology.
445 Ankerst, M., Breunig, M.M., Kriegel, H.-P., and Sander, R. (1999). OPTICS: ordering points to identify
446 the clustering structure. SIGMOD Rec *28*, 49-60.
447 Baell, J.B. (2013). Broad coverage of commercially available lead-like screening space with fewer
448 than 350,000 compounds. Journal of chemical information and modeling *53*, 39-55.
449 Baell, J.B., and Holloway, G.A. (2010). New substructure filters for removal of pan assay interference
450 compounds (PAINS) from screening libraries and for their exclusion in bioassays. Journal of medicinal
451 chemistry *53*, 2719-2740.
452 Becht, E., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Evaluation of
453 UMAP as an alternative to t-SNE for single-cell data. bioRxiv, 298430.
454 Berthold, M.R., Cebron, N., Dill, F., Gabriel, T.R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., and
455 Wiswedel, B. (2008). KNIME: The Konstanz Information Miner (Berlin, Heidelberg: Springer Berlin
456 Heidelberg).
457 Biedler, J.L., Helson, L., and Spengler, B.A. (1973). Morphology and growth, tumorigenicity, and
458 cytogenetics of human neuroblastoma cells in continuous culture. Cancer research *33*, 2643-2652.
459 Birmingham, A., Selfors, L.M., Forster, T., Wrobel, D., Kennedy, C.J., Shanks, E., Santoyo-Lopez, J.,
460 Dunican, D.J., Long, A., Kelleher, D*., et al.* (2009). Statistical methods for analysis of high-throughput
461 RNA interference screens. Nature methods *6*, 569.
462 Bonello, T.T., Janco, M., Hook, J., Byun, A., Appaduray, M., Dedova, I., Hitchcock-DeGregori, S.,
463 Hardeman, E.C., Stehn, J.R., Bocking, T*., et al.* (2016). A small molecule inhibitor of tropomyosin
464 dissociates actin binding from tropomyosin-directed regulation of actin dynamics. Sci Rep *6*, 19816.
465 Bryce, N.S., Hardeman, E.C., Gunning, P.W., and Lock, J.G. (2019). Chemical biology approaches
466 targeting the actin cytoskeleton through phenotypic screening. Curr Opin Chem Biol *51*, 40-47.
467 Bubb, M.R., Senderowicz, A.M., Sausville, E.A., Duncan, K.L., and Korn, E.D. (1994). Jasplakinolide, a
468 cytotoxic natural product, induces actin polymerization and competitively inhibits the binding of
469 phalloidin to F-actin. The Journal of biological chemistry *269*, 14869-14871.
470 Burridge, K., and Wittchen, E.S. (2013). The tension mounts: stress fibers as force-generating
471 mechanotransducers. The Journal of cell biology *200*, 9-19.
472 Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: An R Package for Determining
473 the Relevant Number of Clusters in a Data Set. 2014 *61*, 36.
474 Cleary, P.W., Tomas, D., Bolger, M., Hetherton, L., Rusinski, C., and Watkins, D. (2015). Using
475 Workspace to automate workflow processes for modelling and simulation in engineering. Paper
476 presented at: MODSIM2015, 21st International Congress on Modelling and Simulation.
477 Dunn, J.C. (1974). Well-Separated Clusters and Optimal Fuzzy Partitions. Journal of Cybernetics *4*, 95-
478 104.

479 Hahsler, M., and Peiekenbrock, M. (2017). dbscan: Density Based Clustering of Applications with
480 Noise (DBSCAN) and Related Algorithms. R package version 1.1-1. .
481 Halkidi, M., Vazirgiannis, M., and Batistakis, Y. (2000). Quality Scheme Assessment in the Clustering
482 Process, Vol 1910.
483 Huang, S. (2012). The molecular and mathematical basis of Waddington's epigenetic landscape: a
484 framework for post-Darwinian biology? Bioessays *34*, 149-157.
485 Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D.E. (2005). Cell fates as high-dimensional attractor
486 states of a complex gene regulatory network. Physical review letters *94*, 128701.
487 Huang, S., Ernberg, I., and Kauffman, S. (2009). Cancer attractors: a systems view of tumors from a
488 gene network dynamics and developmental perspective. Semin Cell Dev Biol *20*, 869-876.
489 Huang, S., Li, F., Zhou, J.X., and Qian, H. (2017). Processes on the emergent landscapes of
490 biochemical reaction networks and heterogeneous cell population dynamics: differentiation in living
491 matters. Journal of The Royal Society Interface *14*.
492 Inc., P.T. (2015). Collaborative data science (Montréal, QC: Plotly Technologies Inc.).
493 Janco, M., Rynkiewicz, M.J., Li, L., Hook, J., Eiffe, E., Ghosh, A., Bocking, T., Lehman, W.J., Hardeman,
494 E.C., and Gunning, P.W. (2019). Molecular integration of the anti-tropomyosin compound ATM-3507
495 into the coiled coil overlap region of the cancer-associated Tpm3.1. Sci Rep *9*, 11262.
496 Klapholz, B., and Brown, N.H. (2017). Talin - the master of integrin adhesions. Journal of cell science
497 *130*, 2435-2446.
498 Krijthe, J.H. (2015). {Rtsne}: T-Distributed Stochastic Neighbor Embedding using Barnes-Hut
499 Implementation.
500 Lehtimaki, J., Hakala, M., and Lappalainen, P. (2017). Actin Filament Structures in Migrating Cells.
501 Handbook of experimental pharmacology *235*, 123-152.
502 Leijnse, N., Oddershede, L.B., and Bendix, P.M. (2015). An updated look at actin dynamics in
503 filopodia. Cytoskeleton (Hoboken, NJ) *72*, 71-79.
504 Malo, N., Hanley, J.A., Cerquozzi, S., Pelletier, J., and Nadon, R. (2006). Statistical practice in high-
505 throughput screening data analysis. Nat Biotechnol *24*, 167-175.
506 McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection
507 for dimension reduction. arXiv preprint arXiv:180203426.
508 Moon, K.R., Stanley, J.S., Burkhardt, D., van Dijk, D., Wolf, G., and Krishnaswamy, S. (2018). Manifold
509 learning-based methods for analyzing single-cell RNA-sequencing data. Current Opinion in Systems
510 Biology *7*, 36-46.
511 Murtagh, F., and Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which
512 Algorithms Implement Ward's Criterion? Journal of Classification *31*, 274-295.
513 Pollard, T.D. (2016). Actin and Actin-Binding Proteins. Cold Spring Harbor perspectives in biology *8*.
514 Schevzov, G., Whittaker, S.P., Fath, T., Lin, J.J., and Gunning, P.W. (2011). Tropomyosin isoforms and
515 reagents. Bioarchitecture *1*, 135-164.
516 Senju, Y., and Lappalainen, P. (2018). Regulation of actin dynamics by PI(4,5)P2 in cell migration and
517 endocytosis. Current opinion in cell biology *56*, 7-13.
518 Shaner, N.C., Lambert, G.G., Chammas, A., Ni, Y., Cranfill, P.J., Baird, M.A., Sell, B.R., Allen, J.R., Day,
519 R.N., Israelsson, M.*, et al.* (2013). A bright monomeric green fluorescent protein derived from
520 Branchiostoma lanceolatum. Nature methods *10*, 407-409.
521 Skau, C.T., and Waterman, C.M. (2015). Specification of Architecture and Function of Actin
522 Structures by Actin Nucleation Factors. Annual review of biophysics *44*, 285-310.
523 Spector, I., Shochet, N.R., Kashman, Y., and Groweiss, A. (1983). Latrunculins: novel marine toxins
524 that disrupt microfilament organization in cultured cells. Science (New York, NY) *219*, 493-495.
525 Stehn, J.R., Haass, N.K., Bonello, T., Desouza, M., Kottyan, G., Treutlein, H., Zeng, J., Nascimento,
526 P.R., Sequeira, V.B., Butler, T.L.*, et al.* (2013). A novel class of anticancer compounds targets the actin
527 cytoskeleton in tumor cells. Cancer research *73*, 5169-5182.
528 Steinbacher, T., and Ebnet, K. (2018). The regulation of junctional actin dynamics by cell adhesion
529 receptors. Histochemistry and cell biology *150*, 341-350.

530   Team, R.C. (2014). R: A language and environment for statistical computing. R Foundation for
531   Statistical Computing, Vienna, Austria.
532   Ulyanov, D. (2016). Multicore-TSNE (Github).
533   van der Maaten, L. (2014). Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine
534   Learning Research *15*, 3221-3245.
535   Van der Maaten, L., and Hinton, G.E. (2008). Visualizing High-Dimensional Data Using t-SNE. . Journal
536   of Machine Learning Research *9*, 2579-2605.
537   Vindin, H., Bischof, L., Gunning, P., and Stehn, J. (2014). Validation of an algorithm to quantify
538   changes in actin cytoskeletal organization. J Biomol Screen *19*, 354-368.
539   Waddington, C.H. (1942). Canalization of development and the inheritance of acquired characters.
540   Nature *150*, 563.
541   Ward, J.H. (1963). Hierarchical Grouping to Optimize an Objective Function. Journal of the American
542   Statistical Association *58*, 236-244.
543   Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag).
544   Yang, C., Kwon, S., Kim, S.J., Jeong, M., Park, J.Y., Park, D., Hong, S.J., Jung, J.W., and Kim, C. (2017).
545   Identification of Indothiazinone as a Natural Antiplatelet Agent. Chemical biology & drug design.
546   Yao, M., Goult, B.T., Klapholz, B., Hu, X., Toseland, C.P., Guo, Y., Cong, P., Sheetz, M.P., and Yan, J.
547   (2016). The mechanical response of talin. Nature communications *7*, 11966.
548   Zhou, J.X., Isik, Z., Xiao, C., Rubin, I., Kauffman, S.A., Schroeder, M., and Huang, S. (2016). Systematic
549   drug perturbations on cancer cells reveal diverse exit paths from proliferative state. Oncotarget *7*,
550   7415-7425.

551

STAR Methods

Contact for reagent and resource sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr John Lock (john.lock@unsw.edu.au) or else by Prof. Peter Gunning (p.gunning@unsw.edu.au). This study did not generate any unique reagents. There are restrictions to the availability of compound identifiers for commercial reasons.

Experimental model

Human, female, SK-N-SH neuroblastoma cells (Biedler et al., 1973) were cultured in Dulbecco's Modified Eagle Medium (DMEM) plus 10% fetal bovine serum (Invitrogen) in a humidified environment at 37 °C with 5 % CO_2. Cells were confirmed to be Mycoplasma free with tests performed using the PCR Mycoplasma Test kit (AppliChem GmbH) as per manufacturer's instructions. Cells were transfected with pmNeonGreen-talin (a gift from Nathan Shaner and Jiwu Wang) (Shaner et al., 2013) using Lipofectamine LTX (Invitrogen).

Method details

Unbiased chemical perturbations

Unbiased chemical perturbations were conducted using the 114,400 compound WECC diversity library (Baell, 2013; Baell and Holloway, 2010). SK-N-SH cells were seeded at a density of 1800 cells/well into 384-well ViewPlates (Perkin Elmer) using a Multidrop 384 (Thermo) liquid dispenser and incubated at 37°C for 18 h. Library compounds and negative (vehicle only; DMSO) and phenotype-specific positive controls (8 μM TR100, 40 nM Jasplakinolide, 250 nM Latrunculin A) were dispensed to assay plates (100 nL; each control replicated 8 times per plate) using a Janus liquid handling robot (Perkin Elmer) equipped with a 384-well pintool in a cell::explorer automated workstation (Perkin Elmer). A total of 124,767 experimental conditions were assayed over n = 22 independent experiments. Each library compound was applied as a chemical perturbation at a single fixed concentration of 10 μM. Following 24 h drug exposure cells were fixed by addition of 16% paraformaldehyde (PFA) (ProSciTech) for 30

577 min at room temperature using a Multidrop Combi liquid dispenser (Thermo Scientific). Cells were

578 washed twice with PBS using a Biotek Elx405 plate washer

Cell Labelling, Imaging and Image Analysis

580 Cells were permeabilised and stained concurrently with a solution of phalloidin 488-atto (1:1000, Atto-

581 Tec GmbH), Hoechst 33342 (1:10000, ThermoFisher) and Triton X-100 (0.1%, Sigma) in PBS in the dark

582 for 1 h. Stain solution was dispensed using a Janus liquid handling robot. Cells were then washed twice

583 with PBS prior to imaging. Images were acquired on an Opera LX high-content imaging system

584 equipped with spinning disk confocal optics, using a 20X air objective (numerical aperture 0.45). A

585 total of 8 fields of view were acquired for each well using two excitation sources, a 200mW 488nm

586 solid state laser (320ms exposure) and a Xenon UV lamp (40ms exposure). Each image was captured

587 with a 12-bit high QE CCD camera and no pixel binning was applied to the images. Image data was

588 then uploaded to both the Columbus data storage and analysis server (Perkin Elmer) or departmental

589 servers for analysis via the Columbus or Workspace programs respectively (details provided in

590 methods sections quantification and statistical analysis: image analysis).

G-actin:F-Actin assay

592 Each of 3 independent experiments was performed as follows: 1.5×10^6 cells were plated on 10 cm

593 dishes and incubated overnight. After treatment with DMSO or 25 µM of compound for 24 h, the cells

594 were harvested and processed using the G-actin/F-actin *In Vivo* Assay Biochem Kit (Cytoskeleton) as

595 per the manufacturer's protocol. The positive control of 5 µM Latrunculin A was incubated with cells

596 for 1 h before sample processing. The G-actin and F-actin fractions were run on 10% SDS-PAGE gels

597 and transferred for Western blotting onto a PVDF membrane using the Trans-Blot Turbo (Bio-Rad)

598 transfer system. Actin was detected using the C4 total actin antibody and the blot developed with

599 Luminata Crescendo Western HRP substrate (Merck) and imaged on a ChemiDoc MP imaging system

600 (Bio-Rad).

Tpm3.1 localisation

Each of 3 independent experiments was performed as follows: 2×10^5 cells were plated onto a coverslip in a 6 well plate and incubated at 37°C for 18 h. The cells were then treated with 10 μM of compound for 24 h. The cells were fixed with 4% paraformaldehyde and permeabilized with ice cold methanol, then blocked with 2% FBS for 1 h. Tpm3.1 was detected using the γ9d antibody (Schevzov et al., 2011) and a Alexa488-labelled secondary antibody. Widefield fluorescent images were taken on a Zeiss Axioskop40 using a Plan Apochromat 63X 1.4 Oil DIC lens with a Axiocam 506 mono camera and ZEN 2.5 (blue edition) software (Zeiss).

Expression and purification of full length talin1

E. coli BL21(DE3) were transformed with a pet21a plasmid containing full-length mouse Talin1 (FL-mTalin1) with a non-cleavable C-terminal His-tag. A single colony was used to inoculate a 5 ml overnight LB+ampicillin (100 μg/ml) culture that was then added to 500 ml LB+ampicillin and grown at 37°C to a density (OD_{600}) of 0.4. Protein expression was induced by addition of 200 μM IPTG, at 37°C for 3 h. Harvested cells were resuspended in 50 mM Tris pH 7.5, 500 mM NaCl, 10 mM imidazole, 5mM phenylmethylsulfonyl fluoride (PMSF) and lysed by sonication. Clarified cell lysate was bound to 1 ml Ni-NTA resin using a batch method. FL-mTalin1 was eluted from the Ni-NTA resin in 2 ml of 50 mM Tris pH 7.5, 500 mM NaCl, 150 mM imidazole, 5 mM PMSF. The purified talin was dialysed into 10 mM Phosphate pH 7.4, 75 mM NaCl, 2 mM DTT.

Saturation Transfer Difference Nuclear Magnetic Resonance (STD-NMR)

All NMR experiments were carried out at 25°C, in 10 mM sodium phosphate pH 7.4, 75 mM NaCl, 2 mM DTT, 5% (v/v) D_2O using a Bruker AVANCE III 600 MHz spectrometer equipped with CryoProbe. 600 μl FL-mTalin1 was prepared at a concentration of 1.5 μM, in 5 mm NMR sample tubes. TA Drug compounds (50 mM Stocks in DMSO) were added to FL-mTalin1 at a final concentration of 1.5 mM, 3 % DMSO was added to FL-mTalin1 for the control sample. A 1D reference spectrum was collected of all compounds at a concentration of 1.5 mM. All spectra were processed using Topspin.

626 STD-NMR spectra were acquired with 32k data points and 700 scans. The protein was irradiated at -1

627 ppm (on-resonance) and -30 ppm (off-resonance) with a train of Gaussian shaped pulses (50 ms). The

628 saturation time used in the STD experiments was 2 s. STD-NMR experiments were optimized on ligand-

629 only samples to ensure that the irradiation at the selected frequency for on-resonance scan did not

630 affect the ligand. The protein resonances were suppressed with a 30 ms spin-lock pulse. The final

631 saturation difference spectra were obtained by subtraction of the on-resonance spectra from the off-

632 resonance spectra.

633 **Pyrene assay**

634 Each of 3 independent experiments was performed as follows: Pyrene actin was purchased from

635 Hypermol (Bielefeld, Germany) and diluted to a 1 mg/mL (24 µM) stock solution. Before use,

636 spontaneously formed actin aggregates were removed by ultracentrifugation for 1 h at 40,000 rpm

637 and 4 °C. 50 µl samples for the pyrene assay consisted of: 30 µl H_2O, 10 µl $MgCl_2$ (10 mM), 5 µl F-actin

638 Buffer (100 mM Imidazole-Cl pH 7.4, 10 mM ATP, Hypermol, Germany), as well as 5 µl DMSO

639 (containing the indicated concentrations of the respective compound). 10 µl pyrene actin (24 µM)

640 were rapidly added to start polymerization. Pyrene fluorescence was monitored every 20 s over 1 h in

641 a 96-well fluorescence plate reader (Tecan) with 360 nm excitation and emission at 400 nm in

642 duplicate.

643 **Compounds**

644 Individual compounds for secondary analyses were sourced through Molport (Latvia). Stock solutions

645 of 50 mM in DMSO were aliquoted for single use and stored at -20°C. TR100 (GVKBio, India),

646 Jasplakinolide (Cayman Chemical) and Latrunculin A (Adipogen Life Sciences).

647 **Quantification and statistical analysis**

648 **Image analysis**

649 Cell morphometry and textural properties were measured using a custom workflow in Columbus

650 (Perkin Elmer). Image analysis was developed and carried out within CSIRO's Workspace software

651   platform (Cleary et al., 2015). Workspace is a cross-platform framework for constructing workflows in

652   a graphic drag-drop editor for a range of applications, including image analysis. Workspace's flexible

653   architecture allows users to connect their own specialized plug-ins to the framework. Our custom

654   cellular analysis workflow combined operations such as nucleus detection, cell detection and filament

655   detection from Image Analysis plug-in developed by CSIRO's Quantitative Imaging. In addition to the

656   filament detection previously described (Vindin et al., 2014) the updated algorithm finds and

657   quantifies punctate ("dot-like") structures within each cell, using a dot detector. The punctate

658   structures are detected relative to the background intensity. The dots, or regions of peak intensity,

659   are detected for a range of dot diameters. The peaks are thresholded on basis of relative (i.e. above

660   background) and absolute intensity. Parameters were initially optimized, then kept constant across all

661   chemical perturbations, in order to achieve measurements that correlated closely with actin

662   organisation. This was confirmed by eye using an overlay of software output with the original image.

Data and Code Availability Statement

664   The dataset and code generated during this study are available at Data Dryad,

665   https://doi.org/10.5061/dryad.1cg2dq2

Software and general data visualization

667   Data derived from image quantification were statistically analysed and visualized using Knime

668   (Berthold et al., 2008) with R (Team, 2014) integration. Data were primarily visualized using the R

669   packages "ggplot2" (Wickham, 2009) and "plotly" (Inc., 2015). Sunburst plots were generated using

670   the D3.js implementation in Knime. Workflow schematics were generated using Lucidchart (Lucid

671   Software Inc.).

Data import and initial filtering

673   Quantitative data derived from 328 individual 384-well plates were first filtered to remove missing

674   values (124,767 observations retained). Data were further filtered to remove conditions with less than

675   10 or more than 700 cell measurements as well as duplicated values, leaving 124,343 conditions

676   characterized by 75 measured features.

Normalization

678   Z-score normalization of data was performed per measured feature (1 feature removed due to low

679   variance) using robust statistics (median and median absolute deviation; MAD) (Malo et al., 2006)

680   describing the entire population of unknown drugs grouped per experimental date (25 dates). The

681   unknown drug Treatment population was a more effective reference for normalization than the DMSO

682   negative control because the Treatment population was orders of magnitude larger and because the

683   majority of these drugs induced no phenotype (~90%) (Birmingham et al., 2009). Normalization using

684   the Treatment population therefore achieved the best superimposition of data, enabling detection of

685   distinct phenotypes as opposed to inter-experimental variation.

Cluster Number Estimation

687   The number of data clusters (potential phenotypes) was estimated using the R package "NbClust"

688   (Charrad et al., 2014) based on Manhattan distances as a proximity measure and using Ward's

689   agglomerative hierarchical clustering method (Ward, 1963) to minimize total within-cluster-distances,

690   including a correction criterion where dissimilarities (distances) are squared before iteration updating

691   (Murtagh and Legendre, 2014). Two alternative indices were used to finally estimate the number of

692   clusters: Dunn's validity index, based on optimizing the distance between clusters *versus* the diameter

693   of clusters (Dunn, 1974) and; the SD validity index, based on optimizing the total separation of clusters

694   *versus* the average scattering of clusters (Halkidi et al., 2000). This procedure was performed using a

695   random sampling of approximately 10% of the dataset (12,000 observations) including all 74 Z-

696   normalized quantitative features, and was repeated 20 times. The mean estimates (~24 clusters,

697   Dunn's index; ~28 clusters, SD index) were used to guide parameterization of subsequent dimension

698   reduction and unsupervised clustering steps.

t-SNE Dimension Reduction

To limit data set complexity whilst retaining maximal information, principal components analysis (PCA) was performed prior to t-distributed stochastic neighbour embedding (t-SNE)-based dimension reduction. PCA was performed using all 74 Z-normalized quantitative features and parameterized to retain no less than 98% of total data set variance, resulting in the generation of 26 orthogonal principal components. t-SNE-based dimension reduction to 2 dimensions was then performed via the accelerated Barnes-Hut implementation (van der Maaten, 2014) using the R package "Rtsne" (Krijthe, 2015). As described in Figure 1C and Figure 2C, a quadrangulation strategy was used to guide optimisation of the main t-SNE tuning parameter, perplexity, leading to a final selected value of 60. As part of the quadrangulation strategy, 17 perplexity parameters values were tested, with 11 random seeding states replicated at each perplexity value, resulting in 187 distinct t-SNE projections. Performed directly within R, this process was accelerated through use of the 'Rtsne.multicore' package available from https://github.com/DmitryUlyanov/Multicore-TSNE (Ulyanov, 2016).

UMAP Dimension Reduction

To limit data set complexity whilst retaining maximal information, principal components analysis (PCA) was performed prior uniform manifold approximation and projection (UMAP)-based dimension reduction (McInnes et al., 2018). As with t-SNE, PCA was first performed using all 74 Z-normalized quantitative features and parameterized to retain no less than 98% of total data set variance, resulting in the generation of 26 orthogonal principal components. UMAP-based dimension reduction to 2 dimensions was then performed directly in R via the 'umap' package. As described in Figure 1C and Figure 2E, a quadrangulation strategy was used to guide optimisation of the main UMAP tuning parameter, nearest neighbour, leading to a final selected value of 21. As part of the quadrangulation strategy, 14 nearest neighbour parameter values were tested, with 11 random seeding states replicated at each nearest neighbour value, resulting in 154 distinct UMAP projections.

OPTICS Observation Clustering

To define robust and recurrent phenotypes based on observation clustering within 2D t-SNE or UMAP spaces, the rapid OPTICS (ordering points to identify the clustering structure) (Ankerst et al., 1999) unsupervised clustering algorithm was employed via the R package "dbscan" (Hahsler and Peiekenbrock, 2017). For clustering in the optimised t-SNE space, epsilon (0.3), minimum number of points (18) and cluster threshold (i.e. eps_cl; 0.3) values were selected based results of the quadrangulation strategy (Figure 1C) for t-SNE analysis. 25 observation clusters (phenotypes) were thus defined, with 1,235 non-clustered observations excluded, leaving 123,108 observations retained. This exclusionary approach to non-clustered observations was utilized so as to focus on robust and recurring phenotypes, rather than potentially unique (or noise-induced) perturbation-effects. For clustering in the optimised UMAP space, epsilon (0.15), minimum number of points (18) and cluster threshold (i.e. eps_cl; 0.15) values were selected based results of the quadrangulation strategy (Figure 1C) for UMAP analysis.

Supervised Feature Selection

To identify a limited set of (15) features that are maximally informative about differences between phenotypes, we applied Forward Feature Selection in combination with Random Forest machine learning-based prediction of phenotype (cluster) membership for 123,108 observations defined by all 74 Z-score normalized features. Using Knime, the Random Forest model was optimized by iterating over randomly partitioned data (20% for learning in presence of phenotype membership data; 80% for phenotype prediction in absence of phenotype membership data), using the information gain ratio as a splitting criterion. Model predictions were optimized with respect to maximal overall accuracy, which reached 98.2%, meaning the underlying model is informative.

Inter-feature relationship network mapping

Data summarising phenotypic clusters 1, 4, 5 and 9 across each of the 15 key (feature-selected) Z-normalized quantitative features was first selected. To ensure an equitable final comparison of inter-

748   feature relationship 'strengths' (i.e. uncertainy coefficient values), 1100 observations were randomly

749   sampled from each phenotypic cluster (as constrained by the smallest of these clusters (9), which

750   contains 1166 observations). The R package 'mpmi' was then used to calculate entropy values for each

751   quantitative feature and mutual information values between each pair of features. Mutual

752   information values for each feature pair were then independently normalized to the entropy value of

753   each feature pair-member, thus providing sensitivity to asymmetry in 'information overlap' as a

754   proportion of information captured within each feature. This constitutes an estimate of the

755   uncertainty coefficient in each 'direction' of inter-feature relationships. Uncertainty coefficient values

756   greater than 0.2 (thresholded to limit network connections and enable visual interpretation) were

757   plotted as directed edges in a circular 15 node (15 key features) network layout using the R package

758   'igraph'. Colour-coding and line thickness were defined by uncertainty coefficient values. Comparison

759   of networks using the igraph function 'difference' revealed how networks are re-wired in phenotypic

760   clusters 4, 5 and 9 when compared to cluster 1 (the main non-responsive, spontaneous phenotype

761   cluster).

762   **Time-series analysis of phenotypic trajectories**

763   Data for time-series analysis of phenotypic trajectories was generated by parallel treatment of cells

764   with either negative control (DMSO), or 10 µM of compounds T1 or T2. Cells were treated for periods

765   of 1 min, 5 min, 15 min, 30 min, 45 min, 1 h, 2 h, 3 h, 4 h, 6 h, 20 h or 24 h. Cells were fixed,

766   permeabilised and labelled, imaged and quantified as described for the main chemical perturbation

767   analyses. Statistical analyses mirrored the main chemical perturbation analyses, beginning with

768   exclusion of conditions capturing < 10 > 700 cells, Z-normalisation relative to DMSO controls, and

769   principle components analysis producing 26 orthogonal principle components as input for t-SNE

770   dimension reduction using a perplexity value of 50, as previously. Data were then filtered based on

771   comparison of loess regression (span = 1) in the presence and absence of putative outliers.

772   Independent dynamic phenotypic trajectories for compounds T1 and T2 were then estimated by fitting

773   a Bezier curve to all remaining values ordered by time. Data were plotted using R.

Hierarchical clustering of phenotypes and features

774

775 A heatmap summarising hierarchical clustering of all 25 phenotypic clusters and all 74 Z-normalized

776 quantitative features was generated using the 'heatmap.2' function from the 'gplots' package in R.

777 Agglomerative clustering proceeded via complete linkage as described for the default 'hclust' function.

A

CELL BIOLOGY

SK-N-SH neuroblastoma cells

114 400 compounds + 10 400 control wells

Actin filament and nuclear stain

IMAGING AND IMAGE ANALYSIS

Spinning-Disc Confocal Imaging - 8 images * 2 channels/well; ~200 cells/well

1 996 800 total images

Quantitative Image Analysis - 75 features extracted (Table S1)

SYSTEMS ANAYSIS AND VISUALIZATION

Automated data import into Knime

Normalization

Hierarchical Clustering with Dunn's & SD validity indices

Dimension Reduction - PCA

TSNE Dimension Reduction

UMAP Dimension Reduction

Parameter Optimisation & Comparison

Phenotype Definition - Optics

Phenotype Definition - Optics

Feature Selection - Random Forest Machine Learning

Inter-Feature Network Inference - Uncertainty Coefficents

Integration of molecular structures/manual hits/raw images

Data visualization outputs

Data interaction

B

Phalloidin

Hoechst 33342

Nuclear detection

Cell boundary detection

Line detection

Dot detection

C

Method 1 — Visual Inspection of Images

Hierarchical Clustering & Dunn's or SD validity index (*Fig S2A*) — Method 2

Potential Range for Phenotype Number
2-3 .... >1000

range estimate ~15 ~40

Estimated Range for Phenotype Number
~15 .... ~40

t-SNE estimate of Phenotype Number
~25

UMAP estimate of Phenotype Number
~16

Phenotype Number "Maximum Plateau Point" (*Fig 2C,E; upper*) — Method 3

Positive Control Phenotype "Fragmentation Point" (*Fig 2C,E; lower*) — Method 4

UMAP dimension reduction + OPTICS

t-SNE dimension reduction + OPTICS

A Row of fluorescence microscopy images of cells, with the leftmost labeled "Control".

B t-SNE plot with axes t-SNE 1 and t-SNE 2, showing clusters labeled 01, 02, 04, 05, 06, 09, 11, 17, 21.

C Plots of Phenotype number and Positive control self-clustering versus Perplexity parameter.

D UMAP plot with axes UMAP 1 and UMAP 2.

E Plots of Phenotype number and Positive control self-clustering versus Nearest neighbour parameter.

F Grid of fluorescence microscopy images labeled by cluster: 01, 01, 02, 02, 04, 04 (top row); 05, 05, 06, 06, 09, 09 (middle row); 11, 11, 17, 17, 21, 21 (bottom row).

G Cluster: 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Parallel coordinates plot of Value (0.00 to 1.00) versus Variable. Variables: Cell number.z, Nucleus area.z, Cell area.z, Cell length.z, Cell ratio width/length.z, Cell haralick correlation 1px.z, Cell texture 2 px SER edge.z, Cell texture 2 px SER saddle.z, Spot ratio width/length.z, Spot std dev intensity.z, Spot intensity CV [%].z, Line length.z, Line intensity/cell.z, Line intensity/cell area.z, Dot Number.z

A

B

- Cluster 3 — Jas
  1966 — 2209
  (89.00%)

- Unknown — Cluster 9
  2003 — 1166
  (58.21%)

- Unknown — Cluster 6
  2003 — 228
  (11.38%)

- Unknown — Cluster 2
  2003 — 114
  (5.69%)

- Induced 8176
  - Unknown 2003 (24.50%)
  - TR100 1486 (18.18%)
  - Lat A 2476 (30.28%)
  - Jas 2209 (27.02%)
  - DMSO 2 (0.054%)

- TR100 — Cluster 5
  1486 — 1297
  (87.28%)

- Cluster 4 — Lat
  2354 — 2476
  (95.07%)

C

1166
(58.21%)

- Jasplakinolide-like 20 (1.00%)
- TR100-like 55 (2.75%)
- 61 (3.05%)
- Latrunculin-like 87 (4.34%)
- 94 (4.69%)
- 114 (5.69%)
- 228 (11.38%)

Induced — Unknown
8176 — 2003 (24.50%)

D

Cluster 4 — Unknown (87)

Cluster 5 — Unknown (55)

Cluster 9 — Unknown (1166)

- DMSO
- Jasplakinolide
- Latrunculin A
- TR100

E

Cluster 1 Cluster 4 Cluster 5 Cluster 9

Cluster 1 vs 4 Cluster 1 vs 5 Cluster 1 vs 9

A. Con, DMSO, LatA, L1, L2, L3 with G and F lanes; Total Actin

B. Fluorescence intensity (a.u.) vs time — Baseline, Control, Lat B 10μM, L2 10μM, L3 10 μM, L2 100μM, L3 100μM

C. DMSO D. Tr1 E. Tr2

F. T1 G. T2 H. T3 I. T4

J. t-SNE 1 vs t-SNE 2 — T1, T2, DMSO, 4 h, 24 h, 1 min

K.

Primary Screening

High Perturbation Diversity
(114,400 compounds)

Suggests multiple mechanisms may produce single phenotypes

Effects on Cell and Cytoskeletal Organisation

Low Phenotypic Diversity
(25 total, 20 drug-induced)

4 13 12 18 9 16 10 11 17 6 24 3 7 1 20 25 8 19 22 23 14 21 5 2 15

Secondary Testing

Confirms multiple mechanisms can produce single phenotypes

Distinct mechanisms with same phenotype

Lat A-like phenotype

Lat A-like mechanism | non-Lat A mechanism

Altered adhesion phenotype

Talin binding | non-Talin binding

TR100-like phenotype

TR100-like mechanism | non-TR100 mechanism

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
C4 Total actin	Millipore	Cat# MAB1501, RRID:AB_2223041
γ9d Tpm3.1	Millipore	Cat# MAB2256, RRID:AB_10806918
Chemicals, Peptides, and Recombinant Proteins		
TR100	GVK Bio, India	Made to order
Latrunculin A	Adipogen Life Sciences	Cat# AG-CN2-0027-C500
Latrunculin B	Adipogen Life Sciences	Cat# AG-CN2-0031-M001
Jasplakinolide	Cayman Chemical	Cat#11705
Hoechst 33342	ThermoFisher Scientific	Cat#62249
Pyrene actin	Hypermol	Cat#8112-04
Critical Commercial Assays		
G-actin/F-actin *In Vivo* Assay Biochem Kit	Cytoskeleton	Cat# BK037
Experimental Models: Cell Lines		
Human Female: SK-N-SH neuroblastoma cells	ATCC	Cat# HTB-11, RRID:CVCL_0531
Recombinant DNA		
pmNeonGreen-talin	Shaner et al., 2013	
Software and Algorithms		
Columbus	Perkin Elmer	
Workspace	CSIRO	
Knime	https://www.knime.com/	
R	https://www.r-project.org/	
R-package ggplot2	https://ggplot2.tidyverse.org/	
R-package plotly	https://plot.ly/r/	
R-package NbClust	https://cran.r-project.org/web/packages/NbClust/index.html	
R-package Rtsne	https://cran.r-project.org/web/packages/Rtsne/index.html	
R-package dbscan	https://cran.r-project.org/web/packages/dbscan/index.html	
R-package mpmi	https://CRAN.R-project.org/package=mpmi	
R-package igraph	http://igraph.org/r/	

R-package gplots	https://cran.r-project.org/web/packages/gplots/index.html	
R-package Rtsne.multicore	https://github.com/DmitryUlyanov/Multicore-TSNE	
R-package UMAP	https://cran.r-project.org/web/packages/umap/index.html	
Lucidchart	https://www.lucidchart.com/	
Deposited Data		
DataDryad – Bryce_et_al_Quantitative_Data	doi:10.5061/dryad.1cg2dq2	