

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE  
CURSO DE GRADUAÇÃO EM BIOMEDICINA

Rodrigo Haas Bueno

**METANÁLISE DE DADOS DE TRANSCRIPTOMA REVELA MÓDULOS  
PATO-FISIOLÓGICOS ENVOLVIDOS COM A FIBRILAÇÃO ATRIAL**

Porto Alegre

2019

Rodrigo Haas Bueno

**METANÁLISE DE DADOS DE TRANSCRIPTOMA REVELA MÓDULOS  
PATOFISIOLÓGICOS ENVOLVIDOS COM A FIBRILAÇÃO ATRIAL**

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel em Biomedicina.

Orientadora: Dra. Mariana Recamonde-Mendoza

Porto Alegre  
2019

#### CIP - Catalogação na Publicação

Haas Bueno, Rodrigo  
METANÁLISE DE DADOS DE TRANSCRIPTOMA REVELA MÓDULOS  
PATO-FISIOLÓGICOS ENVOLVIDOS COM A FIBRILAÇÃO ATRIAL /  
Rodrigo Haas Bueno. -- 2019.  
76 f.  
Orientadora: Mariana Recamonde-Mendoza.

Trabalho de conclusão de curso (Graduação) --  
Universidade Federal do Rio Grande do Sul, Instituto  
de Ciências Básicas da Saúde, Curso de Biomedicina,  
Porto Alegre, BR-RS, 2019.

1. Fibrilação Atrial. 2. Transcriptoma. 3.  
Metanálise. 4. Análise de Redes. 5. Análise de  
Enriquecimento Funcional. I. Recamonde-Mendoza,  
Mariana, orient. II. Título.

Rodrigo Haas Bueno

**METANÁLISE DE DADOS DE TRANSCRIPTOMA REVELA MÓDULOS  
PATOFIISIOLÓGICOS ENVOLVIDOS COM A FIBRILAÇÃO ATRIAL**

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel em Biomedicina.

Aprovado em: 02 de dezembro de 2019.

**BANCA EXAMINADORA**

---

Raquel Calloni – IFSUL - Gravataí

---

Santiago Tobar - HCPA

---

Mariana Recamonde-Mendoza – UFRGS (orientadora)

## RESUMO

A fibrilação atrial (FA) é o tipo de arritmia cardíaca mais comum e afeta de forma crescente milhões de pessoas ao redor do mundo. Sua patofisiologia envolve mecanismos complexos de remodelamento estrutural e elétrico dos átrios, muitos dos quais são pouco compreendidos. Dados de transcriptoma têm histórico em serem utilizados para levantar hipóteses de mecanismos biológicos por trás de doenças complexas. Contudo, a análise desse tipo de dado sofre com baixo poder estatístico e problemas de reprodutibilidade. Nesse sentido, este trabalho teve como objetivo realizar uma metanálise de dados de microarranjo envolvendo amostras de pacientes com fibrilação atrial. Os dados dos estudos selecionados foram baixados, analisados quanto a sua qualidade e pré-processados. A seguir, foi realizado uma análise de expressão diferencial de cada conjunto de dados individual. Estes resultados foram integrados por meio de metanálise, resultando numa lista final de 589 transcritos diferencialmente expressos na FA em relação a controles. Uma rede de interação proteína-proteína foi gerada a partir dessa lista e analisada quanto a seus parâmetros topológicos - identificando 11 *hubs-bottlenecks* principais - e por meio de análise de clusterização - identificando 4 módulos funcionais principais. Estes 4 módulos enriqueceram para 23, 54, 33 e 53 vias biológicas, respectivamente, as quais envolvem, entre outros, processos relacionados ao estresse do retículo endoplasmático, estresse oxidativo, catabolismo proteico e inflamação. Os *hubs-bottlenecks*, por sua vez, representam proteínas densamente conectadas e com possível papel chave regulatório da rede principal. A chaperona HSPA5, a proteína de ancoramento ANK2 e a proteína CTNNB1, chave da via Wnt/ $\beta$ -catenina são alguns exemplos. Dessa forma, a metanálise de dados de microarranjo foi útil em identificar mecanismos patofisiológicos e transcritos desregulados envolvidos com a FA, demonstrando seu uso potencial no estudo de doenças complexas. Nossos resultados poderão, portanto, ser utilizados como base sólida para o levantamento de hipóteses de pesquisas futuras envolvendo a FA.

Palavras-chave: Análise de redes. Enriquecimento funcional. Fibrilação atrial. Metanálise. Microarranjo. Transcriptoma.

## ABSTRACT

Atrial fibrillation (FA) is the most common type of cardiac arrhythmia and affects millions of people around the world. Its pathophysiology involves complex mechanisms of structural and electrical remodeling of the atria, many of which are poorly understood. Transcriptome data have a history of being used to hypothesize biological mechanisms behind complex diseases. However, analysis of this type of data suffers from low statistical power and reproducibility problems. Therefore, this study aimed to perform a meta-analysis of microarray data involving samples of patients with atrial fibrillation. Data extracted from the selected studies were downloaded, analyzed for quality and normalized. Next, a differential expression analysis of each individual data set was performed. These results were integrated by meta-analysis, resulting in a final list of 589 differentially expressed transcripts in FA relative to controls. A protein-protein interaction network was generated from this list and analyzed for its topological parameters - identifying 11 main hubs-bottlenecks - and by clustering analysis - identifying 4 main functional modules. These 4 modules enriched for 23, 54, 33 and 53 biological pathways, respectively, which involve, among others, processes related to endoplasmic reticulum stress, oxidative stress, protein catabolism and inflammation. The hubs-bottlenecks, in turn, represent densely connected proteins with possible regulatory key role in the main network. Chaperone HSPA5, ANK2 anchoring protein and CTNNB1 protein, key to the Wnt/ $\beta$ -catenin pathway, are some examples. Thus, the meta-analysis of microarray data was useful in identifying pathophysiological mechanisms and dysregulated transcripts involved with AF, demonstrating their potential use in the study of complex diseases. Our results can therefore be used as a solid basis for raising hypotheses of future research involving AF.

**Keywords:** Network analysis. Functional enrichment. Atrial fibrillation. Meta-analysis. Microarray. Transcriptome.

## **LISTA DE FIGURAS**

Figura 1 – Fases do potencial de ação cardíaco.....	10
Figura 2 – Esquema metodológico de uma metanálise de dados de microarranjo.....	16

## LISTA DE ABREVIATURAS

- ATR1: receptor de angiotensina II do tipo I
- ATR2: receptor de angiotensina II do tipo II
- CaMKII: calmodulina cinase tipo 2
- ERK: *extracellular signal-related kinase*
- FA: fibrilação atrial
- JNK: *C-jun N-terminal kinase*
- NCX: bomba antiporte 3Na<sup>+</sup>/1Ca<sup>2+</sup>
- PDGF: fator de crescimento derivado de plaquetas
- PDPs: pós-despolarizações precoces
- PDTs: pós-despolarizações tardias
- PLB: proteína fosfolambam
- RS: retículo Sarcoplasmático
- RyR2: receptor de rianodina
- TGFβ1: fator de crescimento transformante β

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>8</b>
1.1 A FIBRILAÇÃO ATRIAL.....	8
<b>1.1.1 A Patofisiologia da Fibrilação Atrial .....</b>	<b>9</b>
1.2 TRANSCRIPTOMA .....	14
<b>1.2.1 O Microarranjo.....</b>	<b>15</b>
<b>1.2.2 Metanálise de Dados de Transcriptoma provenientes de Microarranjo.....</b>	<b>15</b>
1.3 JUSTIFICATIVA .....	17
1.4 OBJETIVOS .....	18
<b>1.4.1 Objetivo geral.....</b>	<b>18</b>
<b>1.4.2 Objetivos específicos.....</b>	<b>18</b>
<b>2 ARTIGO CIENTÍFICO.....</b>	<b>19</b>
<b>3 CONCLUSÕES E PERSPECTIVAS .....</b>	<b>51</b>
<b>REFERÊNCIAS .....</b>	<b>52</b>
<b>ANEXO A – NORMAS DE PUBLICAÇÃO DA REVISTA <i>MOLECULAR MEDICINE</i></b> <b>.....</b>	<b>55</b>

## 1 INTRODUÇÃO

A fibrilação atrial é um tipo de arritmia extremamente comum e responsável por crescente morbidade cardiovascular. Possui mecanismos patofisiológicos complexos envolvendo remodelamento estrutural e elétrico dos átrios, muitos dos quais não são compreendidos. Nesse sentido, a análise de dados de microarranjo pode servir de ferramenta para explorar novos mecanismos biológicos envolvidos com a doença. Contudo, análise de experimentos individuais de microarranjo sofrem com baixo poder estatísticos e baixa reprodutibilidade. Por isso, este estudo se propôs a realizar uma metanálise de dados de transcriptoma envolvendo a fibrilação atrial para mitigar esses reverses metodológicos e lançar luz aos mecanismos biológicos que a permeiam.

### 1.1 A FIBRILAÇÃO ATRIAL

A fibrilação atrial (FA) é o tipo de arritmia cardíaca de significância clínica mais comum, e está associada com crescente morbidade cardiovascular, especialmente no que diz respeito à chance de derrames, insuficiência cardíaca e morte súbita (CHUGH et al., 2014). Em 2010, o número estimado de pacientes diagnosticados com FA girava em torno dos 33 milhões de pessoas, afetando principalmente indivíduos com mais de 60 anos (40% dos casos). Até 2030, o número de novos casos diagnosticados por ano deve girar em torno dos 215.000, em função da inversão das pirâmides etárias ao redor do mundo nas próximas décadas (envelhecimento da população) (MCMANUS; RIENSTRA; BENJAMIN, 2012). Além de em idosos, a FA tem alta prevalência em pacientes com hipertensão, insuficiência cardíaca, doença arterial coronariana, anomalias valvulares, diabetes mellitus e doença crônica renal (BALL et al., 2013).

A FA se apresenta de formas distintas de acordo com o início, duração e conclusão dos episódios arrítmicos. Tipicamente, a FA evolui ao longo do tempo de episódios curtos e infrequentes (FA paroxística - termina em até 48 horas) para episódios mais longos e mais frequentes (FA persistente - dura mais que 7 dias ou até que seja realizada a cardioversão - ou permanente - dura mais do que um ano e é refratária à cardioversão) (CHARITOS et al., 2014). Dessa forma, muitos pacientes com FA têm sua qualidade de vida significativamente afetada, experienciando uma variedade de sintomas que incluem principalmente fadiga, mas também palpitações, dispneia, sensação de aperto no peito, problemas para dormir e

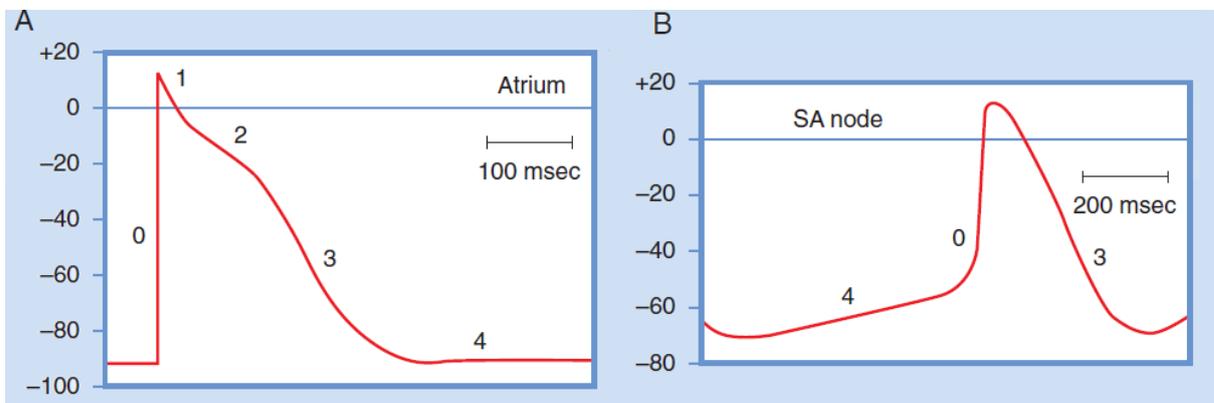
significativo estresse psicológico (THRALL et al., 2006). Hospitalizações também são frequentes entre esses pacientes, principalmente em função das comorbidades associadas à FA, o que também afeta negativamente sua qualidade de vida (DEVORE et al., 2016).

### 1.1.1 A Patofisiologia da Fibrilação Atrial

A FA é uma taquiarritmia supraventricular, caracterizada por ativação descoordenada e irregular dos átrios e consequente inefetividade da contração atrial (FUSTER et al., 2011). O conjunto disso com a decorrente variabilidade no enchimento ventricular associado a um controle pobre da frequência ventricular e a ativação do sistema nervoso simpático por despolarizações irregulares causam as principais alterações hemodinâmicas e sintomas da FA (IKEDA et al., 2012). A FA também confere grande aumento do risco de eventos tromboembólicos devido à formação de trombos, usualmente no apêndice atrial esquerdo. Além disso, o aparecimento da FA está frequentemente associada à exacerbação de outras doenças cardíacas, tanto por ser causa ou consequência da deterioração funcional e/ou estrutural do coração encontrada nessas condições (HEALEY et al., 2012). No paciente com FA, a perda da contração funcional dos átrios também pode diminuir o débito cardíaco, principalmente quando o enchimento dos ventrículos é ainda dificultado por quadros de estenose mitral, hipertensão e hipertrofia cardíaca (BROOKES et al., 1998).

O potencial de ação normal do músculo cardíaco (figura 1) exhibe mudanças típicas de voltagem ao longo do tempo. O *potencial de ação de resposta rápida* é característico de células musculares atriais e ventriculares normais e das fibras de Purkinje e tem início com um potencial de membrana intracelular negativo ( $-90$  mV: potencial de repouso), que se torna rapidamente positivo (depois que atinge o limiar de em torno de  $-65$  mV) durante a despolarização (fase 0), na qual ocorre um influxo rápido de sódio ( $\text{Na}^+$ ) devido à abertura de canais de  $\text{Na}^+$  dependentes de voltagem, que se inativam rapidamente em 1 ou 2 milissegundos. Com a inativação dos canais de  $\text{Na}^+$ , ocorre breve repolarização (fase 1) devido a abertura de canais de potássio ( $\text{K}^+$ ), a chamada *transient outward current*. Durante a fase de despolarização (fase 0), conforme o potencial de membrana torna-se positivo, abrem-se os chamados canais lentos (*L-type*) de cálcio ( $\text{Ca}^{2+}$ ). Estes canais são assim chamados porque, uma vez abertos, inativam-se lentamente. A abertura destes canais promove o influxo de  $\text{Ca}^{2+}$  para a célula cardíaca. Neste momento, o balanço entre o efluxo de  $\text{K}^+$  e o influxo de  $\text{Ca}^{2+}$  forma o chamado platô do potencial de ação (fase 2). Durante o platô, outros canais de  $\text{K}^+$  se abrem lentamente, os chamados *delayed rectifier channels*. Estes canais aumentam

gradualmente o efluxo de  $K^+$  e dão início ao processo de repolarização. Cabe ressaltar que a duração e amplitude do platô é menor nos átrios do que nos ventrículos principalmente devido à maior magnitude da *transient outward current* e da *delayed rectifier current*, uma vez que essas correntes aceleram a saída da fase 2 para a fase de repolarização (fase 3), a qual tem início quando o efluxo de  $K^+$  começa a exceder o influxo de  $Ca^{2+}$ . Conforme o potencial de membrana torna-se mais negativo, uma terceira corrente de  $K^+$  (*inward rectifier channels*) ganha força e passa a contribuir substancialmente para a repolarização. O potencial de repouso (fase 4) é, então, alcançado e mantido principalmente pela ação das bombas de membrana  $Na^+-K^+-ATPase$ , pela antiporte  $3Na^+/1Ca^{2+}$  (NCX) e pela bomba  $Ca^{2+}-ATPase$  (BERS, 2002).



**Figura 1 - Fases do potencial de ação cardíaco.** Átrios (A) e nodo sinoatrial (B). Adaptado de: (TERRENOIRE et al., 2005)

Algumas células do coração, notavelmente as dos nodos sinoatrial e atrioventricular - responsáveis pela automaticidade do coração (principalmente o nodo sinoatrial) -, por sua vez, exibem o chamado *potencial de ação de resposta lenta*. Comparado às células atriais e ventriculares, o potencial de repouso destas células é menos negativo ( $-40$  a  $-70$  mV), a fase 0 é menos íngreme, o platô não é proeminente e a fase 3 é mais gradual. Além disso, a fase 4 é marcada por lenta despolarização espontânea até que o limiar seja atingido e um novo potencial de ação seja disparado. O mecanismo por trás dessa despolarização espontânea está no influxo de  $Na^+$  pelos canais *funny* de  $Na^+$  (que se ativam quando o potencial de membrana se torna hiperpolarizado a  $-50$ mV) e também pelo influxo de  $Ca^{2+}$  por canais que também se abrem espontaneamente por volta do potencial  $-55$ mV. A frequência com que estas células marcapasso disparam, portanto, pode variar de acordo com a velocidade de repolarização da fase 3 e de despolarização da fase 4; com a negatividade máxima da hiperpolarização durante

a fase 4; e com o valor do limiar do potencial de ação. Estes parâmetros são modificados principalmente pelo sistema nervoso autônomo. Fisiologicamente, a partir do nodo sinoatrial, o impulso cardíaco espalha-se radialmente através do átrio direito. O feixe de Bachmann conduz o impulso do nodo sinoatrial diretamente para o átrio esquerdo. O impulso alcança então o nodo atrioventricular, que é normalmente a única porta de entrada para os ventrículos (TERRENOIRE et al., 2005).

A FA ocorre quando modificações estruturais e/ou eletrofisiológicas alteram o tecido atrial, promovendo formação e propagação anormal de impulsos cardíacos pelos átrios. Estas anormalidades são causadas por diversos mecanismos patofisiológicos, muitos ainda não compreendidos. Atualmente, sabe-se da interação entre três mecanismos principais que contribuem para o desenvolvimento da FA: o remodelamento elétrico e da dinâmica do ( $\text{Ca}^{2+}$ ) e o remodelamento estrutural dos átrios. Estes fatores interagem - entre si e com outras comorbidades cardíacas - e se retroalimentam, promovendo a ocorrência de despolarizações ectópicas (que tem início fora do nodo sinoatrial) e reentradas do impulso cardíaco, principais responsáveis pelo início, manutenção e cronificação da FA (NATTEL; HARADA, 2014).

As despolarizações ectópicas nos átrios ocorrem majoritariamente em função das chamadas pós-despolarizações precoces (PDPs) - que ocorrem entre a fase 2 e meio da fase 3 do potencial de ação - e pós-despolarizações tardias (PDTs) - que ocorrem entre o fim da fase 3 e início da fase 4. PDPs ocorrem mais facilmente quando ocorre um prolongamento na duração do potencial de ação, enquanto PDTs estão mais ligados às anormalidades da dinâmica do  $\text{Ca}^{2+}$  (WAKILI et al., 2011). A ocorrência de reentradas estão ligadas ao remodelamento tanto elétrico como estrutural dos átrios, os quais propiciam a propagação anormal e o atraso do impulso cardíaco por tempo suficiente para que regiões do músculo recém excitadas recobrem a excitabilidade (saída do período refratário absoluto) (DOBREV; NATTEL, 2010).

As propriedades eletrofisiológicas dos átrios são mantidas por canais iônicos, bombas e trocadores de membrana, e o remodelamento elétrico dos átrios pode afetar a dinâmica de qualquer um desses componentes, favorecendo o aparecimento de despolarizações ectópicas e reentradas. Na FA, os principais componentes afetados pelo remodelamento elétrico são os canais lentos de  $\text{Ca}^{2+}$  (*L-type*), os canais de  $\text{K}^+$  responsáveis pela repolarização final da fase 3 do potencial de ação (*inward rectifier channels*), os canais de  $\text{K}^+$  dependentes de acetilcolina e as conexinas.

Durante a fase 2 do potencial de ação nos átrios, o  $\text{Ca}^{2+}$  entra nos cardiomiócitos através dos canais lentos (*L-type*) de  $\text{Ca}^{2+}$  fazendo com que ainda mais  $\text{Ca}^{2+}$  seja liberado do

interior do retículo sarcoplasmático (RS) dessas células através dos receptores de rianodina (RyR2s). O  $\text{Ca}^{2+}$  liberado promove a contração dos cardiomiócitos e, então, é sequestrado novamente para dentro do RS através da bomba SERCA2a - que, fisiologicamente, tem uma parcela de sua ação inibida pela proteína fosfolambam (PLB) - ou removido da célula pela bomba antiporte  $3\text{Na}^+/1\text{Ca}^{2+}$  (NCX) (WAKILI et al., 2011). Durante estimulação adrenérgica pelo sistema simpático ou adrenal, a calmodulina cinase tipo 2 (CaMKII) e a proteína quinase A tornam-se ativas, fosforilando os canais de  $\text{Ca}^{2+}$  *L-type* (que permitem maior entrada de  $\text{Ca}^{2+}$ ), RyR2s (que passa a liberar mais  $\text{Ca}^{2+}$  do RS) e PBL (que deixa de inibir a SERCA2a, bombeando e acumulando mais  $\text{Ca}^{2+}$  no RS), permitindo maior força e velocidade de contração cardíaca. Embora adaptativo em situações de estresse que demandam mais do sistema cardiovascular, a liberação continuada de mais  $\text{Ca}^{2+}$  para o citoplasma tende a estimular vazamentos diastólicos da RyR2. A ativação, então, da NCX para bombear o excesso de  $\text{Ca}^{2+}$  para fora da célula causa um influxo de  $\text{Na}^+$  em troca da extrusão do  $\text{Ca}^{2+}$  que pode vir a causar PDTs, dando início a uma despolarização ectópica e, conseqüentemente, a um episódio de FA. Com o passar do tempo, a maior frequência de contração atrial presente na FA suprime a expressão dos canais *L-type* de  $\text{Ca}^{2+}$ , resultando no menor influxo de  $\text{Ca}^{2+}$  e redução da duração do potencial de ação, facilitando, por sua vez, episódios de reentradas (NATTEL; DOBREV, 2012). Algumas outras outras modificações no que diz respeito à dinâmica do  $\text{Ca}^{2+}$  observadas na FA são a superexpressão do trocador NCX (propiciando PDTs) e a superativação da proteína cinase A (hiperfosforilando canais de  $\text{Ca}^{2+}$  *L-type*, RyR2 e PBL, culminando na ocorrência do descrito acima) (YUE; XIE; NATTEL, 2011); e o aumento de 10 vezes da expressão da proteína inibidora da proteína fosfatase 1, a qual, por sua vez, desfosforila os canais de  $\text{Ca}^{2+}$ , RyR2 e PBL (PATHAK et al., 2005).

Reentradas também são facilitadas pela maior expressão de canais de  $\text{K}^+$  (*inward rectifier channels* e canais dependentes de acetilcolina) verificadas na FA, uma vez que o maior efluxo de  $\text{K}^+$  durante a repolarização diminui a duração do potencial de ação e, por consequência, o período refratário absoluto das células, propiciando que possam despolarizar-se novamente mais rápido, culminando em possíveis conduções anômalas do impulso cardíaco (PANDIT et al., 2005). Outro componente celular afetado pela fibrilação atrial são as conexinas. As conexinas formam as junções *gap*, responsáveis pela comunicação elétrica (passagem de íons  $\text{Na}^+$ ) entre os cardiomiócitos. Na FA, a condução elétrica entre as células é diminuída, o que favorece o remodelamento atrial e reentradas. Além disso, diversas mutações somáticas envolvendo genes de conexinas já foram relacionadas à FA (CHALDOUPI et al., 2009).

O remodelamento estrutural dos átrios - especialmente a fibrose intersticial - também é importante no início, cronificação e resistência farmacológica da FA, como diversos estudos têm demonstrado (BURSTEIN; NATTEL, 2008). A fibrose promove FA interrompendo feixes condutores do impulso cardíaco e causando distúrbios de condução locais. Além disso, interações entre cardiomiócitos e fibroblastos podem causar mudanças arritmogênicas na dinâmica elétrica dos cardiomiócitos (YUE; XIE; NATTEL, 2011). Contudo, os mecanismos que medeiam o remodelamento da matriz extracelular dos átrios não são completamente compreendidos e podem variar, por exemplo, de acordo com outras comorbidades cardíacas concomitantes à FA, uma vez que fibrose é desfecho de diversas delas (OAKES et al., 2009). Diversos fatores secretados tem ação pró-fibrótica. Entre eles, a angiotensina II e o TGF $\beta$ 1 (fator de crescimento transformante  $\beta$ 1) são fatores bem estabelecidos, enquanto evidências mais recentes sugerem também papel do fator de crescimento derivado de plaquetas (PDGF) (BURSTEIN et al., 2007).

A angiotensina II tem seus níveis aumentados na FA e medeia fibrose cardíaca em diversas patologias, incluindo a hipertensão, a insuficiência cardíaca e o infarto (XIAO et al., 2004). Age ligando-se a dois tipos de receptores: tipo 1 (AT1R) e tipo 2 (AT2R). A via de sinalização mediada pelo AT1R é pró-fibrótica e estimula a proliferação de fibroblastos, hipertrofia dos cardiomiócitos e apoptose através da ativação da ERK 1 e 2 (extracellular signal-related kinase) e da JNK (c-Jun N-terminal kinase), as quais ativam diversos fatores de transcrição (Elk-1, c-jun, and c-fos), modulando a expressão gênica. A via do AT1R também estimula a fosfolipase C, culminando na ativação da proteína cinase C e na liberação intracelular de Ca<sup>2+</sup>, que também estimulam o remodelamento estrutural e elétrico dos átrios (HUNYADY; CATT, 2006). O TGF $\beta$ 1 é secretado pelos cardiomiócitos e fibroblastos sob ação da angiotensina II e tem ação autócrina e parácrina, estimulando ainda mais a produção de angiotensina II (*feedback* positivo), a ativação de fibroblastos e a deposição de colágeno (ROSENKRANZ, 2004). A superexpressão de TGF- $\beta$ 1 causa FA através de distúrbios de condução do impulso cardíaco em modelo animal (VERHEULE et al., 2004). O PDGF, por sua vez, estimula a proliferação e diferenciação de fibroblastos. Seus receptores dão início a uma cascata de sinalização que ativa os mesmos fatores de transcrição da cascata da angiotensina II. A cascata ativada pelo PDGF parece ser muito mais ativa nos átrios, o que explica em parte a maior susceptibilidade dos átrios à fibrose do que os ventrículos (BURSTEIN et al., 2008).

Apesar dos importantes avanços na compreensão acerca dos mecanismos patofisiológicos da FA nos últimos anos, ainda há muitos outros a serem elucidados.

Processos como dilatação dos átrios, edema e amiloidose são todos observados na FA, mas sua importância no desenvolvimento da doença segue indefinida. Além disso, as interações físicas e via sinalização celular entre cardiomiócitos e fibroblastos têm papel obscuro na iniciação e cronificação da FA e estão sendo foco recente de pesquisa (YUE; XIE; NATTEL, 2011). Recentemente, uma vez que, no estudo da FA, modelos animais apresentam importantes limitações - como o fato de não ocorrer FA espontânea nos animais modelo como ocorre em humanos e também o fato de terapias efetivas em animais não serem em humanos -, diversos métodos em bioinformática - análise de transcriptômica, de proteômica e de biologia de sistemas - também têm sido empregados para integrar grandes quantidades de dados no intuito de obter novos *insights* envolvendo as complexas vias de sinalização presentes na FA, sendo importantes também para o desenvolvimento de novas estratégias terapêuticas (NISHIDA et al., 2010). A combinação de diversos fatores patofisiológicos e sua interação com outras doenças cardíacas é possivelmente a chave para o avanço de nosso conhecimento biológico e terapêutico acerca da FA.

## 1.2 TRANSCRIPTOMA

O transcriptoma se refere a soma de tudo que é transcrito em células ou tecidos e baseia-se na quantificação relativa de todas as moléculas de RNA neles expressas em um dado momento ou condição. Nesse contexto, a quantificação das moléculas de RNA, principalmente RNA mensageiro, permite que a expressão dos genes seja mensurada e avaliada de forma indireta. O estudo da expressão dos genes, por sua vez, fornece uma quantidade enorme de informações a respeito da biologia do alvo ou da condição experimental em estudo (DONG; CHEN, 2013). A posterior análise desses dados de expressão nos permite, então, identificar genes diferencialmente expressos e outros padrões de expressão gênica; elucidar mecanismos por trás da regulação dos genes; identificar as vias biológicas enriquecidas; obter informações acerca da co-expressão dos genes entre outras informações relevantes (LOWE et al., 2017). Há mais de uma década, as informações obtidas por meio de experimentos de transcriptoma têm sido usadas para alavancar o progresso das ciências biológicas e biomédicas, com destaque para a descoberta de mecanismos importantes envolvendo doenças multifatoriais complexas de humanos, como diferentes tipos de câncer, doenças cardiovasculares, doenças neurológicas, entre muitas outras (GUINDALINI; TUFIK, 2007).

### 1.2.1 O Microarranjo

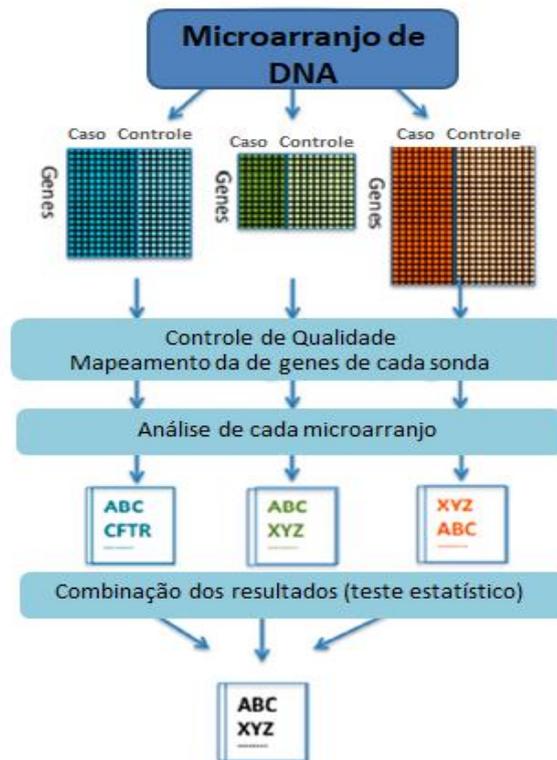
Os tipos de experimentos mais comuns envolvendo o transcriptoma são o RNA-seq e o microarranjo. Este trabalho foca na análise de dados provenientes de experimentos de microarranjo. A tecnologia envolvida nos experimentos de microarranjo é capaz de quantificar simultaneamente todos os quase todos os transcritos conhecidos (e, dependendo da plataforma, até mesmo as isoformas dos transcritos) do genoma (LOWE et al., 2017). Chips de microarranjo consistem basicamente de milhares de sondas (oligonucleotídeos) sobre uma placa (chip) de vidro. A mensuração da expressão dos transcritos é determinada pela hibridização dos transcritos (que são marcados com fluoróforos) isolados dos tecidos de interesse com as sondas específicas para eles no chip de microarranjo. Após a hibridização entre os transcritos e as sondas, o chip é posto em um scanner que mede a fluorescência em cada ponto do chip. Assume-se, então, que a fluorescência de uma sonda é proporcional a quantidade de transcritos que hibridizam nela (e, portanto, ao nível de atividade do gene correspondente) (BARBULOVIC-NAD et al., 2006). Microarranjos podem ser *two-color* (geralmente mais antigos) e *one-color* (geralmente mais modernos). Ambos os tipos, porém, permitem que as mesmas análises sobre os dados sejam realizadas. As empresas Affymetrix, Agilent e Illumina são as principais fabricantes dos principais modelos de chips utilizados atualmente.

### 1.2.2 Metanálise de Dados de Transcriptoma provenientes de Microarranjo

Apesar de essenciais na elucidação de diversas questões biológicas ao longo de mais de uma década, dados de transcriptoma não estão livres de problemas metodológicos. Por serem experimentos caros e envolverem a mensuração da expressão de milhares de transcritos, dados de transcriptoma provenientes de microarranjo geralmente sofrem com problemas de reprodutibilidade e de baixo poder estatístico, muito em função do pequeno número de amostras envolvidas nos experimentos quando comparado às dezenas de milhares de sondas mensuradas (RAMASAMY et al., 2008). Nesse contexto, combinar os dados de diferentes estudos existentes é um passo lógico a seguir no sentido de mitigar esses reverses e permitir que o máximo de informação biologicamente relevante seja detectada. Atualmente, os dados de milhares de estudos envolvendo transcriptoma estão disponíveis gratuitamente

em bancos de dados como o Gene Expression Omnibus (CLOUGH; BARRETT, 2016). É comum, portanto, que os dados de múltiplos estudos envolvendo a mesma doença ou condição de interesse estejam disponíveis. Esses dados podem, portanto, ser submetidos a técnicas estatísticas integrativas de análise, configurando a metodologia de metanálise de dados de transcriptoma. Resumidamente, o processo de metanálise como um todo passa pela busca dos dados em repositórios digitais; o controle de qualidade e pré-processamento individual de cada estudo; análise individual de cada estudo; a combinação dos resultados individuais (metanálise propriamente dita) e, por fim, a análise dos resultados.

Diversas técnicas foram desenvolvidas ao longo dos anos no sentido de possibilitar a integração de dados de microarranjo. Em geral, elas subdividem-se de acordo com o parâmetro estatístico a ser integrado: valores de  $p$ , tamanhos de efeito ou *ranks*. As metodologias existentes também diferem no que diz respeito ao quão conservadores são no momento de apresentar um resultado estatisticamente significativo (CHANG et al., 2013). A escolha do método passa, portanto, pela análise da estrutura dos dados, a sua heterogeneidade e pelo objetivo final do estudo. No caso deste projeto, cujo objetivo é mais exploratório, uma metodologia muito conservadora poderia não detectar achados biológicos interessantes. Contudo, como a estrutura dos dados utilizados é razoavelmente heterogênea, metodologias pouco conservadoras poderiam resultar num excesso de resultados falsos-positivos. O método escolhido chama-se *r-th ordered p-value* - um método que faz uso de valores de  $p$  para integrar os dados (SONG; TSENG, 2014) e é considerado intermediário quanto ao seu conservadorismo (CHANG et al., 2013). O esquema resumido da metodologia de dados de microarranjo encontra-se na figura 2.



**Figura 2: Esquema metodológico de uma metanálise de dados de microarranjo.** Os estudos individuais selecionados passam por controle de qualidade e pré-processamento. Depois, são analisados individualmente para determinar expressão diferencial. Os resultados individuais são, então, combinados por método estatístico metanalítico, resultando em uma lista final de transcritos diferencialmente expressos que podem ser analisados. Adaptado de (WALSH et al., 2015).

### 1.3 JUSTIFICATIVA

A fibrilação atrial é o tipo de arritmia cardíaca mais comum e está associada à crescente morbidade cardiovascular e diminuição da qualidade de vida dos pacientes afetados. Além disso, muitos dos mecanismos envolvidos com a doença ainda não são conhecidos. Nesse sentido, dados de transcriptoma são frequentemente e historicamente usados na elucidação ou levantamento de hipóteses importantes sobre mecanismos biológicos envolvidos em doenças complexas. Contudo, como a análise individual de apenas um conjunto de dados geralmente sofre com importantes reverses metodológicos e estatísticos, a metanálise de um número maior de conjuntos de dados surge como alternativa e passo lógico a ser dado no sentido de obter dados mais confiáveis, fidedignos e reprodutíveis.

## 1.4 OBJETIVOS

Os objetivos deste trabalho são apresentados a seguir, divididos em objetivos gerais e objetivos específicos.

### 1.4.1 Objetivo geral

Realizar metanálise de dados de microarranjo envolvendo a fibrilação atrial e, sobre esses resultados, utilizar métodos de bioinformática para obter *insights* sobre o perfil de expressão gênica e mecanismos biológicos envolvidos com a doença.

### 1.4.2 Objetivos específicos

- Revisar a literatura envolvendo a doença fibrilação atrial;
- realizar busca de dados de microarranjo em repositórios digitais de estudos de transcriptoma;
- realizar controle de qualidade de análises apropriadas sobre o conjunto de dados brutos de todos os estudos elencados;
- realizar metanálise sobre os dados de microarranjo previamente analisados utilizando o método *r-th ordered p-value*;
- sobre os resultados da metanálise, utilizar metodologias de análise em bioinformática, como a análise de redes biológicas e análise de enriquecimento funcional;
- analisar e discutir apropriadamente os resultados biologicamente interessantes (genes-chave, perfil de expressão gênica ou vias biológicas, por exemplo) que possam agregar conhecimento sobre a fibrilação atrial.

## 2 ARTIGO CIENTÍFICO

*OBS: Este artigo não está finalizado e está sujeito a mudanças até sua publicação final.*

# METANÁLISE DE DADOS DE TRANSCRIPTOMA REVELA MÓDULOS PATOFIISIOLÓGICOS ENVOLVIDOS COM A FIBRILAÇÃO ATRIAL

**Rodrigo Haas Bueno, Mariana Recamonde-Mendoza**

## RESUMO

### Introdução

A fibrilação atrial é uma doença complexa afetando milhões de pessoas ao redor do mundo. Os mecanismos biológicos que a permeiam são complexos e precisam ser melhor compreendidos. Nesse sentido, realizamos uma metanálise de dados de transcriptoma para explorar esses mecanismos com resultados mais sensíveis e fidedignos.

### Métodos

Dados públicos de transcriptoma foram baixados, analisados quanto a sua qualidade e pré-processados individualmente. Sobre cada estudo individual, uma análise de expressão diferencial foi realizada e esses resultados foram, por fim, agregados metanaliticamente utilizando o método *r-th ordered p-value*. Analisamos a lista final de transcritos diferencialmente expressos por meio de análises rede (análises topológicas e de modularidade) e análise de enriquecimento funcional.

### Resultados

A metanálise resultou em 589 transcritos diferencialmente expressos, cuja rede de interação proteína-proteína apresentou 11 *hubs-bottlenecks* e 4 módulos funcionais principais. Estes módulos, por fim, enriqueceram para, respectivamente, 23, 54, 33 e 53 vias biológicas envolvidas com a patofisiologia da fibrilação atrial.

### Conclusões

Esta metanálise identificou transcritos diferencialmente expressos na fibrilação atrial, cuja rede de interação apresentou módulos funcionais relacionados aos processos de remodelamento estrutural e elétricos presentes na doença. Estresse no retículo endoplasmático, catabolismo proteico, estresse oxidativo e inflamação são alguns deles. Altamente conectados e com provável papel chave na regulação de processos envolvidos na patofisiologia da doença estão os *hubs-bottlenecks* como HSPA5, ANK2, CTNNA1 e VWF.

**Palavras-chave**

Análise de redes, enriquecimento funcional, fibrilação atrial, metanálise, microarranjo, transcriptoma

**BACKGROUND**

A fibrilação atrial (FA) é o tipo mais comum de arritmia cardíaca, afetando a qualidade de vida de milhões de pessoas ao redor do mundo. Com o envelhecimento da população, o número de diagnósticos de FA deve aumentar consideravelmente, uma vez que sua incidência aumenta com a idade (1). A FA está relacionada com o aumento de eventos tromboembólicos (derrames e doença cardíaca isquêmica) e tem alta prevalência em pacientes com outras doenças cardiovasculares ou metabólicas como a hipertensão, a insuficiência cardíaca ou diabetes mellitus (2). Tipicamente, apresenta-se como doença crônica e progressiva, e sua patofisiologia envolve modificações estruturais e eletrofisiológicas do tecido atrial que acabam por promover a formação e propagação de impulsos anormais pelos átrios (3).

Apesar do avanço sobre o conhecimento dos mecanismos por trás da FA, ainda há inúmeras e importantes lacunas a serem preenchidas (4). Nesse sentido, há mais de uma década, a análise de dados de microarranjo tem se mostrado útil na identificação de genes e mecanismos biológicos envolvidos em doenças complexas como câncer, doenças cardiovasculares e neurológicas, e há quantidades enormes desses dados depositados em banco de dados de forma gratuita. Contudo, a análise de dados provenientes de microarranjo geralmente sofre com a baixa reprodutibilidade dos resultados e com baixo poder estatístico em função do pequeno número amostral dos estudos em relação às milhares de sondas mensuradas. Por isso, combinar os resultados de diferentes estudos através de métodos de metanálise surge como alternativa para mitigar os problemas metodológicos envolvendo esse tipo de dado. Neste estudo, realizamos uma metanálise de dados de transcriptoma envolvendo amostras de tecido cardíaco com fibrilação atrial e lançamos luz a possíveis novos mecanismos patofisiológicos da doença, bem como a novos alvos terapêuticos.

**MÉTODOS**

Neste estudo metanalítico, os seguintes passos foram realizados: 1) busca manual de dados públicos de transcriptoma; 2) Pré-processamento dos dados de transcriptoma e análise

de expressão diferencial; 3) Metanálise dos dados de transcriptoma; 4) Análise de redes biológicas; 5) Interpretação dos resultados. Os itens 1 e 2 são componentes indiretos da análise, visando a avaliação conjunta de múltiplos dados de expressão. As etapas 2 e 3 foram implementadas no ambiente R (versão 3.6.0 para Windows) (5). A etapa 4 foi implementada no software cytoscape (versão 3.7.1) (6), utilizando-se de aplicativos específicos de análise. Para a etapa 5, o software Cytoscape (versão 3.7.1), bem como ferramentas web de domínio público já bem consolidadas para as respectivas análises foram utilizados.

### **Busca manual de dados públicos**

A busca por dados de transcriptoma foi feita na base de dados públicos Gene Expression Omnibus (GEO). As palavras-chave utilizadas incluíram os termos “atrial fibrillation”, “arrhythmia”, “heart”, “cardiac”, “disease”, “gene expression profiling”, “transcriptome” e seus respectivos entre termos. A seleção se deu pelos seguintes critérios: a) estudos transcriptômicos baseados nas técnicas de microarranjo; b) realizado em humanos; c) com grupos e amostras claramente descritos e com no mínimo três amostras por grupo; d) estudos que dispunham dos dados brutos; e) estudos com data posterior a 2001, priorizando-se estudos publicados após 2007 em razão da melhor precisão de tecnologias mais recentes. A busca foi realizada em abril de 2019.

### **Pré-processamento dos dados de transcriptoma**

**Avaliação da qualidade dos dados brutos:** a avaliação da qualidade dos dados brutos é importante para as análises posteriores dos dados de microarranjo, uma vez que aumenta a capacidade de detecção de genes diferentemente expressos (7). O processo de avaliação da qualidade dependeu do tipo de plataforma:

1) Plataforma Affymetrix: para chips da affymetrix, o pacote *arrayQualityMetrics* para o R foi utilizado (8). Este pacote computa, a partir dos dados brutos, 6 métricas comumente utilizadas para a avaliação da qualidade de dados brutos de microarranjo, a constar: *Distances between arrays*, *Principal Component Analysis (PCA)*, *Boxplots*, *Density plots*, *Standard deviation versus rank of the mean*, e *MA-plots*. Em nosso estudo, amostras consideradas outliers pelo software em pelo menos duas métricas foram excluídas.

2) Plataforma Agilent: para chips da Agilent, a avaliação da qualidade se deu a partir da avaliação de *MA-plots*, *Background intensities boxplots* e *PCA*, como preconizado pelo

*Limma user`s guide* (9). Amostras extremamente desviantes em qualquer uma das métricas foram excluídas.

**Normalização:** a normalização tem por objetivo ajustar os efeitos sobre os dados provenientes de variações técnicas e ambientais como reações de hibridização ou marcação errôneas; problemas físicos ou estruturais; problemas com os reagentes ou com as condições laboratoriais e, por fim, com o escaneamento dos chips hibridizados. Estas correções permitem que as amostras de um mesmo estudo sejam comparáveis entre si. O processo de normalização neste estudo dependeu do tipo de plataforma:

1) Plataforma Affymetrix: os dados brutos provenientes da base de dados foram importados para o R e processados pelo algoritmo RMA (*Robust Multi-array Average*) (10) através do pacote *oligo* (11). Este algoritmo realiza etapas distintas de pré-processamento: correção de background, transformação dos dados por aplicação de logaritmo de base 2, normalização pelo método *quantile* e sumarização das sondas (*probesets*) em um valor único de expressão para cada gene. A anotação sonda - *gene symbol* foi feita através de pacotes de anotação específicos para cada modelo de chip disponíveis no repositório *Bioconductor*. Sondas com anotação para mais de um gene ou que não possuíam anotação foram excluídas da análise. Para modelar e corrigir possíveis *batch effects* (fonte de variação proveniente principalmente de heterogeneidade técnica - diferentes espaços temporais de experimentação, diferentes lotes de reagentes, etc), foi aplicado sobre os dados o algoritmo de correção *SVA* (*surrogate variable analysis*) (12).

2) Plataforma Agilent: os dados brutos provenientes da base de dados foram importados para o R e as seguintes etapas de pré-processamento foram realizadas através do pacote *limma*: correção de background (algoritmo *normexp*, método *mle*, *offset* de 50) (13,14) e normalização *between-array* pelo método *quantile* (15). A anotação sonda - *gene symbol* foi feita a partir de tabelas fornecidas pela empresa responsável pela plataforma. Sondas com anotação para mais de um gene ou sem anotação foram excluídas da análise. Para a modelagem de *batch effects*, o algoritmo *SVA* também foi utilizado.

### **Análise de expressão diferencial**

Para a obtenção dos valores de  $p$ , provenientes de testes de hipótese acerca da expressão diferencial de cada gene em cada um dos *datasets* posteriormente utilizados na metanálise, o protocolo bem estabelecido de análise de expressão diferencial do pacote *limma*

foi utilizado. Resumidamente, a ideia foi criar um modelo linear a partir dos dados de expressão para cada gene e utilizá-los para realizar um teste-*t* “moderado” (16), que estabiliza e aumenta o poder estatístico das análises para experimentos com baixo número amostral, como costumam ser os estudos que fazem uso de microarranjo.

### **Metanálise**

A metanálise dos dados de transcriptoma foi feita com o pacote *MetaDE* para o ambiente R (17). Foram utilizados na metanálise os genes em comum em no mínimo 80% dos estudos individuais, os valores de *p* calculados para cada gene em cada estudo e a direção de efeito (sinal) obtida a partir do tamanho de efeito estimado como descrito em (18) para cada gene de cada estudo. O método escolhido de metanálise foi o *r-th ordered p-value* (roP). O parâmetro *r-th* refere-se, de forma simplificada, à proporção de *datasets* em que um gene deve ter sido considerado diferencialmente expresso para que o esse método de metanálise o considere como tal (19). Neste estudo, foi aplicado parâmetro *rth* igual a 0.5 (19). Os valores de *p* finais obtidos para cada gene foram corrigidos pelo método FDR (*false discovery rate*) (20). Apenas os genes considerados diferencialmente expressos (FDR<0,05) pelo método roP e com direção de efeito divergente em no máximo 30% dos estudos que continham informações sobre o gene foram considerados para as análises posteriores (figura 1).

### **Análise de redes**

Com a lista de genes com direções de efeito semelhantes e considerados diferencialmente expressos pela metanálise, uma busca com parâmetro de confiabilidade alta (0.7) foi feita no STRING (21). A rede PPI (*protein-protein interaction*) básica foi importada para o aplicativo Cytoscape (6), à qual adicionou-se informações complementares a respeito da regulação (*up* ou *down-regulated*) de cada gene. Para avaliação dos parâmetros topológicos da rede, o plugin NetworkAnalyzer (22) do Cytoscape foi utilizado. Duas medidas de centralidade foram avaliadas para determinar os nós centrais da rede: grau e *betweenness*. Enquanto o grau quantifica o número de conexões de um nó, o *betweenness* reflete número de caminhos mínimos entre dois nós da rede que passam por um determinado nó. Por fim, a análise de clusterização para identificação de nodos altamente conectados dentro da rede – módulos funcionais - foi realizada através do aplicativo MCODE (23) para o Cytoscape. Este algoritmo identifica na rede-alvo regiões densamente conectadas, candidatas a representar complexos de interação funcionais.

## **Análise de enriquecimento funcional no GO**

A análise de enriquecimento funcional no GO (24,25) – com ontologia *biological process* - foi realizada sobre os genes dos quatro clusters identificados pelo MCODE com o maior score. Para realizá-la, o pacote clusterProfiler para o R foi utilizado e o teste estatístico utilizado foi o *over-representation test*.

## **RESULTADOS**

### **Dados incluídos na metanálise**

De acordo com os critérios de inclusão, 8 estudos foram selecionados para esta metanálise, com um dos estudos contendo dois *datasets*, totalizando nove *datasets* individuais analisados. Dos nove *datasets*, seis foram gerados com chips da plataforma Affymetrix e três com chips da plataforma Agilent. No total, os nove *datasets* continham 79 amostras de pacientes com fibrilação atrial e 75 amostras de pacientes com ritmo sinusal (controles). Amostras de tecido atrial foram utilizados em oito *datasets*, enquanto apenas um *dataset* utilizou sangue como amostra. As informações relevantes acerca dos estudos selecionados para a metanálise estão sumarizadas na tabela 1 (número de amostras inclui apenas amostras que passaram no controle de qualidade dos dados brutos). Foi feito o *download* dos dados brutos dos *datasets* selecionados e os mesmos foram importados para a plataforma R para as análises posteriores.

### **Controle da qualidade dos dados brutos**

A partir da análise da qualidade dos dados brutos como descrito na metodologia, amostras de qualidade duvidosa foram retiradas das análises posteriores na busca por resultados mais sensíveis, a constar: GSE2240 – GPL96, 6 amostras; GSE2240 – GPL97, 5 amostras; GSE75092, 1 amostra; GSE41177, 4 amostras e GSE14975, 1 amostra.

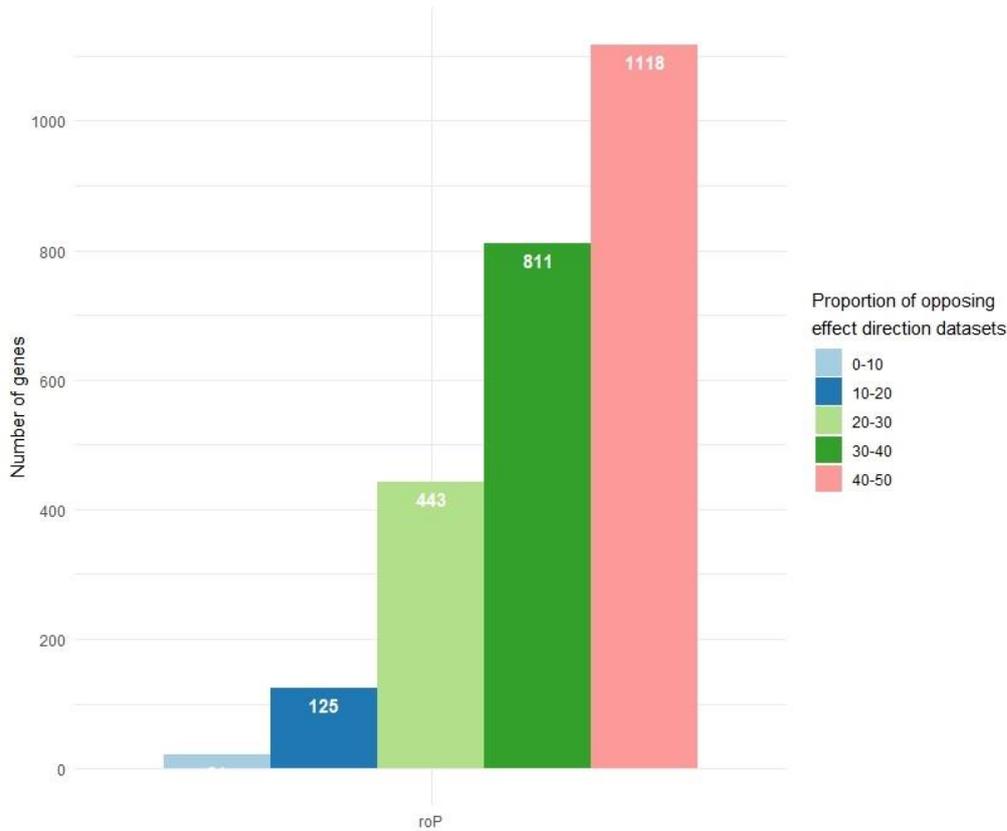
GEO accession number	Sample size		Sample type	Plataform	Publication
	Atrial fibrillation	Sinus rhythm			
GSE2240 - GPL96	10	19	Atrial myocardium	Affymetrix Human Genome U133A Array	May 13, 2005
GSE2240 - GPL97	10	20	Atrial myocardium	Affymetrix Human Genome U133B Array	May 13, 2005
GSE31821	4	2	Atrial myocardium	Affymetrix Human Genome U133 Plus 2.0 Array	Apr 20, 2018
GSE113013	5	5	Atrial appendage tissue	Agilent-045997 Arraystar human lncRNA microarray V3	Apr 12, 2018
GSE62871	7	9	Right atrial myocardium	Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray	Oct 30, 2014
GSE75092	5	3	Blood	Agilent-045997 Arraystar human lncRNA microarray V3	Jul 31, 2017
GSE76899	6	6	Myocardial sleeves of pulmonary veins	Affymetrix Human Transcriptome Array 2.0	Jan 16, 2016
GSE41177	28	6	Left atrial appendage	Affymetrix Human Genome U133 Plus 2.0 Array	Sep 18, 2013
GSE14975	4	5	Left atrial myocardium	Affymetrix Human Genome U133 Plus 2.0 Array	Feb 11, 2010

**Tabela 1 - Característica dos dados selecionados para metanálise.**

### **Análise individual dos datasets e metanálise**

Após o pré-processamento dos dados brutos e a análise de expressão gênica diferencial, como descrito na metodologia, o resultado foram nove listas (uma lista para cada *dataset*) de genes com seus respectivos valores de  $p$  (teste de hipótese acerca da expressão diferencial de cada gene) e direções de efeito. Como as diferentes plataformas diferem no número de transcritos mensurados, realizamos a metanálise apenas sobre os dados de 14582 genes representados em no mínimo 80% dos *datasets* (sete dos nove).

Existem diversas técnicas estatísticas que possibilitam a integração de dados de microarranjo. Elas diferem no parâmetro estatístico de que fazem uso e no que diz respeito ao quão conservador são seus resultados (26). De forma geral, a escolha do método se baseia nos objetivos e hipóteses do estudo e no quão heterogêneo são os dados a serem utilizados (27). O objetivo de natureza exploratória deste estudo seria limitado por métodos muito conservadores; contudo, o uso de diferentes plataformas de microarranjo, bem como diferentes tipos de amostras biológicas agregam considerável heterogeneidade aos dados. Nossa escolha foi, portanto, o uso do método o *r-th ordered p-value* (roP), considerado robusto, consistente e apenas moderadamente conservador pelos testes realizados em revisão de métodos de metanálise (26). Nosso parâmetro  $r$ -th de escolha foi 0.5, ou seja, 50% dos *datasets* contendo dados sobre um determinado gene devem considerá-lo diferencialmente expresso de forma significativa para que no resultado final da metanálise esse gene seja considerado diferencialmente expresso. Em nossa metanálise, 2518 genes foram considerados diferencialmente expressos no resultado final. Entretanto, muitos genes apresentaram direções de efeito – *up* ou *downregulation* – destoantes entre os *datasets*. Nesse contexto, permaneceram para as análises posteriores 589 genes (material suplementar - tabela 1) – 356 *upregulated* e 233 *downregulated* - cuja direção de efeito discordou em no máximo 30% dos *datasets* que continham dados sobre eles (figura 1).



**Figura 1 - Proporção (%) de datasets com direções de efeito divergentes para os genes diferencialmente expressos identificados na metanálise:** Apenas genes com direções de efeito opostas em no máximo 30% dos estudos individuais contendo informação sobre eles seguiram para as análises posteriores. Isto é, apenas os genes incluídos nas barras azul clara, azul e verde claro foram selecionados, totalizando 589 genes.

### Análise de redes

A rede de interações proteína-proteína importada do STRING possuía 527 nodos (representando as proteínas codificadas pelos genes diferencialmente expressos; nodos não conectados à rede, contudo, foram excluídos) e 1750 arestas. A análise topológica de distribuição de grau revelou boa adequação ( $r\text{-squared} = 0.875$ ) ao modelo exponencial, caracterizando uma rede *scale-free*, como a maioria das redes biológicas (28). A intersecção entre os 15 nodos com maior grau e os 15 nodos com maior *betweenness* resultou em 11 *hubs-bottlenecks*, os quais são elementos chaves na análise funcional de uma rede, uma vez que tendem a ser proteínas funcionalmente importantes e evolutivamente conservadas nos processos representados pela rede biológica (29). Destes, nove representam proteínas *upregulated* e 2 representam proteínas *downregulated* (tabela 2). Outra característica importante das redes biológicas são seus módulos funcionais, cuja identificação pode revelar

processos biológicos coordenados, uma vez que proteínas que ocorrem num mesmo cluster tendem a enriquecer para funções biológicas em comum (30). Da rede principal, 13 clusters funcionais emergiram a partir do programa MCODE. Os quatro módulos com maior score – cluster 1: 7.74; cluster 2: 5.96; cluster 3: 5.72 e cluster 4: 4.90 – contêm, respectivamente, 132, 98, 162 e 94 genes.

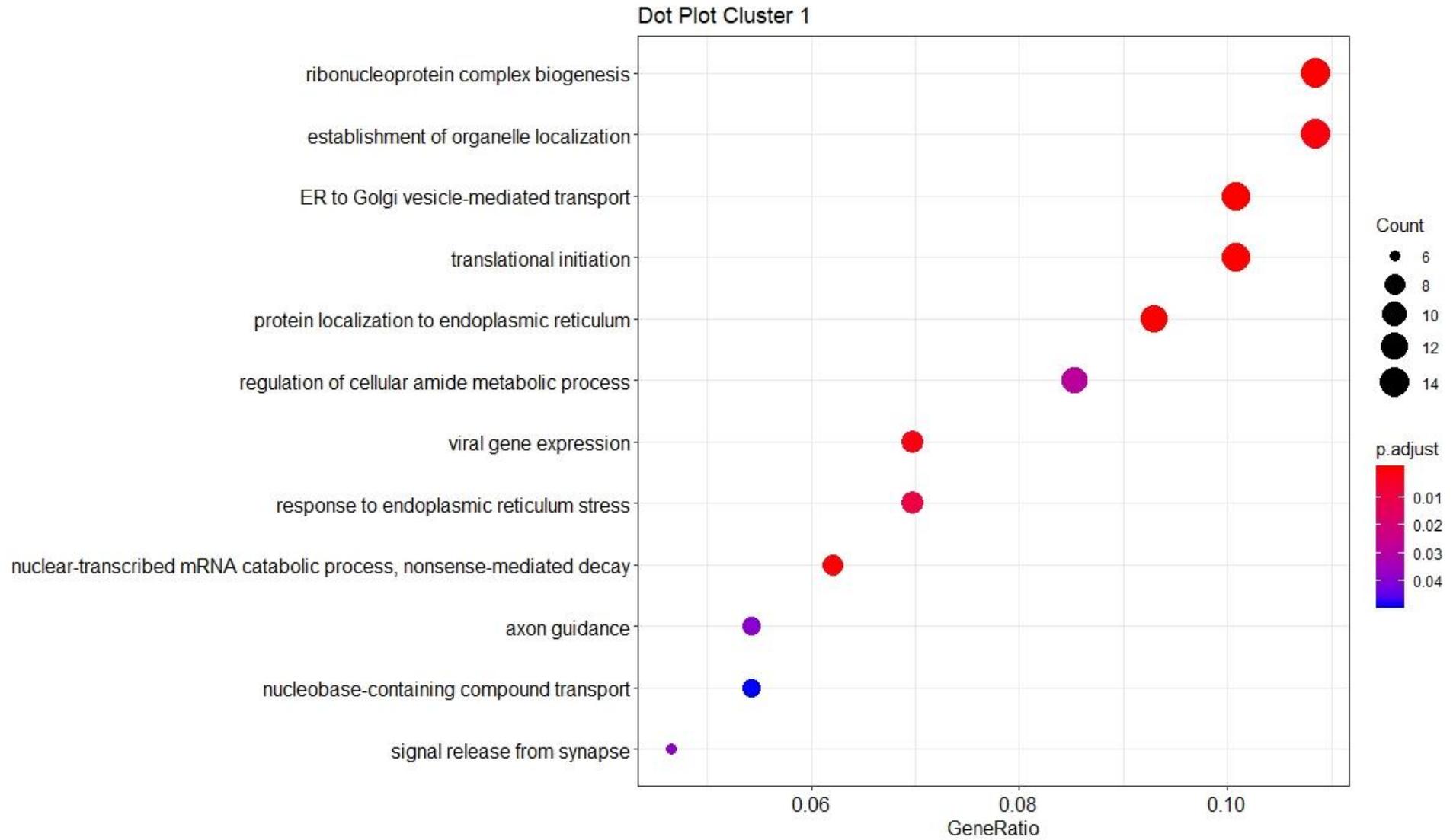
Protein - node	BetweennessCentrality	Degree	Regulation
SRSF1	0.05092252	23	Upregulated
NPM1	0.05727506	25	Upregulated
CLTC	0.03537033	24	Upregulated
CCNB1	0.04814839	30	Upregulated
BDNF	0.04563939	28	Upregulated
ANK2	0.02687843	22	Downregulated
GART	0.06351884	29	Upregulated
VWF	0.02946177	22	Upregulated
HSPA5	0.09752871	38	Downregulated
NRAS	0.04645906	34	Upregulated
CTNNB1	0.09820713	45	Upregulated

**Tabela 2 - Hubs-bottlenecks identificados na rede principal de transcritos diferencialmente expressos.** Devido à suas propriedades topológicas, representam candidatos interessantes a possuírem papel de destaque na patofisiologia da FA.

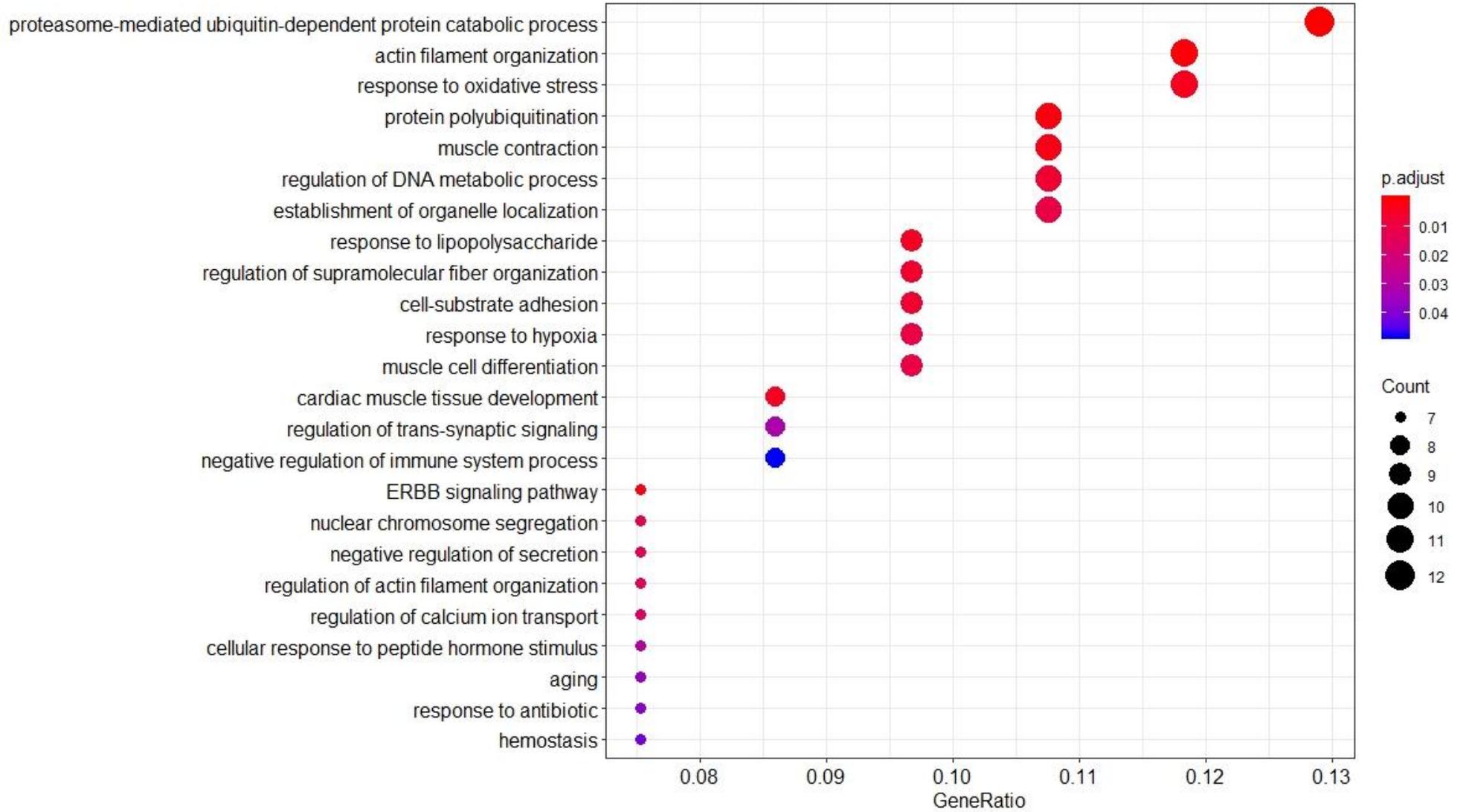
### Enriquecimento funcional dos clusters

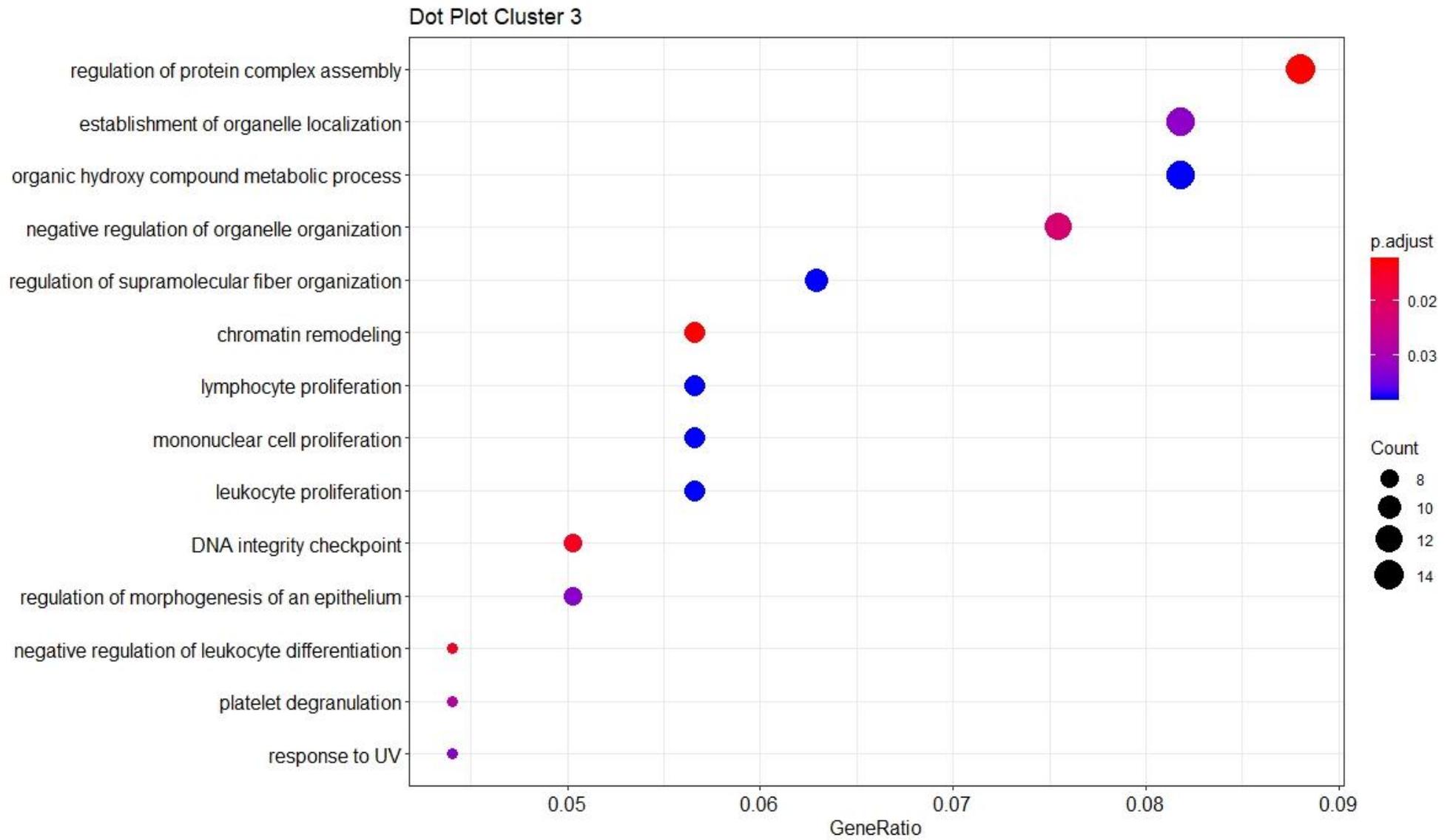
Para identificar as vias enriquecidas pelos módulos na rede, o *over-representation test* foi aplicado sobre os constituintes de cada módulo utilizando a base de dados do GO – Biological Process. De forma estatisticamente significativa ( $FDR < 0,05$ ), os módulos 1, 2, 3 e 4 enriqueceram, respectivamente, para 2832, 3367, 3657 e 2796 vias biológicas. Para simplificar a interpretação dos resultados, um algoritmo (31) de sumarização de termos redundantes dos nomes das vias foi utilizado através do pacote para o R ClusterProfiler (32). Dessa forma, o número final de vias enriquecidas e sumarizadas para os módulos 1, 2, 3 e 4 foram, respectivamente, 23, 54, 33 e 53. Para eliminar a presença de vias enriquecidas com poucos genes em relação ao total de genes componentes dos respectivos módulos (*gene ratio*), foram excluídas as vias de cada módulo com *gene ratio* menor que o trigésimo percentil da

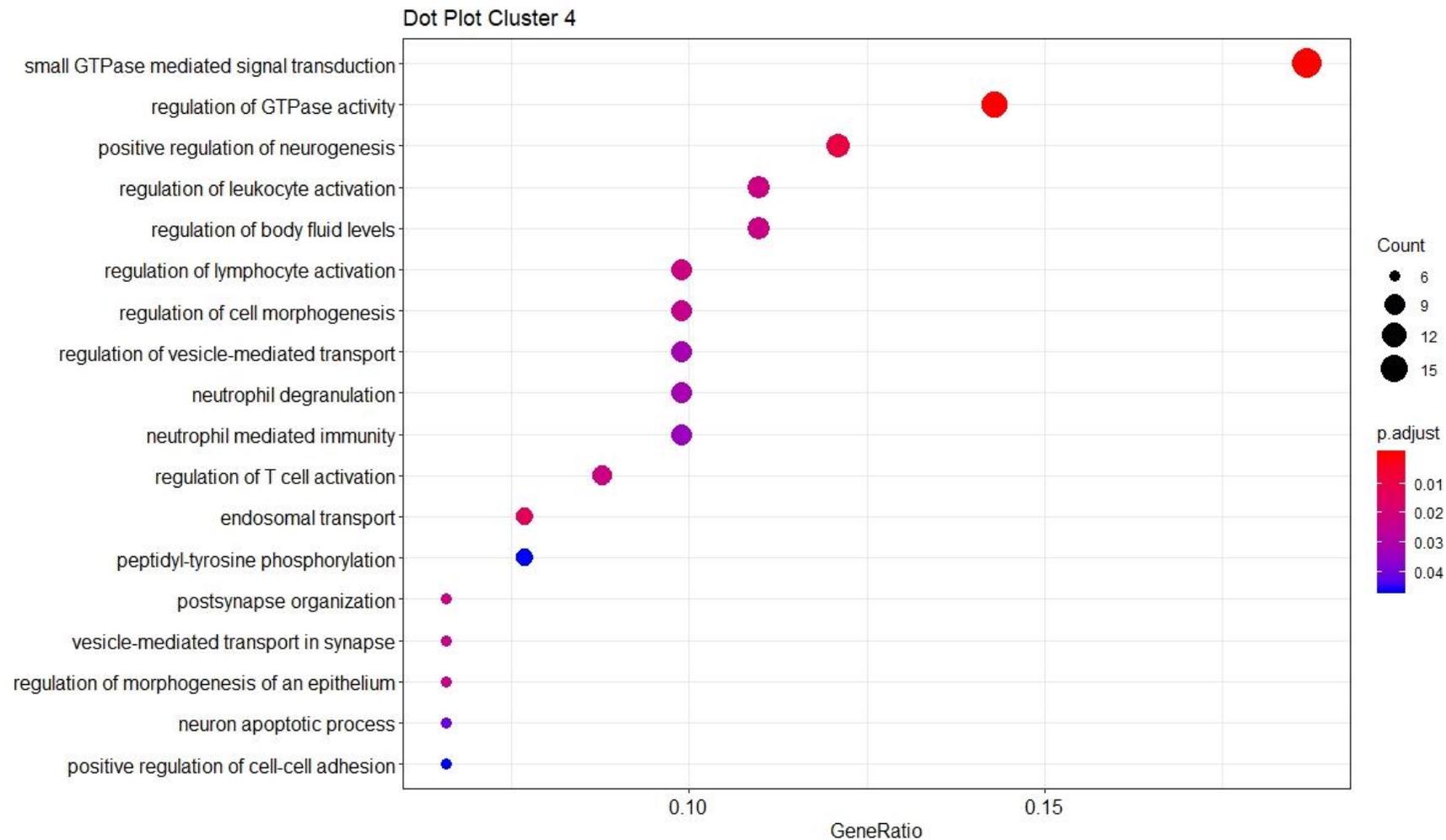
distribuição de *gene ratios* de cada módulo. A relação final de vias enriquecidas para cada módulo, seus respectivos *gene ratios* e valores de p ajustados estão na figura 2. A relação de todas as vias enriquecidas por cluster e os transcritos desregulados pertencentes a elas estão no material suplementar - tabela 2.



Dot Plot Cluster 2







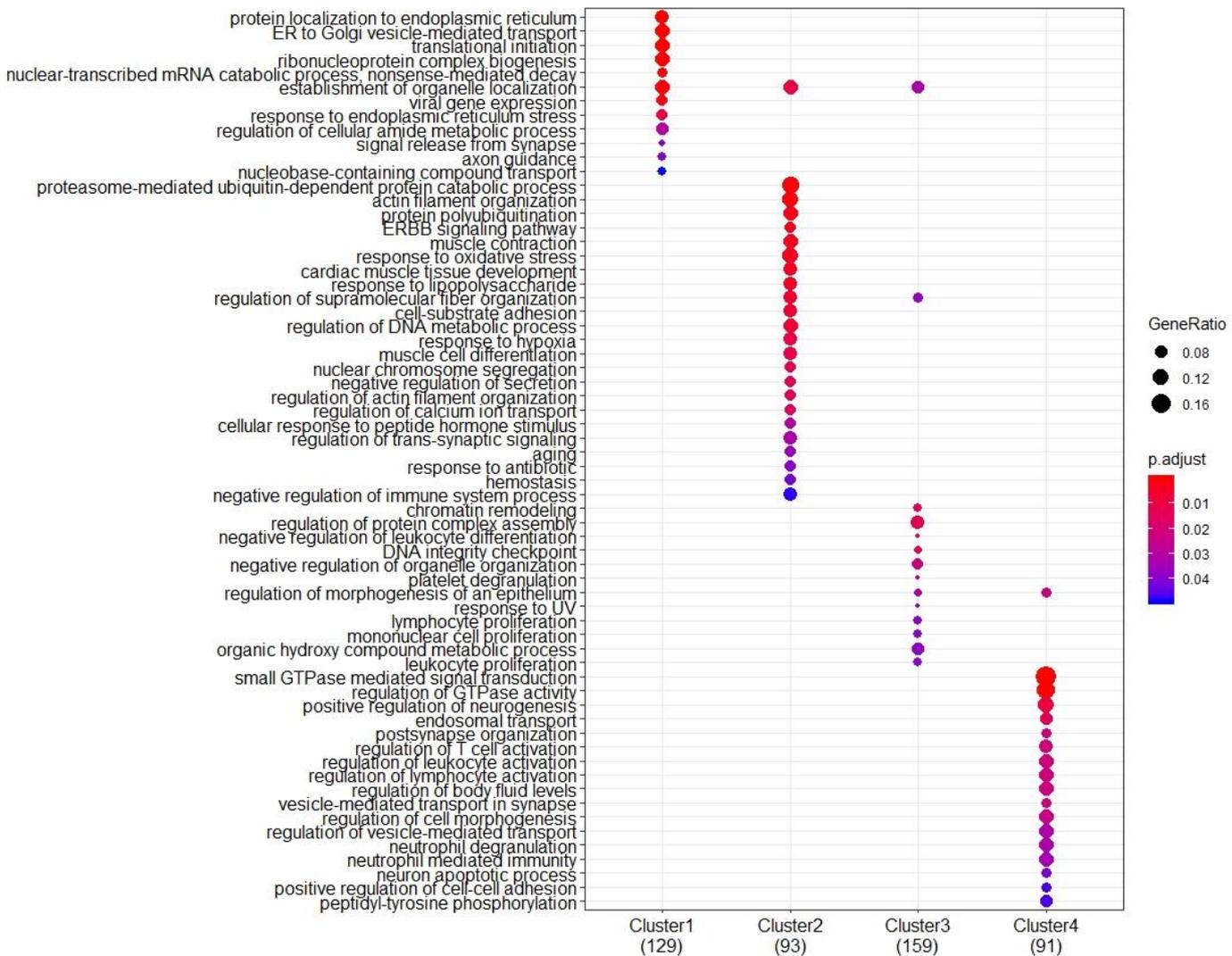
**Figura 2 - Dot plots dos quatro módulos principais enriquecidos pela análise de clusterização.** Vias biológicas enriquecidas pelos respectivos transcritos diferencialmente expressos que os compõem à esquerda. Tamanho do círculo representa o número de transcritos enriquecendo para a via biológica e a cor do círculo representa a significância estatística. GeneRatio é a proporção entre o número de transcritos que enriqueceram para determinada via e o número total de transcritos que pertencem ao respectivo módulo.

## DISCUSSÃO

A FA é uma doença complexa e progressiva que envolve múltiplas vias de remodelamento elétrico e estrutural do tecido atrial, as quais culminam no surgimento de atividade ectópica e condução disfuncional do impulso cardíaco nos átrios. Para avançar no conhecimento a respeito dos mecanismos que a permeiam, este estudo se utilizou de dados públicos de transcriptoma e realizou uma metanálise, obtendo uma lista de 589 transcritos diferencialmente expressos de forma consistente entre os conjuntos de dados individuais. A rede de interação entre os produtos proteicos desses transcritos, por sua vez, revelou módulos funcionais cujos processos biológicos enriquecidos representam processos importantes e inter-relacionados no contexto tecidual e patofisiológico da FA.

Os processos biológicos enriquecidos para os produtos proteicos dos transcritos pertencentes aos diferentes módulos identificados na rede estão representados na figura 3. Como esperado, os módulos apresentam pouca sobreposição de vias biológicas, indicando um comportamento modular e funcional dos processos patofisiológicos envolvidos com a FA, mas ainda apresentando vias de interface entre os diferentes módulos, possíveis elos de comunicação entre eles. O módulo 1 aparece representando, entre outras, vias biológicas relacionadas à síntese proteica – início da tradução e biogênese de ribossomos – e à fisiologia do retículo endoplasmático (RE) – localização de proteínas, transporte de vesículas para o complexo de Golgi e resposta ao estresse do RE. As funções celulares de síntese e dobramento proteico e homeostase do cálcio desempenhadas pelo retículo endoplasmático são indissociáveis dos processos patológicos de remodelamento estrutural dos cardiomiócitos presentes na FA (33). Uma série de estímulos danosos relacionadas à FA como depleção do cálcio, hipóxia, estresse oxidativo, inflamação, exacerbação da síntese proteica e miólise, bem como o envelhecimento, podem desencadear estresse do RE e respostas celulares subsequentes, culminando, por exemplo, na expressão de genes de *heat-shock proteins* (HSPs) (34,35). Estas proteínas atenuam o processo de desregulação da dinâmica de síntese/degradação e da função de proteínas estruturais e contráteis dos cardiomiócitos na FA (36), facilitando o dobramento/redobramento de proteínas e a degradação de agregados proteicos que podem levar à apoptose dos cardiomiócitos (35). HSPA1A e membros da família HSPB são exemplos de chaperonas cuja presença protege contra o remodelamento estrutural dos cardiomiócitos e para as quais a indução farmacológica da expressão têm sido avaliada (37). A análise topológica de nossa rede principal revelou a chaperona BiP (binding

immunoglobulin protein), codificada pelo gene HSPA5, como *hub-bottleneck*, desempenhando provável papel chave na dinâmica patofisiológica da FA. A BiP desempenha, além de funções clássicas de uma chaperona, papel sinalizador em respostas celulares protetoras contra o estresse do RE, bem como papel regulatório da dinâmica do cálcio intracelular (38). Nosso estudo identificou o transcrito do gene HSPA5 como consistentemente *downregulated*. A superexpressão desta proteína foi capaz de atenuar a apoptose induzida por estresse do RE, enquanto que a ausência da mesma causa vazamento de cálcio do RE via complexo-canal Sec61 (39,40). Desta forma, a baixa expressão do transcrito do gene HSPA5 pode estar deixando de efetuar papel protetor contra o remodelamento estrutural e a apoptose em cardiomiócitos submetidos ao estresse na FA, assim como estar contribuindo nas despolarizações ectópicas do tecido atrial via *delayed afterdepolarizations* (DADs) induzidos pelo vazamento de cálcio do retículo sarcoplasmático (RS). A modulação da expressão da BiP, portanto, pode vir a ser mecanismo promissor na interrupção do remodelamento estrutural e elétrico do tecido atrial, interrompendo o progresso da doença.



**Figura 3 - Comparação entre as vias enriquecidas para cada módulo identificado.** Satisfatoriamente, as vias enriquecidas para cada módulo apresentam pouca sobreposição, indicando que são funcionalmente distintos.

Outro transcrito *downregulated* relacionado às vias enriquecidas para a fisiologia do RE/localização de proteínas e considerado *hub-bottleneck* pela análise topológica da rede de transcritos desregulados identificados pela metanálise é o ANK2. Este gene codifica para uma série de isoformas da proteína ankirina-B. No contexto celular, essas proteínas são elos adaptadores entre proteínas integrais de membrana e o citoesqueleto celular, organizando redes proteicas estruturais, eletrogênicas e de sinalização. A perda de função do gene ANK2 já foi associado a diversas patologias cardíacas, incluindo anormalidades de condução do impulso cardíaco e arritmias, inclusive a FA (41). Em cardiomiócitos, a ankirina-B parece organizar um complexo envolvendo a isoforma  $\alpha 2$  da bomba  $\text{Na}^+/\text{K}^+$  (NKA) e o trocador

$\text{Na}^+/\text{Ca}^{2+}$  (NCX) em um microambiente em que a atividade da NCX é regulada pela atividade da NKA. Nesse contexto, a perda de função ou baixa expressão da ankirina-B perturba a regulação e a localização na membrana destes trocadores, aumentando as taxas de ondas de  $\text{Ca}^{2+}$  associadas a vazamentos do RS via receptor de rianodina (RyR) (42), mecanismo arritmogênico importante e causador de DADs (43). O aumento dessas taxas está provavelmente associado à hiperfosforilação de RyRs pela *Ca<sup>2+</sup>/calmodulin-dependent protein kinase II* (CaMKII), cuja atividade é aumentada devido à maior  $[\text{Ca}^{2+}]$  na fenda entre a membrana plasmática e o RS resultante da localização e do funcionamento alterados do complexo NKA/NCX/ANK2 (44). Além disso, recentemente, a ankirina-B foi descrita interagindo com a via WNT/ $\beta$ -catenina através da modulação do estado de fosforilação da  $\beta$ -catenina. Neste estudo, a ativação farmacológica da via WNT/ $\beta$ -catenina foi capaz de resgatar modelo animal [Ank2-cKO] de cardiomiopatia arritmogênica (45). O transcrito da  $\beta$ -catenina também foi identificada como *hub-bottleneck* em nossa rede principal, de forma consistentemente *upregulated* na FA, agregando evidência sobre o papel da desregulação da expressão dessas proteínas na patofisiologia da FA.

O segundo módulo identificado pela análise de clusterização enriqueceu para diversas vias relacionadas à fisiologia cardíaca e à patofisiologia da FA, incluindo regulação e organização de filamentos de actina; catabolismo de proteínas; resposta à hipóxia e ao estresse oxidativo; diferenciação de células musculares e envelhecimento. A autofagia e o catabolismo de proteínas tem relação direta com o remodelamento estrutural dos cardiomiócitos presentes na FA, uma vez que a degradação de conexinas, proteínas contráteis e do citoesqueleto impedem o funcionamento adequado das células cardíacas (33,46). Além disso, o catabolismo proteico na FA também está ligado ao remodelamento elétrico das células através da degradação de canais de  $\text{Ca}^{2+}$  do tipo L (LTCC), cuja baixa expressão favorece o encurtamento do potencial de ação cardíaco e consequente surgimento de DADs (47,48). Apesar de ajudarem a manter a homeostase celular durante episódios de hipóxia e estresse oxidativo, mecanismos autofágicos e catabólicos também mostram-se citotóxicos caso exacerbados (49,50). Nesse sentido, não espanta que estudos avaliando a eficácia farmacológica da inibição desses mecanismos no contexto de doenças cardíacas apresentam resultados divergentes, sugerindo que o potencial protetivo e citotóxico destas vias são tenuamente separados (51–54). Ligado à indução das vias autofágicas está o dano proteico resultante do estresse oxidativo. Assim como em outras patologias cardíacas, o balanço entre a produção de espécies reativas de oxigênio (ROS) e a capacidade antioxidante dos átrios têm

sido associados ao desenvolvimento e à progressão da FA (55). Uma série de comorbidades comumente associadas a FA levam ao aumento da produção de ROS, tais como hiperglicemia, dislipidemia, hipertensão, inflamação e o aumento nos níveis teciduais de angiotensina II (4,55,56). O estresse oxidativo predispõe o tecido atrial à atividade ectópica através da alteração da fisiologia de diversas correntes iônicas, como a do  $\text{Na}^+$  e  $\text{K}^+$ , mas principalmente da do  $\text{Ca}^{2+}$ , afetando a atividade de LTCCs e RyR, propiciando vazamentos de  $\text{Ca}^{2+}$  do RS e seus efeitos arritmogênicos subsequentes (57–60). Também contribui no estabelecimento de distúrbios de condução característicos da FA pela exacerbação da fibrose cardíaca por meio da ativação de miofibroblastos por TGF- $\beta$  (61). O tratamento farmacológico de distúrbios arritmicos via inibição do estresse oxidativo, portanto, vem sendo apontado como promissor (55) e, nesta metanálise, apontamos os transcritos desregulados pertencentes a essa via (material suplementar - tabela 2), os quais configuram potenciais alvos de estudo na patofisiologia da FA. O transcrito desregulado NOX4, por exemplo, está *upregulated* na via de resposta ao estresse oxidativo identificada no módulo 2 e codifica a isoforma 4 da enzima NADPH oxidase (62). Junto com a NOX2, a NOX4 é a isoforma mais expressa no tecido cardíaco e está relacionada com a indução de apoptose de células cardíacas e ativação de miofibroblastos a partir da produção de  $\text{H}_2\text{O}_2$  (62,63). Sua alta expressão na FA, portanto, sugere a contínua injúria oxidativa à qual o tecido atrial está submetido e também surge como potencial alvo terapêutico.

Enriquecendo para diversas vias dos módulos 2 e 3 está o transcrito CTNNB1, que codifica para a  $\beta$ -catenina. Este transcrito está *upregulated* e configura um *hub-bottleneck* na rede principal de transcritos desregulados. A via Wnt/ $\beta$ -catenina regula os níveis citosólicos da  $\beta$ -catenina que, na ausência de sinal pelas proteínas ligantes Wnt, encontra-se associada ao complexo proteico de destruição APC/CK-1/GSK-3 que rapidamente a fosforila, marcando-a para ubiquitinação e, posteriormente, degradação. A ativação da via, por sua vez, desestrutura o complexo proteico de destruição, permitindo a acumulação da  $\beta$ -catenina e sua translocação para o núcleo, onde interage e ativa o fator de transcrição TCF/Lef, permitindo que seus diversos genes alvo sejam expressos (64). A via Wnt/ $\beta$ -catenina é evolutivamente conservada nos animais e controla processos de diferenciação, proliferação e sobrevivência celular e de desenvolvimento e homeostase tecidual. Os genes e vias que regula, contudo, são tecido-dependentes e modificam-se de acordo com o contexto fisiológico (64–66). Uma vez que a via Wnt/ $\beta$ -catenina está envolvida com a homeostasia de diferentes tecidos e que regula processos tão fundamentais, sua desregulação está envolvida com diversas doenças e tipos de

câncer (67). No coração, tem papel crucial ao longo do desenvolvimento embrionário, mas está quiescente no adulto. Contudo, estímulos estressores como dano isquêmico, estresse oxidativo e altas cargas pressóricas podem reativá-la, contribuindo no desenvolvimento de doenças como hipertrofia, insuficiência cardíaca e arritmias (64,68,69). No contexto da fibrilação atrial, uma série de alterações celulares induzidas pela via Wnt/ $\beta$ -catenina podem estar contribuindo com seu estabelecimento e progresso. O aumento dos níveis de  $\beta$ -catenina nuclear piorou o dano oxidativo induzido em modelo animal de cardiomiopatia diabética via interação  $\beta$ -catenina/c-Myc (70). Dessa mesma forma na FA, a ativação desta via pode estar contribuindo com o dano oxidativo exacerbado. Recentemente, observou-se também que o fator de transcrição TCF/Lef ativado pela  $\beta$ -catenina regula negativamente a expressão do canal de sódio Nav1.5, codificado pelo gene SCN5a, além de modular sua cinética, promovendo taquicardia e/ou fibrilação ventricular em modelo animal (71). No que diz respeito ao remodelamento estrutural dos átrios, a Wnt/ $\beta$ -catenina regula a transcrição de genes relacionados à fibrose na FA e em diversas outras doenças fibróticas (72). Há evidências, portanto, para indicar que o aumento da sinalização pela  $\beta$ -catenina é uma das vias responsáveis pela via “*chromatin remodeling*” enriquecida no módulo 3, uma vez que regula diversos processos e genes intrinsecamente ligados à patofisiologia da FA. Sob estresse, o aumento da expressão da  $\beta$ -catenina nos átrios pode estar ligado a uma resposta protetiva frustrada por meio da desdiferenciação celular e posterior transcrição de genes envolvidos com a proliferação, desenvolvimento e sobrevivência das células do tecido cardíaco. Esta hipótese é sustentada por esta metanálise através da identificação da expressão diferencial de transcritos e enriquecimento de vias na rede formada por eles relacionadas ao desenvolvimento tecidual como “*cardiac muscle tissue development*”; “*cell-substrate adhesion*”; “*muscle cell differentiation*”; “*nuclear chromosome segregation*”; “*establishment of organelle localization*”; “*regulation of morphogenesis of an epithelium*”; “*positive regulation of neurogenesis*”; “*regulation of cell morphogenesis*” e “*regulation of morphogenesis of an epithelium*”. A resistência do tecido cardíaco à reativação de vias embrionárias e proliferativas, contudo, bem como a exacerbação observada de diversos mecanismos patológicos, acabam por sugerir que é a inibição da sinalização pela  $\beta$ -catenina que deve ser estudada no âmbito terapêutico da FA. Por fim, a ativação de vias proliferativas também podem estar explicando a alta expressão identificada nesta metanálise do transcrito CNNB1 - outro *hub-bottleneck* de nossa rede -, que codifica para a proteína reguladora do ciclo celular ciclina B1. Em cardiomiócitos adultos, os níveis desse transcrito são extremamente baixos, uma vez que quase a totalidade dos cardiomiócitos estão presos nas

fases G1/S ou G2/M do ciclo celular (73,74). A expressão forçada do complexo ciclina B1/CDK1 promove a cariocinese em modelos *in vivo* e *in vitro*, mas a capacidade de promover citocinese e, por conseguinte, regeneração tecidual efetiva parece limitada (75). A importância topológica na rede e a desregulação da expressão deste transcrito, porém, ainda podem ser exploradas no sentido de representar um importante biomarcador na FA.

No quarto módulo identificado, é notável o enriquecimento de vias relacionadas a ativação, regulação e migração leucocitária, lançando luz à atividade inflamatória crônica ocorrendo na FA. Diversas comorbidades cardíacas associadas à FA como hipertensão, doença isquêmica ou aterosclerose, assim como a sinalização por angiotensina II, podem ter papel na promoção do estado pró-inflamatório sistêmico ou nos átrios (76,77). Instaurada a doença, a próprio dano e remodelamento tecidual presente na FA promove inflamação pelo extravasamento de “*damage-associated molecular patterns*” (DAMPs) (78). Análises citopatológicas de tecido atrial de pacientes com FA revelam intenso infiltrado leucocitário e necrose de cardiomiócitos adjacentes (79), o que é congruente com a via de regulação de “*cell-cell adhesion*” enriquecida no módulo 4, a qual atua sobre a expressão de proteínas de adesão nos linfócitos. Uma vez no tecido atrial, estas células secretam altos níveis de TGF- $\beta$ , o qual contribui no remodelamento estrutural pela ativação de miofibroblastos e IL-6, a qual induz remodelamento elétrico através da diminuição da expressão de conexinas, por exemplo (80). Sabe-se também que a IL-6 também induz a expressão de proteínas protrombóticas (81,82). Nesta metanálise, o transcrito do fator de von Willebrand foi identificado *upregulated* e classificado como *hub-bottleneck*, enriquecendo em vias relacionadas à sua função fisiológica como adesão celular, homeostase e desgranulação de plaquetas. Portanto, além do fluxo lento e turbulento do sangue favorecendo a formação de trombos e eventos tromboembólicos em pacientes com FA, a alta expressão do VWF contribui com esse quadro. Além disso, interage de forma central com várias outras proteínas na rede de transcritos diferencialmente expressos na FA, sugerindo que a modulação de seus níveis pode surgir como alternativa terapêutica envolvendo vias bem além das clássicas relacionadas ao uso de antiagregantes plaquetários.

## CONCLUSÕES

Nesta metanálise, integramos os dados de 9 *datasets* provenientes de experimentos de microarranjo envolvendo a FA, resultando numa lista de 589 transcritos diferencialmente expressos. A análise da rede de interação proteína-proteína destes transcritos revelou 4

módulos e 11 *hubs-bottlenecks* relacionados à patofisiologia da FA. A análise dos processos biológicos enriquecidos nestes módulos demonstraram boa congruência com a literatura recente e lançam luz a vias e a proteínas-chave no contexto da doença. Dessa forma, a metanálise de dados de transcriptoma mostra seu potencial no estudo dos mecanismos que permeiam doenças complexas. Portanto, uma análise mais profunda envolvendo a validação experimental dos nossos resultados pode trazer avanços significativos no que diz respeito à FA e a sua terapêutica.

## **LISTA DE ABREVIÇÕES**

CaMKII - *Ca<sup>2+</sup>/calmodulin-dependent protein kinase II*

DADs - *delayed afterdepolarizations*

DAMPs - *damage associated molecular patterns*

FA - fibrilação atrial

HSPs - *heat-shock proteins*

LTCC - canais de Ca<sup>2+</sup> do tipo L

NCX - bomba Na<sup>+</sup>/K<sup>+</sup>

NKA - trocador Na<sup>+</sup>/Ca<sup>2+</sup>

RE - retículo endoplasmático

ROS - espécies reativas de oxigênio

RS - retículo sarcoplasmático

RyR - receptor de rianodina

## **DECLARAÇÕES**

**Conflito de interesses**

Os autores declaram que este estudo foi realizado na ausência de quaisquer relações financeiras ou comerciais que possam configurar conflito de interesse

### **Financiamento**

Não há financiamento a declarar.

### **Contribuição dos autores**

R.H.B e M.R.M projetaram o estudo. R.H.B implementou os métodos, conduziu os experimentos e analisou os dados. R.H.B escreveu o manuscrito. M.R.M supervisionou o estudo. Todos os autores leram e aprovaram a versão final do manuscrito.

## **MATERIAL SUPLEMENTAR**

Tabela 1 - Todos os transcritos diferencialmente expressos identificados na metanálise e sua regulação

Tabela 2 - Vias biológicas enriquecidas em cada módulo, seus transcritos diferencialmente expressos e respectivas regulações

## REFERÊNCIAS

1. Heeringa J, van der Kuip DAM, Hofman A, Kors JA, van Herpen G, Stricker BHC, et al. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. *Eur Heart J* [Internet]. 2006 Apr [cited 2019 Sep 27];27(8):949–53. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16527828>
2. Ball J, Carrington MJ, McMurray JJ V, Stewart S. Atrial fibrillation: profile and burden of an evolving epidemic in the 21st century. *Int J Cardiol* [Internet]. 2013 Sep 1 [cited 2019 Sep 27];167(5):1807–24. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23380698>
3. Nattel S, Harada M. Atrial remodeling and atrial fibrillation: recent advances and translational perspectives. *J Am Coll Cardiol* [Internet]. 2014 Jun 10 [cited 2019 Sep 27];63(22):2335–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24613319>
4. Heijman J, Voigt N, Nattel S, Dobrev D. Cellular and molecular electrophysiology of atrial fibrillation initiation, maintenance, and progression. *Circ Res* [Internet]. 2014 Apr 25 [cited 2019 Sep 27];114(9):1483–99. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24763466>
5. R: The R Project for Statistical Computing [Internet]. [cited 2019 Sep 26]. Available from: <https://www.r-project.org/>
6. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003 Nov;13(11):2498–504.
7. Kauffmann A, Huber W. Microarray data quality control improves the detection of differentially expressed genes. *Genomics* [Internet]. 2010 Mar [cited 2019 Sep 26];95(3):138–42. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20079422>
8. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. *Bioinformatics* [Internet]. 2009 Feb 1 [cited 2019 Sep 26];25(3):415–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19106121>
9. Smyth G, Hu Y, Ritchie M, Silver J, Wettenhall J, McCarthy D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. [Internet]. *Nucleic Acids Research. Bioconductor*; 2015 [cited 2019 Sep 26]. p. Available from: <http://bioconductor.org/packages/release/bioc/html/limma.html>
10. Irizarry RA. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* [Internet]. 2003 Apr 1 [cited 2019 Sep 26];4(2):249–64. Available from: <https://academic.oup.com/biostatistics/article-lookup/doi/10.1093/biostatistics/4.2.249>
11. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* [Internet]. 2010 Oct 1 [cited 2019 Sep 26];26(19):2363–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20688976>

12. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* [Internet]. 2007 Sep [cited 2019 Sep 26];3(9):1724–35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17907809>
13. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, et al. A comparison of background correction methods for two-colour microarrays. *Bioinformatics*. 2007 Oct 15;23(20):2700–7.
14. Silver JD, Ritchie ME, Smyth GK. Microarray background correction: Maximum likelihood estimation for the normal-exponential convolution. *Biostatistics*. 2009 Apr;10(2):352–63.
15. Yang YH, Thorne NP. Normalization for two-color cDNA microarray data. In 2003. p. 403–18.
16. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* [Internet]. 2004 [cited 2019 Sep 26];3:Article3. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16646809>
17. Wang X, Kang DD, Shen K, Song C, Lu S, Chang LC, et al. An r package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. *Bioinformatics*. 2012 Oct;28(19):2534–6.
18. Marot G, Foulley J-L, Mayer C-D, Jaffrézic F. Moderated effect size and P-value combinations for microarray meta-analyses. *Bioinformatics* [Internet]. 2009 Oct 15 [cited 2019 Sep 26];25(20):2692–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19628502>
19. Song C, Tseng GC. HYPOTHESIS SETTING AND ORDER STATISTIC FOR ROBUST GENOMIC META-ANALYSIS. *Ann Appl Stat* [Internet]. 2014 [cited 2019 Sep 26];8(2):777–800. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25383132>
20. Benjamini, Yoav ; Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* 1995.pdf. *J R Stat Soc Ser B* [Internet]. 1995 [cited 2019 Sep 26];57(1):289–300. Available from: <http://www.jstor.org/stable/2346101>
21. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015 Jan 28;43(D1):D447–52.
22. Assenov Y, Ramírez F, Schelhorn SESE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. *Bioinformatics*. 2008 Jan 15;24(2):282–4.
23. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003 Jan 13;4.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. Vol. 25, *Nature Genetics*. 2000. p. 25–9.

25. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 2019 Jan 8;47(D1):D330–8.
26. Chang LC, Lin HM, Sibille E, Tseng GC. Meta-analysis methods for combining multiple expression profiles: Comparisons, statistical characterization and an application guideline. *BMC Bioinformatics.* 2013 Dec 21;14(1).
27. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. Vol. 5, *PLoS Medicine.* 2008. p. 1320–32.
28. Barabási AL, Oltvai ZN. Network biology: Understanding the cell's functional organization. Vol. 5, *Nature Reviews Genetics.* 2004. p. 101–13.
29. Park K, Kim D. Localized network centrality and essentiality in the yeast-protein interaction network. *Proteomics [Internet].* 2009 Nov [cited 2019 Oct 6];9(22):5143–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19771559>
30. Charitou T, Bryan K, Lynn DJ. Using biological networks to integrate, visualize and analyze genomics data. Vol. 48, *Genetics Selection Evolution.* BioMed Central Ltd.; 2016.
31. Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF. A new method to measure the semantic similarity of GO terms. *Bioinformatics.* 2007 May 15;23(10):1274–81.
32. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi A J Integr Biol.* 2012 May 1;16(5):284–7.
33. Wiersma M, Meijering RAM, Qi XY, Zhang D, Liu T, Hoogstra-Berends F, et al. Endoplasmic reticulum stress is associated with autophagy and cardiomyocyte remodeling in experimental and human atrial fibrillation. *J Am Heart Assoc.* 2017 Oct 1;6(10).
34. Nattel S, Harada M. Atrial remodeling and atrial fibrillation: recent advances and translational perspectives. *J Am Coll Cardiol [Internet].* 2014 Jun 10 [cited 2019 Oct 31];63(22):2335–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24613319>
35. Amen OM, Sarker SD, Ghildyal R, Arya A. Endoplasmic Reticulum Stress Activates Unfolded Protein Response Signaling and Mediates Inflammation, Obesity, and Cardiac Dysfunction: Therapeutic and Molecular Approach. *Front Pharmacol.* 2019 Sep 10;10.
36. Allessie MA, de Groot NMS, Houben RPM, Schotten U, Boersma E, Smeets JL, et al. Electropathological substrate of long-standing persistent atrial fibrillation in patients with structural heart disease: longitudinal dissociation. *Circ Arrhythm Electrophysiol [Internet].* 2010 Dec [cited 2019 Oct 31];3(6):606–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20719881>
37. Hoogstra-Berends F, Meijering RAM, Zhang D, Heeres A, Loen L, Seerden J-P, et al. Heat shock protein-inducing compounds as therapeutics to restore proteostasis in atrial

- fibrillation. *Trends Cardiovasc Med* [Internet]. 2012 Apr [cited 2019 Oct 19];22(3):62–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22863365>
38. Wang J, Lee J, Liem D, Ping P. HSPA5 Gene encoding Hsp70 chaperone BiP in the endoplasmic reticulum. Vol. 618, *Gene*. Elsevier B.V.; 2017. p. 14–23.
  39. Fu HY, Minamino T, Tsukamoto O, Sawada T, Asai M, Kato H, et al. Overexpression of endoplasmic reticulum-resident chaperone attenuates cardiomyocyte death induced by proteasome inhibition. *Cardiovasc Res* [Internet]. 2008 Sep 1 [cited 2019 Nov 2];79(4):600–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18508854>
  40. Schäuble N, Lang S, Jung M, Cappel S, Schorr S, Ulucan Ö, et al. BiP-mediated closing of the Sec61 channel limits Ca<sup>2+</sup> leakage from the ER. *EMBO J* [Internet]. 2012 Aug 1 [cited 2019 Nov 2];31(15):3282–96. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22796945>
  41. Cunha SR, Mohler PJ. Ankyrin protein networks in membrane formation and stabilization. *J Cell Mol Med* [Internet]. [cited 2019 Nov 2];13(11–12):4364–76. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19840192>
  42. Skogestad J, Aronsen JM, Tovsrud N, Wanichawan P, Hougen K, Stokke MK, et al. Coupling of the Na<sup>+</sup>/K<sup>+</sup>-ATPase to Ankyrin B controls Na<sup>+</sup>/Ca<sup>2+</sup> exchanger activity in cardiomyocytes. *Cardiovasc Res*. 2019 Apr 5;
  43. Shiferaw Y, Aistrup GL, Wasserstrom JA. Intracellular Ca<sup>2+</sup> waves, afterdepolarizations, and triggered arrhythmias. Vol. 95, *Cardiovascular Research*. 2012. p. 265–8.
  44. Popescu I, Galice S, Mohler PJ, Despa S. Elevated local [Ca<sup>2+</sup>] and CaMKII promote spontaneous Ca<sup>2+</sup> release in ankyrin-B-deficient hearts. *Cardiovasc Res*. 2016 Aug 1;111(3):287–94.
  45. Roberts JD, Murphy NP, Hamilton RM, Lubbers ER, James CA, Kline CF, et al. Ankyrin-B dysfunction predisposes to arrhythmogenic cardiomyopathy and is amenable to therapy. *J Clin Invest* [Internet]. 2019 [cited 2019 Nov 3];129(8):3171–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31264976>
  46. Zhang D, Wu CT, Qi XY, Meijering RAM, Hoogstra-Berends F, Tadevosyan A, et al. Activation of histone deacetylase-6 induces contractile dysfunction through derailment of  $\alpha$ -tubulin proteostasis in experimental and human atrial fibrillation. *Circulation*. 2014 Jan 21;129(3):346–58.
  47. Yuan Y, Zhao J, Gong Y, Wang D, Wang X, Yun F, et al. Autophagy exacerbates electrical remodeling in atrial fibrillation by ubiquitin-dependent degradation of L-type calcium channel. *Cell Death Dis*. 2018 Sep 1;9(9).
  48. Heinzel FR, Bito V, Biesmans L, Wu M, Detre E, Von Wegner F, et al. Remodeling of T-tubules and reduced synchrony of Ca<sup>2+</sup> release in myocytes from chronically ischemic myocardium. *Circ Res*. 2008 Feb;102(3):338–46.

49. Nakai A, Yamaguchi O, Takeda T, Higuchi Y, Hikoso S, Taniike M, et al. The role of autophagy in cardiomyocytes in the basal state and in response to hemodynamic stress. *Nat Med*. 2007 May;13(5):619–24.
50. Pagan J, Seto T, Pagano M, Cittadini A. Role of the ubiquitin proteasome system in the heart. *Circ Res* [Internet]. 2013 Mar 29 [cited 2019 Nov 5];112(7):1046–58. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23538275>
51. Hedhli N, Lizano P, Hong C, Fritzky LF, Dhar SK, Liu H, et al. Proteasome inhibition decreases cardiac remodeling after initiation of pressure overload. *Am J Physiol Circ Physiol* [Internet]. 2008 Oct [cited 2019 Nov 5];295(4):H1385–93. Available from: <https://www.physiology.org/doi/10.1152/ajpheart.00532.2008>
52. Hedhli N, Wang L, Wang Q, Rashed E, Tian Y, Sui X, et al. Proteasome activation during cardiac hypertrophy by the chaperone H11 Kinase/Hsp22. *Cardiovasc Res* [Internet]. 2007 Nov 12 [cited 2019 Nov 5];77(3):497–505. Available from: <https://academic.oup.com/cardiovasces/article-lookup/doi/10.1093/cvr/cvm054>
53. Tang M, Li J, Huang W, Su H, Liang Q, Tian Z, et al. Proteasome functional insufficiency activates the calcineurin–NFAT pathway in cardiomyocytes and promotes maladaptive remodelling of stressed mouse hearts. *Cardiovasc Res* [Internet]. 2010 Dec 1 [cited 2019 Nov 5];88(3):424–33. Available from: <https://academic.oup.com/cardiovasces/article-lookup/doi/10.1093/cvr/cvq217>
54. Carrier L. Too much of a good thing is bad: proteasome inhibition induces stressed hearts to fail. *Cardiovasc Res* [Internet]. 2010 Dec 1 [cited 2019 Nov 5];88(3):389–90. Available from: <https://academic.oup.com/cardiovasces/article-lookup/doi/10.1093/cvr/cvq315>
55. Korantzopoulos P, Letsas K, Fragakis N, Tse G, Liu T. Oxidative stress and atrial fibrillation: an update. Vol. 52, *Free Radical Research*. Taylor and Francis Ltd; 2018. p. 1199–209.
56. Chiang CE, Naditch-Brûlé L, Murin J, Goethals M, Inoue H, O'Neill J, et al. Distribution and risk profile of paroxysmal, persistent, and permanent atrial fibrillation in routine clinical practice insight from the real-life global survey evaluating patients with atrial fibrillation international registry. *Circ Arrhythmia Electrophysiol*. 2012;5(4):632–9.
57. Bukowska A, Schild L, Keilhoff G, Hirte D, Neumann M, Gardemann A, et al. Mitochondrial dysfunction and redox signaling in atrial tachyarrhythmia. *Exp Biol Med*. 2008 May;233(5):558–74.
58. Xie W, Santulli G, Reiken SR, Yuan Q, Osborne BW, Chen B-X, et al. Mitochondrial oxidative stress promotes atrial fibrillation. *Sci Rep* [Internet]. 2015 Sep 14 [cited 2019 Nov 5];5(1):11427. Available from: <http://www.nature.com/articles/srep11427>
59. Anzai K, Ogawa K, Kuniyasu A, Ozawa T, Yamamoto H, Nakayama H. Effects of hydroxyl radical and sulfhydryl reagents on the open probability of the purified cardiac ryanodine receptor channel incorporated into planar lipid bilayers. *Biochem Biophys Res Commun*. 1998 Aug 28;249(3):938–42.

60. Sovari AA. Cellular and Molecular Mechanisms of Arrhythmia by Oxidative Stress. Vol. 2016, Cardiology Research and Practice. Hindawi Limited; 2016.
61. Friedrichs K, Baldus S, Klinke A. Fibrosis in atrial fibrillation - Role of reactive species and MPO. Vol. 3 JUN, *Frontiers in Physiology*. 2012.
62. Youn JY, Zhang J, Zhang Y, Chen H, Liu D, Ping P, et al. Oxidative stress in atrial fibrillation: An emerging role of NADPH oxidase. Vol. 62, *Journal of Molecular and Cellular Cardiology*. 2013. p. 72–9.
63. Zhang J, Youn JY, Kim AY, Ramirez RJ, Gao L, Ngo D, et al. NOX4-dependent hydrogen peroxide overproduction in human atrial fibrillation and HL-1 atrial cells: Relationship to hypertension. *Front Physiol*. 2012;3 JUN.
64. Barker N. The canonical Wnt/beta-catenin signalling pathway. *Methods Mol Biol* [Internet]. 2008 [cited 2019 Oct 15];468:5–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19099242>
65. MacDonald BT, Tamai K, He X. Wnt/ $\beta$ -Catenin Signaling: Components, Mechanisms, and Diseases. Vol. 17, *Developmental Cell*. 2009. p. 9–26.
66. Frietze S, Wang R, Yao L, Tak Y, Ye Z, Gaddis M, et al. Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol* [Internet]. 2012 [cited 2019 Nov 7];13(9):R52. Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2012-13-9-r52>
67. Nusse R, Clevers H. Wnt/ $\beta$ -Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* [Internet]. 2017 Jun [cited 2019 Nov 7];169(6):985–99. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867417305470>
68. Malekar P, Hagenmueller M, Anyanwu A, Buss S, Streit MR, Weiss CS, et al. Wnt signaling is critical for maladaptive cardiac hypertrophy and accelerates myocardial remodeling. *Hypertension*. 2010 Apr;55(4):939–45.
69. Dawson K, Aflaki M, Nattel S. Role of the Wnt-Frizzled system in cardiac pathophysiology: A rapidly developing, poorly understood area with enormous potential. Vol. 591, *Journal of Physiology*. 2013. p. 1409–32.
70. Liu P, Su J, Song X, Wang S. Activation of nuclear  $\beta$ -catenin/c-Myc axis promotes oxidative stress injury in streptozotocin-induced diabetic cardiomyopathy. *Biochem Biophys Res Commun*. 2017 Dec 2;493(4):1573–80.
71. Huo R, Hu C, Zhao L, Sun L, Wang N, Lu Y, et al. Enhancement of  $\beta$ -catenin/T-cell factor 4 signaling causes susceptibility to cardiac arrhythmia by suppressing NaV1.5 expression in mice. *Hear Rhythm*. 2019 May;
72. Lv X, Li J, Hu Y, Wang S, Yang C, Li C, et al. Overexpression of miR-27b-3p Targeting Wnt3a Regulates the Signaling Pathway of Wnt/ $\beta$ -Catenin and Attenuates Atrial Fibrosis in Rats with Atrial Fibrillation. *Oxid Med Cell Longev*. 2019;2019:5703764.

73. Ponnusamy M, Li PF, Wang K. Understanding cardiomyocyte proliferation: an insight into cell cycle activity. Vol. 74, Cellular and Molecular Life Sciences. Birkhauser Verlag AG; 2017. p. 1019–34.
74. Ahuja P, Sdek P, MacLellan WR. Cardiac myocyte cell cycle control in development, disease, and regeneration. Vol. 87, Physiological Reviews. 2007. p. 521–44.
75. Bicknell KA, Coxon CH, Brooks G. Forced expression of the cyclin B1-CDC2 complex induces proliferation in adult rat cardiomyocytes. *Biochem J*. 2004 Sep 1;382(2):411–6.
76. Andrade J, Khairy P, Dobrev D, Nattel S. The Clinical Profile and Pathophysiology of Atrial Fibrillation. *Circ Res* [Internet]. 2014 Apr 25 [cited 2019 Nov 11];114(9):1453–68. Available from: <https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.114.303211>
77. da Silva RMFL. Influence of Inflammation and Atherosclerosis in Atrial Fibrillation. *Curr Atheroscler Rep* [Internet]. 2017 Jan 19 [cited 2019 Nov 11];19(1):2. Available from: <http://link.springer.com/10.1007/s11883-017-0639-0>
78. Hu YF, Chen YJ, Lin YJ, Chen SA. Inflammation and the pathogenesis of atrial fibrillation. Vol. 12, *Nature Reviews Cardiology*. Nature Publishing Group; 2015. p. 230–43.
79. Frustaci A, Chimenti C, Bellocci F, Morgante E, Russo MA, Maseri A. Histological substrate of atrial biopsies in patients with lone atrial fibrillation. *Circulation* [Internet]. 1997 Aug 19 [cited 2019 Nov 11];96(4):1180–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9286947>
80. Jalloul Y, Refaat MM. IL-6 Rapidly Induces Reversible Atrial Electrical Remodeling by Downregulation of Cardiac Connexins. *J Am Heart Assoc*. 2019 Aug 20;8(16).
81. Kerr R, Stirling D, Ludlam CA. Interleukin 6 and Haemostasis. *Br J Haematol* [Internet]. 2001 Oct [cited 2019 Nov 12];115(1):3–12. Available from: <http://doi.wiley.com/10.1046/j.1365-2141.2001.03061.x>
82. Inflammation and Thrombosis in Atrial Fibrillation [Internet]. [cited 2019 Nov 12]. Available from: <https://www.revespcardiol.org/en-pdf-S1885585711003069>

### 3 CONCLUSÕES E PERSPECTIVAS

Esta metanálise de dados de transcriptoma lançou luz a módulos funcionais enriquecendo para diversos processos biológicos envolvidos com a patofisiologia da fibrilação atrial e identificou 589 transcritos diferencialmente expressos no contexto da doença. Estes resultados estão totalmente disponíveis na tabela 2 do material suplementar do artigo científico e esperamos que seu conteúdo inspire muitos estudos futuros. Além disso, ficamos satisfeitos em ver que a metanálise de dados de transcriptoma serviu ao propósito de obter resultados mais sensíveis do que a análise individual desse tipo de dado. Prova disso é o enriquecimento de vias biológicas pelos transcritos diferencialmente expressos há muito tempo estudadas na literatura no contexto da fibrilação atrial.

Contudo, uma série de melhorias ainda podem ser realizadas. Dados sobre a expressão de miRNAs, por exemplo, podem ser integrados a rede, aumentando o espectro e profundidade das análises realizadas. O uso de outros algoritmos de metanálise e/ou para identificação dos módulos funcionais na rede também poderia ser testado e seus resultados comparados. Por fim, o uso de testes *in vitro* ou *in vivo* seriam a prova final da veracidade dos resultados obtidos por esse estudo. Por isso, a escolha de algumas vias ou genes/proteínas mais interessantes poderia ser feita e, assim, o estudo poderia prosseguir, almejando, por exemplo, identificar novas metodologias terapêuticas a serem utilizadas no tratamento da fibrilação atrial.

## REFERÊNCIAS

- BADER, Gary D.; HOGUE, Christopher W. V. An automated method for finding molecular complexes in large protein interaction networks. **BMC bioinformatics**, [s. l.], v. 4, p. 2, 2003.
- BALL, Jocasta et al. Atrial fibrillation: profile and burden of an evolving epidemic in the 21st century. **International journal of cardiology**, [s. l.], v. 167, n. 5, p. 1807–1824, 2013.
- BARBULOVIC-NAD, Irena et al. Bio-Microarray Fabrication Techniques—A Review. **Critical reviews in biotechnology**, [s. l.], v. 26, n. 4, p. 237–259, 2006.
- BERS, Donald M. Cardiac excitation–contraction coupling. **Nature**, [s. l.], v. 415, n. 6868, p. 198–205, 2002. . Acesso em: 28 jul. 2019.
- BROOKES, C. I. et al. Myocardial contractility is not constant during spontaneous atrial fibrillation in patients. **Circulation**, [s. l.], v. 98, n. 17, p. 1762–1768, 1998.
- BURSTEIN, Brett et al. Atrial cardiomyocyte tachycardia alters cardiac fibroblast function: a novel consideration in atrial remodeling. **Cardiovascular research**, [s. l.], v. 76, n. 3, p. 442–452, 2007.
- BURSTEIN, Brett et al. Differential behaviors of atrial versus ventricular fibroblasts: a potential role for platelet-derived growth factor in atrial-ventricular remodeling differences. **Circulation**, [s. l.], v. 117, n. 13, p. 1630–1641, 2008.
- BURSTEIN, Brett; NATTEL, Stanley. Atrial fibrosis: mechanisms and clinical relevance in atrial fibrillation. **Journal of the American College of Cardiology**, [s. l.], v. 51, n. 8, p. 802–809, 2008.
- CHALDOUPI, Sevasti-Maria et al. The role of connexin40 in atrial fibrillation. **Cardiovascular research**, [s. l.], v. 84, n. 1, p. 15–23, 2009.
- CHANG, Lun-Ching et al. Meta-analysis methods for combining multiple expression profiles: comparisons, statistical characterization and an application guideline. **BMC bioinformatics**, [s. l.], v. 14, p. 368, 2013.
- CHARITOS, Efstratios I. et al. Clinical classifications of atrial fibrillation poorly reflect its temporal persistence: insights from 1,195 patients continuously monitored with implantable devices. **Journal of the American College of Cardiology**, [s. l.], v. 63, n. 25 Pt A, p. 2840–2848, 2014.
- CHUGH, Sumeet S. et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. **Circulation**, [s. l.], v. 129, n. 8, p. 837–847, 2014.
- CLOUGH, Emily; BARRETT, Tanya. The Gene Expression Omnibus Database. **Methods in molecular biology**, [s. l.], v. 1418, p. 93–110, 2016.
- DEVORE, Adam D. et al. Hospitalizations in patients with atrial fibrillation: an analysis from ROCKET AF. **Europace: European pacing, arrhythmias, and cardiac electrophysiology:**

**journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology**, [s. l.], v. 18, n. 8, p. 1135–1142, 2016.

DOBREV, Dobromir; NATTEL, Stanley. New antiarrhythmic drugs for treatment of atrial fibrillation. **The Lancet**, [s. l.], v. 375, n. 9721, p. 1212–1223, 2010.

DONG, Zhicheng; CHEN, Yan. Transcriptomics: advances and approaches. **Science China. Life sciences**, [s. l.], v. 56, n. 10, p. 960–967, 2013.

FUSTER, Valentin et al. 2011 ACCF/AHA/HRS focused updates incorporated into the ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. **Circulation**, [s. l.], v. 123, n. 10, p. e269–367, 2011.

GUINDALINI, Camila; TUFIK, Sergio. Uso de microarrays na busca de perfis de expressão gênica: aplicação no estudo de fenótipos complexos. **Brazilian Journal of Psychiatry**, [s. l.], v. 29, n. 4, p. 370–374, 2007. . Acesso em: 25 set. 2019.

HEALEY, Jeff S. et al. Subclinical atrial fibrillation and the risk of stroke. **The New England journal of medicine**, [s. l.], v. 366, n. 2, p. 120–129, 2012.

HUNYADY, László; CATT, Kevin J. Pleiotropic AT1 receptor signaling pathways mediating physiological and pathogenic actions of angiotensin II. **Molecular endocrinology**, [s. l.], v. 20, n. 5, p. 953–970, 2006.

IKEDA, Tatsunori et al. Augmented single-unit muscle sympathetic nerve activity in heart failure with chronic atrial fibrillation. **The Journal of physiology**, [s. l.], v. 590, n. 3, p. 509–518, 2012.

LOWE, Rohan et al. Transcriptomics technologies. **PLoS computational biology**, [s. l.], v. 13, n. 5, p. e1005457, 2017.

MCMANUS, David D.; RIENSTRA, Michiel; BENJAMIN, Emelia J. An update on the prognosis of patients with atrial fibrillation. **Circulation**, [s. l.], v. 126, n. 10, p. e143–6, 2012.

NATTEL, Stanley; DOBREV, Dobromir. The multidimensional role of calcium in atrial fibrillation pathophysiology: mechanistic insights and therapeutic opportunities. **European heart journal**, [s. l.], v. 33, n. 15, p. 1870–1877, 2012.

NATTEL, Stanley; HARADA, Masahide. Atrial remodeling and atrial fibrillation: recent advances and translational perspectives. **Journal of the American College of Cardiology**, [s. l.], v. 63, n. 22, p. 2335–2345, 2014.

NISHIDA, Kunihiko et al. Animal models for atrial fibrillation: clinical insights and scientific opportunities. **Europace: European pacing, arrhythmias, and cardiac electrophysiology: journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology**, [s. l.], v. 12, n. 2, p. 160–172, 2010.

- OAKES, Robert S. et al. Detection and quantification of left atrial structural remodeling with delayed-enhancement magnetic resonance imaging in patients with atrial fibrillation. **Circulation**, [s. l.], v. 119, n. 13, p. 1758–1767, 2009.
- PANDIT, Sandeep V. et al. Ionic determinants of functional reentry in a 2-D model of human atrial cells during simulated chronic atrial fibrillation. **Biophysical journal**, [s. l.], v. 88, n. 6, p. 3806–3821, 2005.
- PATHAK, Anand et al. Enhancement of cardiac function and suppression of heart failure progression by inhibition of protein phosphatase 1. **Circulation research**, [s. l.], v. 96, n. 7, p. 756–766, 2005.
- RAMASAMY, Adaikalavan et al. Key Issues in Conducting a Meta-Analysis of Gene Expression Microarray Datasets. **PLoS medicine**, [s. l.], v. 5, n. 9, p. e184, 2008. . Accesso em: 26 set. 2019.
- ROSENKRANZ, Stephan. TGF-beta1 and angiotensin networking in cardiac remodeling. **Cardiovascular research**, [s. l.], v. 63, n. 3, p. 423–432, 2004.
- SONG, Chi; TSENG, George C. HYPOTHESIS SETTING AND ORDER STATISTIC FOR ROBUST GENOMIC META-ANALYSIS. **The annals of applied statistics**, [s. l.], v. 8, n. 2, p. 777–800, 2014.
- TERRENOIRE, Cecile et al. Autonomic control of cardiac action potentials: role of potassium channel kinetics in response to sympathetic stimulation. **Circulation research**, [s. l.], v. 96, n. 5, p. e25–34, 2005.
- THRALL, Graham et al. Quality of life in patients with atrial fibrillation: a systematic review. **The American journal of medicine**, [s. l.], v. 119, n. 5, p. 448.e1–19, 2006.
- VERHEULE, Sander et al. Increased vulnerability to atrial fibrillation in transgenic mice with selective atrial fibrosis caused by overexpression of TGF-beta1. **Circulation research**, [s. l.], v. 94, n. 11, p. 1458–1465, 2004.
- WAKILI, Reza et al. Recent advances in the molecular pathophysiology of atrial fibrillation. **The Journal of clinical investigation**, [s. l.], v. 121, n. 8, p. 2955–2968, 2011.
- WALSH, Christopher J. et al. Microarray Meta-Analysis and Cross-Platform Normalization: Integrative Genomics for Robust Biomarker Discovery. **Microarrays (Basel, Switzerland)**, [s. l.], v. 4, n. 3, p. 389–406, 2015.
- XIAO, Hong D. et al. Mice with cardiac-restricted angiotensin-converting enzyme (ACE) have atrial enlargement, cardiac arrhythmia, and sudden death. **The American journal of pathology**, [s. l.], v. 165, n. 3, p. 1019–1032, 2004.
- YUE, Lixia; XIE, Jia; NATTEL, Stanley. Molecular determinants of cardiac fibroblast electrical function and therapeutic implications for atrial fibrillation. **Cardiovascular research**, [s. l.], v. 89, n. 4, p. 744–753, 2011.

# Molecular Medicine - Submission Guidelines

Disponível em:

<https://molmed.biomedcentral.com/submission-guidelines/preparing-your-manuscript/research-article>

## Research article

### Criteria

Research articles should report on original primary research. The Discussion section should discuss the application of the results to the understanding of disease and the development of tools for disease diagnosis, treatment and prevention. The main findings should be placed in context to highlight the advance and describe how it moves the field forward.

*Molecular Medicine* strongly encourages that all datasets on which the conclusions of the paper rely should be available to readers. We encourage authors to ensure that their datasets are either deposited in publicly available repositories (where available and appropriate) or presented in the main manuscript or additional supporting files whenever possible. Please see Springer Nature's information on recommended repositories. Where a widely established research community expectation for data archiving in public repositories exists, submission to a community-endorsed, public repository is mandatory. A list of data where deposition is required, with the appropriate repositories, can be found on the Editorial Policies Page.

The Editors of *Molecular Medicine* welcome authors to send any pre-submission inquiries they might have here. We ask that you please include "Pre-submission inquiry" in the subject line of your email and that you include the title and abstract of your proposed submission within the message. Our Editors will do their best to respond to you in a timely manner.

Please note: we ask that you review the sections on this entire page before submitting your work, as omission of any necessary content will delay the processing of your manuscripts. Some information which is often forgotten includes (but is not limited to) the following, but

please go through the entire page to make sure you include all necessary sections/information:

- The Declarations section, and its sub-sections
- Corresponding author information (it must stay consistent throughout the process; *Molecular Medicine* strongly advises having only one Corresponding Author, for ease of communication during the editorial workflow and production of your article)
- Co-author email addresses
- A list of abbreviations

## Cover letter

You may suggest up to three potential reviewers in your cover letter. You should not suggest recent collaborators or colleagues who work in the same institution as yourselves. You should provide institutional email addresses where possible, or information which will help the Editor to verify the identity of the reviewer (for example an ORCID or Scopus ID). See our Editorial policies for guidance on suggesting peer reviewers, and Submission Guidelines for further information on what you should include in your cover letter.

## Preparing your manuscript

The information below details the section headings that you should include in your manuscript and what information should be within each section.

Please note that your manuscript must include a 'Declarations' section including all of the subheadings (please see below for more information).

### Title page

The title page should:

- present a title that includes, if appropriate, the study design  
e.g.:
  - "A versus B in the treatment of C: a randomized controlled trial", "X is a risk factor for Y: a case control study", "What is the impact of factor X on subject Y: A systematic review"
  - or for non-clinical or non-research studies a description of what the article reports
- list the full names and institutional addresses for all authors
  - if a collaboration group should be listed as an author, please list the Group name as an author. If you would like the names of the individual members of the Group to be searchable through their individual PubMed records, please include this information in the "Acknowledgements" section in accordance with the instructions below

- indicate the corresponding author (Please note: *Molecular Medicine* strongly advises having only one Corresponding Author, for ease of communication during the editorial workflow and Production of your article)

### Abstract

The Abstract should not exceed 350 words. Please minimize the use of abbreviations and do not cite references in the abstract. Reports of randomized controlled trials should follow the CONSORT extension for abstracts. The abstract must include the following separate sections:

- **Background:** the context and purpose of the study
- **Methods:** how the study was performed and statistical tests used
- **Results:** the main findings
- **Conclusions:** brief summary and potential implications
- **Trial registration:** If your article reports the results of a health care intervention on human participants, it must be registered in an appropriate registry and the registration number and date of registration should be stated in this section. If it was not registered prospectively (before enrollment of the first participant), you should include the words 'retrospectively registered'. See our editorial policies for more information on trial registration

### Keywords

Three to ten keywords representing the main content of the article.

### Background

The Background section should explain the background to the study, its aims, a summary of the existing literature and why this study was necessary or its contribution to the field.

### Methods

The methods section should include:

- the aim, design and setting of the study
- the characteristics of participants or description of materials
- a clear description of all processes, interventions and comparisons. Generic drug names should generally be used. When proprietary brands are used in research, include the brand names in parentheses
- the type of statistical analysis used, including a power calculation if appropriate

### Results

This should include the findings of the study including, if appropriate, results of statistical analysis which must be included either in the text or as tables and figures.

#### Discussion

This section should discuss the implications of the findings in context of existing research and highlight limitations of the study.

#### Conclusions

This should state clearly the main conclusions and provide an explanation of the importance and relevance of the study reported.

#### List of abbreviations

If abbreviations are used in the text they should be defined in the text at first use, and a list of abbreviations should be provided.

### Declarations

All manuscripts must contain the following sections under the heading 'Declarations':

- Ethics approval and consent to participate
- Consent for publication
- Availability of data and materials
- Competing interests
- Funding
- Authors' contributions
- Acknowledgements
- Authors' information (optional)

Please see below for details on the information to be included in these sections.

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

#### *Ethics approval and consent to participate*

Manuscripts reporting studies involving human participants, human data or human tissue must:

- include a statement on ethics approval and consent (even where the need for approval was waived)
- include the name of the ethics committee that approved the study and the committee's reference number if appropriate

Studies involving animals must include a statement on ethics approval. See our editorial policies for more information.

If your manuscript does not report on or involve the use of any animal or human data or tissue, please state "Not applicable" in this section.

#### *Consent for publication*

If your manuscript contains any individual person's data in any form (including any individual details, images or videos), consent for publication must be obtained from that person, or in the case of children, their parent or legal guardian. All presentations of case reports must have consent for publication.

You can use your institutional consent form or our consent form if you prefer. You should not send the form to us on submission, but we may request to see a copy at any stage (including after publication).

See our editorial policies for more information on consent for publication.

If your manuscript does not contain data from any individual person, please state "Not applicable" in this section.

*Availability of data and materials*

All manuscripts must include an 'Availability of data and materials' statement. Data availability statements should include information on where data supporting the results reported in the article can be found including, where applicable, hyperlinks to publicly archived datasets analysed or generated during the study. By data we mean the minimal dataset that would be necessary to interpret, replicate and build upon the findings reported in the article. We recognise it is not always possible to share research data publicly, for instance when individual privacy could be compromised, and in such instances data availability should still be stated in the manuscript along with any conditions for access.

Data availability statements can take one of the following forms (or a combination of more than one if required for multiple datasets):

- The datasets generated and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS]
- The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.
- All data generated or analysed during this study are included in this published article [and its supplementary information files].
- The datasets generated and/or analysed during the current study are not publicly available due [REASON WHY DATA ARE NOT PUBLIC] but are available from the corresponding author on reasonable request.
- Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.
- The data that support the findings of this study are available from [third party name] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however

available from the authors upon reasonable request and with permission of [third party name].

- Not applicable. If your manuscript does not contain any data, please state 'Not applicable' in this section.

More examples of template data availability statements, which include examples of openly available and restricted access datasets, are available here.

BioMed Central also requires that authors cite any publicly available data on which the conclusions of the paper rely in the manuscript. Data citations should include a persistent identifier (such as a DOI) and should ideally be included in the reference list. Citations of datasets, when they appear in the reference list, should include the minimum information recommended by DataCite and follow journal style.

Dataset identifiers including DOIs should be expressed as full URLs. For example:

Hao Z, AghaKouchak A, Nakhjiri N, Farahmand A. Global integrated drought monitoring and prediction system (GIDMaPS) data sets.

figshare. 2014. <http://dx.doi.org/10.6084/m9.figshare.853801>

With the corresponding text in the Availability of data and materials statement:

The datasets generated during and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].<sup>[Reference number]</sup>

If you wish to co-submit a data note describing your data to be published in *BMC Research Notes*, you can do so by visiting our submission portal. Data notes support open data and help authors to comply with funder policies on data sharing. Co-published data notes will be linked to the research article the data support (example).

For more information please email our Research Data Team.

#### *Competing interests*

All financial and non-financial competing interests must be declared in this section.

See our editorial policies for a full explanation of competing interests.

If you are unsure whether you or any of your co-authors have a competing interest please contact the editorial office.

Please use the authors initials to refer to each authors' competing interests in this section.

If you do not have any competing interests, please state "The authors declare that they have no competing interests" in this section.

#### *Funding*

All sources of funding for the research reported should be declared.

The role of the funding body in the design of the study and collection,

analysis, and interpretation of data and in writing the manuscript should be declared.

#### *Authors' contributions*

The individual contributions of authors to the manuscript should be specified in this section. Guidance and criteria for authorship can be found in our editorial policies.

Please use initials to refer to each author's contribution in this section, for example: "FC analyzed and interpreted the patient data regarding the hematological disease and the transplant. RH performed the histological examination of the kidney, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript."

#### *Acknowledgements*

Please acknowledge anyone who contributed towards the article who does not meet the criteria for authorship including anyone who provided professional writing services or materials.

Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section.

See our editorial policies for a full explanation of acknowledgements and authorship criteria.

If you do not have anyone to acknowledge, please write "Not applicable" in this section.

Group authorship (for manuscripts involving a collaboration group): if you would like the names of the individual members of a collaboration Group to be searchable through their individual PubMed records, please ensure that the title of the collaboration Group is included on the title page and in the submission system and also include collaborating author names as the last paragraph of the "Acknowledgements" section. Please add authors in the format First Name, Middle initial(s) (optional), Last Name. You can add institution or country information for each author if you wish, but this should be consistent across all authors.

Please note that individual names may not be present in the PubMed record at the time a published article is initially included in PubMed as it takes PubMed additional time to code this information.

#### *Authors' information*

This section is optional.

You may choose to use this section to include any relevant information about the author(s) that may aid the reader's interpretation of the article, and understand the standpoint of the author(s). This may include details about the authors' qualifications, current positions they hold at institutions or societies, or any other relevant background information. Please refer to authors using their initials. Note this section should not be used to describe any competing interests.

#### *Footnotes*

Footnotes can be used to give additional information, which may include the citation of a reference included in the reference list. They should not consist solely of a reference citation, and they should never include the bibliographic details of a reference. They should also not contain any figures or tables.

Footnotes to the text are numbered consecutively; those to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data). Footnotes to the title or the authors of the article are not given reference symbols.

Always use footnotes instead of endnotes.

## References

Examples of the Vancouver reference style are shown below.

See our editorial policies for author guidance on good citation practice  
**Web links and URLs:** All web links and URLs, including links to the authors' own websites, should be given a reference number and included in the reference list rather than within the text of the manuscript. They should be provided in full, including both the title of the site and the URL, as well as the date the site was accessed, in the following format: The Mouse Tumor Biology Database.

<http://tumor.informatics.jax.org/mtbwi/index.do>. Accessed 20 May 2013. If an author or group of authors can clearly be associated with a web link, such as for weblogs, then they should be included in the reference.

Example reference style:

*Article within a journal*

Smith JJ. The world of science. *Am J Sci.* 1999;36:234-5.

*Article within a journal (no page numbers)*

Rohrmann S, Overvad K, Bueno-de-Mesquita HB, Jakobsen MU, Egeberg R, Tjønneland A, et al. Meat consumption and mortality - results from the European Prospective Investigation into Cancer and Nutrition. *BMC Medicine.* 2013;11:63.

*Article within a journal by DOI*

Slifka MK, Whitton JL. Clinical implications of dysregulated cytokine production. *Dig J Mol Med.* 2000; doi:10.1007/s801090000086.

*Article within a journal supplement*

Frumin AM, Nussbaum J, Esposito M. Functional asplenia: demonstration of splenic activity by bone marrow scan. *Blood* 1979;59 Suppl 1:26-32.

*Book chapter, or an article within a book*

Wyllie AH, Kerr JFR, Currie AR. Cell death: the significance of apoptosis. In: Bourne GH, Danielli JF, Jeon KW, editors. *International review of cytology.* London: Academic; 1980. p. 251-306.

*OnlineFirst chapter in a series (without a volume designation but with a DOI)*

Saito Y, Hyuga H. Rate equation approaches to amplification of enantiomeric excess and chiral symmetry breaking. *Top Curr Chem*. 2007. doi:10.1007/128\_2006\_108.

*Complete book, authored*

Blenkinsopp A, Paxton P. Symptoms in the pharmacy: a guide to the management of common illness. 3rd ed. Oxford: Blackwell Science; 1998.

*Online document*

Doe J. Title of subordinate document. In: The dictionary of substances and their effects. Royal Society of Chemistry. 1999.

[http://www.rsc.org/dose/title of subordinate document](http://www.rsc.org/dose/title%20of%20subordinate%20document). Accessed 15 Jan 1999.

*Online database*

Healthwise Knowledgebase. US Pharmacopeia, Rockville. 1998.

<http://www.healthwise.org>. Accessed 21 Sept 1998.

*Supplementary material/private homepage*

Doe J. Title of supplementary material. 2000.

<http://www.privatehomepage.com>. Accessed 22 Feb 2000.

*University site*

Doe, J: Title of preprint. <http://www.uni-heidelberg.de/mydata.html> (1999). Accessed 25 Dec 1999.

*FTP site*

Doe, J: Trivial HTTP, RFC2169. <ftp://ftp.isi.edu/in-notes/rfc2169.txt> (1999). Accessed 12 Nov 1999.

*Organization site*

ISSN International Centre: The ISSN register. <http://www.issn.org> (2006). Accessed 20 Feb 2007.

*Dataset with persistent identifier*

Zheng L-Y, Guo X-S, He B, Sun L-J, Peng Y, Dong S-S, et al. Genome data from sweet and grain sorghum (*Sorghum bicolor*). GigaScience Database. 2011. <http://dx.doi.org/10.5524/100012>.

## **Figures, tables and additional files**

See General formatting guidelines for information on how to format figures, tables and additional files.

## **General formatting guidelines**

- Preparing main manuscript text
- Preparing illustrations and figures
- Preparing tables
- Preparing additional files

## Preparing main manuscript text

Back to top

Quick points:

- Use double line spacing
- Include line and page numbering
- Use SI units: Please ensure that all special characters used are embedded in the text, otherwise they will be lost during conversion to PDF
- Do not use page breaks in your manuscript

File formats

The following word processor file formats are acceptable for the main manuscript document:

- Microsoft word (DOC, DOCX)
- Rich text format (RTF)
- TeX/LaTeX (use BioMed Central's TeX template)

Please note: editable files are required for processing in production. If your manuscript contains any non-editable files (such as PDFs) you will be required to re-submit an editable file when you submit your revised manuscript, or after editorial acceptance in case no revision is necessary.

Additional information for TeX/LaTeX users

Please use BioMed Central's TeX template and BibTeX stylefile if you use TeX format. Submit your references using either a bib or bbl file. When submitting TeX submissions, please submit both your TeX file and your bib/bbl file as manuscript files. Please also convert your TeX file into a PDF (please do not use a DIV file) and submit this PDF as a supplementary file with the name 'Reference PDF'. This PDF will be used by our production team as a reference point to check the layout of the article as the author intended.

The Editorial Manager system checks for any errors in the TeX files. If an error is present then the system PDF will display LaTeX code and highlight and explain the error in a section beginning with an exclamation mark (!).

All relevant editable source files must be uploaded during the submission process. Failing to submit these source files will cause unnecessary delays in the production process.

TeX templates
BioMedCentral_article (ZIP format) - preferred template
article (part of the standard TeX distribution)
amsart (part of the standard TeX distribution)

### Style and language

For editors and reviewers to accurately assess the work presented in your manuscript you need to ensure the English language is of sufficient quality to be understood. If you need help with writing in English you should consider:

- Visiting the English language tutorial which covers the common mistakes when writing in English.
- Asking a colleague who is a native English speaker to review your manuscript for clarity.
- Using a professional language editing service where editors will improve the English to ensure that your meaning is clear and identify problems that require your review. Two such services are provided by our affiliates Nature Research Editing Service and American Journal Experts. BMC authors are entitled to a 10% discount on their first submission to either of these services. To claim 10% off English editing from Nature Research Editing Service, [click here](#). To claim 10% off American Journal Experts, [click here](#).

Please note that the use of a language editing service is not a requirement for publication in the journal and does not imply or guarantee that the article will be selected for peer review or accepted.

为便于编辑和评审专家准确评估您稿件中陈述的研究工作，您需要确保文稿英语语言质量足以令人理解。如果您需要英文写作方面的帮助，您可以考虑：

- 查看一些有关英语写作中常见语言错误的教程。
- 请一位以英语为母语的同事审阅您的稿件是否表意清晰。
- 使用专业语言编辑服务，编辑人员会对英语进行润色，以确保您的意思表达清晰，并提出需要您复核的问题。例如我们的附属机构 Nature Research Editing Service 以及合作伙伴 American Journal Experts 都可以提供此类专业服务。BMC作者享受首次订单10%优惠，该优惠同时适用于两家公司。您只需点击以下链接即可开始。使用 Nature Research Editing Service的编辑润色10%的优惠服务，请点击[这里](#)。使用 American Journal Experts的10%优惠服务，请点击[这里](#)。

请注意，使用语言编辑服务并非在期刊上发表文章的必要条件，这也并不意味着或保证文章将被选中进行同行评议或被接受。

エディターと査読者があなたの論文を正しく評価するには、使用されている英語の質が十分であることが必要とされます。英語での論文執筆に際してサポートが必要な場合には、次のオプションがあります：

- 英語で執筆する際によくある間違いに関する英語のチュートリアルを参照する。
- 英語を母国語とする同僚に、原稿内の英語が明確であるかをチェックしてもらう。
- プロの英文校正サービスを利用する。校正者が原稿の意味を明確にしたり、問題点を指摘し、英語を向上させます。Nature Research Editing Service と American Journal Experts の2つは弊社と提携しているサービスです。BMCのジャーナルの著者は、いずれかのサービスを初めて利用する際に、10%の割引を受けることができます。Nature Research Editing Serviceの10%割引を受けるには、[こちら](#)をクリックしてください。American Journal Expertsの10%割引を受けるには、[こちら](#)をクリックしてください。

英文校正サービスの利用は、このジャーナルに掲載されるための条件ではないこと、また論文審査や受理を保証するものではないことに留意してください。

영어 원고의 경우, 에디터 및 리뷰어들이 귀하의 원고에 실린 결과물을 정확하게 평가할 수 있도록, 그들이 충분히 이해할 수 있을 만한 수준으로 작성되어야 합니다. 만약 영작문과 관련하여 도움을 받기를 원하신다면 다음의 사항들을 고려하여 주십시오:

- 영어 튜토리얼 페이지에 방문하여 영어로 글을 쓸 때 자주하는 실수들을 확인합니다.
- 귀하의 원고의 표현을 명확히 해줄 영어 원어민 동료들 찾아서 리뷰를 의뢰합니다
- 리뷰에 대비하여, 원고의 의미를 명확하게 해주고 리뷰에서 요구하는 문제점들을 식별해서 영문 수준을 향상시켜주는 전문 영문 교정 서비스를 이용합니다. Nature Research Editing Service와 American Journal Experts에서 저희와 협약을 통해 서비스를 제공하고 있습니다. BMC에서는 위의 두 가지의 서비스를 첫 논문 투고를 위해 사용하시는 경우, 10%의 할인을 제공하고 있습니다. Nature Research Editing Service이용시 10% 할인을 요청하기 위해서는 [여기를 클릭해 주시고](#), American Journal Experts 이용시 10% 할인을 요청하기 위해서는 [여기를 클릭해 주십시오](#).

영문 교정 서비스는 게재를 위한 요구사항은 아니며, 해당 서비스의 이용이 피어 리뷰에 논문이 선택되거나 게재가 수락되는 것을 의미하거나 보장하지 않습니다.

#### Data and materials

For all journals, BioMed Central strongly encourages all datasets on which the conclusions of the manuscript rely to be either deposited in publicly available repositories (where available and appropriate) or presented in the main paper or additional supporting files, in machine-readable format (such as spread sheets rather than PDFs) whenever

possible. Please see the list of recommended repositories in our editorial policies.

For some journals, deposition of the data on which the conclusions of the manuscript rely is an absolute requirement. Please check the Instructions for Authors for the relevant journal and article type for journal specific policies.

For all manuscripts, information about data availability should be detailed in an 'Availability of data and materials' section. For more information on the content of this section, please see the Declarations section of the relevant journal's Instruction for Authors. For more information on BioMed Central's policies on data availability, please see our [editorial policies].

*Formatting the 'Availability of data and materials' section of your manuscript*

The following format for the 'Availability of data and materials' section of your manuscript should be used:

"The dataset(s) supporting the conclusions of this article is(are) available in the [repository name] repository, [unique persistent identifier and hyperlink to dataset(s) in http:// format]."

The following format is required when data are included as additional files:

"The dataset(s) supporting the conclusions of this article is(are) included within the article (and its additional file(s))."

BioMed Central endorses the Force 11 Data Citation Principles and requires that all publicly available datasets be fully referenced in the reference list with an accession number or unique identifier such as a DOI.

For databases, this section should state the web/ftp address at which the database is available and any restrictions to its use by non-academics.

For software, this section should include:

- Project name: e.g. My bioinformatics project
- Project home page: e.g. <http://sourceforge.net/projects/mged>
- Archived version: DOI or unique identifier of archived software or code in repository (e.g. enodo)
- Operating system(s): e.g. Platform independent
- Programming language: e.g. Java
- Other requirements: e.g. Java 1.3.1 or higher, Tomcat 4.0 or higher
- License: e.g. GNU GPL, FreeBSD etc.
- Any restrictions to use by non-academics: e.g. licence needed

Information on available repositories for other types of scientific data, including clinical data, can be found in our editorial policies.

### References

See our editorial policies for author guidance on good citation practice.

Please check the submission guidelines for the relevant journal and article type.

#### *What should be cited?*

Only articles, clinical trial registration records and abstracts that have been published or are in press, or are available through public e-print/preprint servers, may be cited.

Unpublished abstracts, unpublished data and personal communications should not be included in the reference list, but may be included in the text and referred to as "unpublished observations" or "personal communications" giving the names of the involved researchers.

Obtaining permission to quote personal communications and unpublished data from the cited colleagues is the responsibility of the author. Only footnotes are permitted. Journal abbreviations follow Index Medicus/MEDLINE.

Any in press articles cited within the references and necessary for the reviewers' assessment of the manuscript should be made available if requested by the editorial office.

#### How to format your references

Please check the Instructions for Authors for the relevant journal and article type for examples of the relevant reference style.

**Web links and URLs:** All web links and URLs, including links to the authors' own websites, should be given a reference number and included in the reference list rather than within the text of the manuscript. They should be provided in full, including both the title of the site and the URL, as well as the date the site was accessed, in the following format: The Mouse Tumor Biology Database.

<http://tumor.informatics.jax.org/mtbwi/index.do>. Accessed 20 May 2013. If an author or group of authors can clearly be associated with a web link, such as for weblogs, then they should be included in the reference.

Authors may wish to make use of reference management software to ensure that reference lists are correctly formatted.

### Preparing figures

#### Back to top

When preparing figures, please follow the formatting instructions below.

- Figures should be numbered in the order they are first mentioned in the text, and uploaded in this order. Multi-panel

figures (those with parts a, b, c, d etc.) should be submitted as a single composite file that contains all parts of the figure.

- Figures should be uploaded in the correct orientation.
- Figure titles (max 15 words) and legends (max 300 words) should be provided in the main manuscript, not in the graphic file.
- Figure keys should be incorporated into the graphic, not into the legend of the figure.
- Each figure should be closely cropped to minimize the amount of white space surrounding the illustration. Cropping figures improves accuracy when placing the figure in combination with other elements when the accepted manuscript is prepared for publication on our site. For more information on individual figure file formats, see our detailed instructions.
- Individual figure files should not exceed 10 MB. If a suitable format is chosen, this file size is adequate for extremely high quality figures.
- Please note that it is the responsibility of the author(s) to obtain permission from the copyright holder to reproduce figures (or tables) that have previously been published elsewhere. In order for all figures to be open access, authors must have permission from the rights holder if they wish to include images that have been published elsewhere in non open access journals. Permission should be indicated in the figure legend, and the original source included in the reference list.

### Figure file types

We accept the following file formats for figures:

- EPS (suitable for diagrams and/or images)
- PDF (suitable for diagrams and/or images)
- Microsoft Word (suitable for diagrams and/or images, figures must be a single page)
- PowerPoint (suitable for diagrams and/or images, figures must be a single page)
- TIFF (suitable for images)
- JPEG (suitable for photographic images, less suitable for graphical images)
- PNG (suitable for images)
- BMP (suitable for images)
- CDX (ChemDraw - suitable for molecular structures)

For information and suggestions of suitable file formats for specific figure types, please see our author academy.

Figure size and resolution

Figures are resized during publication of the final full text and PDF versions to conform to the BioMed Central standard dimensions, which are detailed below.

Figures on the web:

- width of 600 pixels (standard), 1200 pixels (high resolution).

Figures in the final PDF version:

- width of 85 mm for half page width figure
- width of 170 mm for full page width figure
- maximum height of 225 mm for figure and legend
- image resolution of approximately 300 dpi (dots per inch) at the final size

Figures should be designed such that all information, including text, is legible at these dimensions. All lines should be wider than 0.25 pt when constrained to standard figure widths. All fonts must be embedded.

*Figure file compression*

- Vector figures should if possible be submitted as PDF files, which are usually more compact than EPS files.
- TIFF files should be saved with LZW compression, which is lossless (decreases file size without decreasing quality) in order to minimize upload time.
- JPEG files should be saved at maximum quality.
- Conversion of images between file types (especially lossy formats such as JPEG) should be kept to a minimum to avoid degradation of quality.

If you have any questions or are experiencing a problem with figures, please contact the customer service team at [info@biomedcentral.com](mailto:info@biomedcentral.com).

## Preparing tables

Back to top

When preparing tables, please follow the formatting instructions below.

- Tables should be numbered and cited in the text in sequence using Arabic numerals (i.e. Table 1, Table 2 etc.).
- Tables less than one A4 or Letter page in length can be placed in the appropriate location within the manuscript.
- Tables larger than one A4 or Letter page in length can be placed at the end of the document text file. Please cite and indicate

where the table should appear at the relevant location in the text file so that the table can be added in the correct place during production.

- Larger datasets, or tables too wide for A4 or Letter landscape page can be uploaded as additional files. Please see [below] for more information.
- Tabular data provided as additional files can be uploaded as an Excel spreadsheet (.xls ) or comma separated values (.csv). Please use the standard file extensions.
- Table titles (max 15 words) should be included above the table, and legends (max 300 words) should be included underneath the table.
- Tables should not be embedded as figures or spreadsheet files, but should be formatted using 'Table object' function in your word processing program.
- Color and shading may not be used. Parts of the table can be highlighted using superscript, numbering, lettering, symbols or bold text, the meaning of which should be explained in a table legend.
- Commas should not be used to indicate numerical values.

If you have any questions or are experiencing a problem with tables, please contact the customer service team at [info@biomedcentral.com](mailto:info@biomedcentral.com).

## Preparing additional files

Back to top

As the length and quantity of data is not restricted for many article types, authors can provide datasets, tables, movies, or other information as additional files.

All Additional files will be published along with the accepted article.

Do not include files such as patient consent forms, certificates of language editing, or revised versions of the main manuscript document with tracked changes. Such files, if requested, should be sent by email to the journal's editorial email address, quoting the manuscript reference number. Please do not send completed patient consent forms unless requested.

Results that would otherwise be indicated as "data not shown" should be included as additional files. Since many web links and URLs rapidly become broken, BioMed Central requires that supporting data are included as additional files, or deposited in a recognized repository. Please do not link to data on a personal/departmental website. Do not include any individual participant details. The maximum file size for additional files is 20 MB each, and files will be virus-scanned on submission. Each additional file should be cited in sequence within the main body of text.

If additional material is provided, please list the following information in a separate section of the manuscript text:

- File name (e.g. Additional file 1)
- File format including the correct file extension for example .pdf, .xls, .txt, .pptx (including name and a URL of an appropriate viewer if format is unusual)
- Title of data
- Description of data

Additional files should be named "Additional file 1" and so on and should be referenced explicitly by file name within the body of the article, e.g. 'An additional movie file shows this in more detail [see Additional file 1]'.

For further guidance on how to use Additional files or recommendations on how to present particular types of data or information, please see How to use additional files.